

Retraction

Retracted: A K-Means Clustering Algorithm for Early Warning of Financial Risks in Agricultural Industry

Security and Communication Networks

Received 8 January 2024; Accepted 8 January 2024; Published 9 January 2024

Copyright © 2024 Security and Communication Networks. 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.

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

- (1) Discrepancies in scope
- (2) Discrepancies in the description of the research reported
- (3) Discrepancies between the availability of data and the research described
- (4) Inappropriate citations
- (5) Incoherent, meaningless and/or irrelevant content included in the article
- (6) Manipulated or compromised peer review

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

In addition, our investigation has also shown that one or more of the following human-subject reporting requirements has not been met in this article: ethical approval by an Institutional Review Board (IRB) committee or equivalent, patient/participant consent to participate, and/or agreement to publish patient/participant details (where relevant).

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

- [1] X.-T. Li and X.-H. Duan, "A K-Means Clustering Algorithm for Early Warning of Financial Risks in Agricultural Industry," *Security and Communication Networks*, vol. 2022, Article ID 3751539, 9 pages, 2022.

Research Article

A K-Means Clustering Algorithm for Early Warning of Financial Risks in Agricultural Industry

Xue-Tong Li¹ and Xiao-Hua Duan² 

¹College of Economics and Management, Xi'an Kedagaoxin University, Xi'an 710109, China

²School of Finance, Xi'an Eurasia University, Xi'an 710065, China

Correspondence should be addressed to Xiao-Hua Duan; duanxiaohua@eurasia.edu.cn

Received 6 March 2022; Revised 24 March 2022; Accepted 4 April 2022; Published 28 April 2022

Academic Editor: Fang Liu

Copyright © 2022 Xue-Tong Li and Xiao-Hua Duan. 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.

As the primary industry, agricultural industry is the basis of guaranteeing people's basic life and national economic development. Agricultural industrial finance/financial is a weak link in the financial system, which seriously hinders the emergence of agricultural scale effect and the improvement in agricultural production efficiency. In order to find the financial risk of agricultural industry in time, this article proposes an agricultural industry financial risk early warning system based on improved K-means clustering algorithm. Because the traditional K-means algorithm is easy to fall into local optimization in the clustering process, the clustering effect is not reliable. In this article, the idea of immune cloning and particle swarm optimization location update is added to the grey wolf optimization algorithm. Grey wolf optimization algorithm and K-means algorithm are combined to solve the problem that K-means algorithm is easy to fall into local optimization. In the experimental part, through comparative verification, it can be found that the prediction performance of this system is superior to other models. Its practical application value is higher. Therefore, choosing this system for early warning of agricultural industry finance can effectively improve the accuracy of early warning and provide guarantee for the economic development of agricultural industry.

1. Introduction

Agricultural industry finance is a relatively weak link in China's financial system. The financial demand in rural areas mainly includes the demand for productive loans, including planting, breeding, animal husbandry, deep processing of agricultural products, etc. It has long cycle and low income and is greatly affected by natural factors. Agriculture-related loans have instability risk and high service cost. The supply of rural formal finance is insufficient, and the structural contradiction is more prominent. Its financial ecology is fragile and the risk of guarantee chain is large. At the same time, the insufficient supply of rural formal finance provides space for the development of underground finance. Some individuals are engaged in insurance and deposit business under the guise of Internet finance and farmer cooperatives, so that illegal fundraising problems often occur. Therefore,

the study of rural financial risk has very important practical significance.

There are four main causes of rural financial risk in China [1]. First, the uncertainty of agricultural economy. At present, there are some problems in China's agricultural economy, such as scattered scale, ineffective formation of production chain, low level of agricultural science and technology, and unpredictable natural disasters. Therefore, farmers' production and income are reduced, their ability to repay loans is weakened, and the risk of small agricultural loans is formed. Second, the resource allocation of agricultural financial institutions is unreasonable. The cost and income of financial service supply are upside down. Some state-owned commercial banks are shrinking in rural areas, and the service object and scope of policy banks are relatively single. The governance mechanism of rural credit cooperatives has not been fundamentally solved. The level of

internal control management is relatively low, the electro-nization of rural banks and the construction of core business system are backward, and the quality of employees needs to be improved. Third, the rural credit environment is imperfect and lacks an effective punishment mechanism for dishonesty. Intermediary services are not standardized, providing false credit certificates. Meanwhile, fraudulent loans and malicious evasion of debts occur from time to time. Fourth, the construction of risk early warning system and prevention mechanism of rural financial institutions is not in place, so it is difficult to find and warn rural financial risks in time.

In order to predict the agricultural industry finance, domestic and overseas scholars at home and abroad have made positive contributions [2]. As for the risk of accounts receivable financing, some scholars have proposed that most of the customers of global accounts receivable financing are small and medium-sized enterprises, and there are hidden risks in the international accounts receivable factoring business [3]. The longer the maturity time of accounts receivable, the higher the risk. The provider of accounts receivable financing funds should also pay attention to the risk of changes in the value of suppliers' accounts receivable [4]. Meanwhile, there are some problems in accounts receivable financing, such as legal supervision, bank enterprise information asymmetry, etc [5]. The smooth implementation of accounts receivable financing of commercial banks needs to reduce the moral hazard of core enterprises and carry out dynamic risk monitoring of financing enterprises [6]. Because China is a large agricultural country, the development of agricultural industry is largely limited by the supply chain transportation industry. Therefore, the development of supply chain finance business affects agricultural finance business. As for the research on risk assessment with supply chain financial business as a whole, scholars choose the assessment indicators of supply chain financial risk from the aspects of core enterprises, financing enterprises, financing projects, and supply chain operation [7]. At the same time, the methods of supply chain financial risk assessment mainly include objective methods, support vector machine (SVM), fuzzy comprehensive evaluation method, game model, entropy weight TOPSIS model [8–11]. Due to the different methods, indicators and samples used by scholars in the research of supply chain financial risk, there are some differences in the composition and evaluation results of supply chain financial risk. However, it is generally believed that it is necessary to pay attention to the possible risks in the supply chain financial business and take corresponding preventive measures. Under different supply chain financial models, the focus of risk evaluation indicators may be different. Therefore, this article takes agricultural listed enterprises of small and medium-sized enterprises as an example and establishes a risk evaluation index system of accounts receivable financing of small and medium-sized enterprises based on fully considering the characteristics of accounts receivable financing mode. In this way, the credit risk of accounts receivable financing of small and medium-sized enterprises can be truly reflected and the decision-making basis can be provided for the practice of accounts receivable financing of small and medium-sized enterprises.

In order to improve the early warning ability of agricultural financial risk, this article improves the problem that K-means clustering algorithm is easy to local optimal solution and proposes an agricultural financial risk early warning system based on improved K-means clustering. The innovations and contributions of this article are listed below:

- (1) The elite individuals in the iterative process of grey wolf optimization algorithm are deeply mined to improve the in-depth exploration ability of grey wolf optimization algorithm and avoid the premature convergence of grey wolf optimization algorithm.
- (2) In order to expand the scope of prey search in the field of elite individuals and give full play to the residual value of elite individuals, the idea of monomer position update of particle swarm optimization algorithm is combined with the original grey wolf position update. The improved grey wolf optimization algorithm is combined with the classical K-means algorithm to solve the problem that K-means algorithm is easy to fall into local optimization.
- (3) The financial risk early warning experiment of agricultural industry is carried out to verify the effectiveness of this system.

The rest of the article is organized as follows: Section 2 details the financial risk prediction model of agricultural industry based on K-means clustering algorithm, Section 3 illustrates the system design, Section 4 is about experimental simulation and analysis, and Section 5 is the final concluding section of the article.

2. Financial Risk Prediction Model of Agricultural Industry Based on K-Means Clustering Algorithm

2.1. Grey Wolf Optimization Algorithm (GWO). GWO algorithm is a mathematical expression of social hierarchy and predation behaviour in grey wolf population. The hierarchy of GWO algorithm is divided into four, namely α , β , δ , and ω Wolf. Let the grey wolf population with population size n be: $I = \{i_1, i_2, \dots, i_T\}$. The best value of the candidate solution in the grey wolf population is taken as α Wolf, second best value as β Wolf, the third best value as δ Wolf, the remaining candidate solutions are set to ω Wolf. In GWO algorithm, by α , β , and δ as a leader, the wolf searches for the optimal solution within the specified range, and ω under the leadership of the three wolves, the wolves updated their positions. The mathematical model of GWO algorithm is as follows:

In the process of grey wolf searching prey, the distance between each wolf and prey can be expressed by the following formula:

$$D = |C \cdot I_u(n) - I(n)|, \quad (1)$$

$$I(n+1) = I_u(n) - G \cdot D, \quad (2)$$

where $I_u(n)$ represents the position vector of prey, $I(n)$ represents the position vector of grey wolf, D represents the distance vector between grey wolf and prey, and N represents the number of iterations. The calculation formulas of coefficients G and C are as follows:

$$G = 2g \cdot r_1 - g, \quad (3)$$

$$C = 2 \cdot r_2. \quad (4)$$

In the process of encirclement, the value of g linearly from 2 to 0, and r_1 and r_2 are random vectors in $[0, 1]$.

Grey wolves have the ability to judge the location of prey and surround prey. Preservation α , β , and δ wolves are the first three best solutions obtained in the population. And force other search agents (ω (Wolf)) according to α , β , δ the wolf's position updates its position. Other search agents and α , β , δ the wolf distance can be expressed by the following formula:

$$\begin{cases} D_\alpha = |C_1 \cdot I_\alpha - I|, \\ D_\beta = |C_2 \cdot I_\beta - I|, \\ D_\delta = |C_3 \cdot I_\delta - I|. \end{cases} \quad (5)$$

After the distance of each wolf is obtained, the individual position of grey wolf is updated through the following formulas:

$$\begin{cases} I_1 = I_\alpha - G_1 \cdot (D_\alpha), \\ I_2 = I_\beta - G_2 \cdot (D_\beta), \\ I_3 = I_\delta - G_3 \cdot (D_\delta), \end{cases} \quad (6)$$

$$I(n+1) = \frac{I_1 + I_2 + I_3}{3}. \quad (7)$$

Although GWO algorithm has strong search ability, the population diversity is decreasing with the increase of iteration times. The difference between individuals is getting smaller and smaller, and the optimal value cannot be found in the search space. Premature convergence may occur, which will affect the performance of GWO algorithm. Therefore, GWO algorithm is improved based on immune cloning theory and particle swarm optimization position update idea. Immune cloning selects elite individuals from the population and performs cloning and mutation operation on them to increase the diversity of the population and avoid premature convergence of the algorithm. Then, the position change idea of a single grey wolf is introduced to increase a certain mutation ability to the change of grey wolf position, so as to improve the global search ability of the algorithm.

2.2. Immune Clone Selection Operation. The essence of immune clonal selection is to select elite individuals from the population according to individual fitness value, and clone and mutate the elite individuals to form a new species group. Then elite individuals are selected from the new population

to enter the next iteration until the maximum iteration number of immune clone selection is reached. Applying it to the grey wolf optimization algorithm is a more in-depth exploration of the elite individuals in the original grey wolf population, so as to expand the search scope and improve the population diversity. The detailed steps of clone selection are as follows:

Step 1: select m individuals with good fitness from the grey wolf population according to the fitness function value to form an elite population (set the m value to 1/4 of the number of grey wolf individuals).

Step 2: clone all grey wolf individuals in the elite population. The clone size is directly proportional to the number m of the selected elite population to form a temporary population T with the size of NC. The calculation of NC is as follows:

$$T_c = \sum_{x=1}^w \text{round}\left(\frac{\lambda w}{x} + h\right). \quad (8)$$

The round () function is a rounding function. λ is a random number between $[0, 1]$. H is an integer constant and $h \geq 1$. Compared with the original population t , the size of T_c is positively correlated with the value of h , which can ensure that each individual in the elite population has a certain number of clones.

In step 3, high-frequency mutation is implemented for each individual in the population to obtain better candidate solutions near elite individuals. The variation operation is shown in the following formulas:

$$n_x^{\text{new}} = n_x + u \cdot \eta \cdot n_x \cdot r_4 - u \cdot \eta \cdot n_x \cdot r_5, \quad (9)$$

$$u = \begin{cases} 1, & r_6 \leq 0.5, \\ 0, & \text{other}, \end{cases} \quad (10)$$

$$\eta = 1 - (1/(\exp(1) - 1)) \times (\exp(x/(x_{\max} - 1)) - 1), \quad (11)$$

where n_x is the individual of the x th iteration of population n ; n_x^{new} is a new individual produced by n_x after mutation operation; r_4, r_5, r_6 are random numbers between $[0, 1]$; X stands for iteration x ; x_{\max} represents the maximum number of iterations of immune cloning operation; and η is a clonal variation parameter. It can be seen from formula (11) that the number of iterations is negatively correlated with the clone variation parameter η . η is close to 1 at the beginning, with a wide range of variation. At this point, a global range search is performed to ensure population diversity. As the number of iterations increases, the value of η gets closer and closer to 0, indicating that local searches are performed within a small range to increase fine-tuning capability and ensure the accuracy of the search.

Step 4: select w better individuals from N as the elite individual population for the next iteration until the maximum number of iterations of immune cloning operation is reached.

2.3. Particle Swarm Location Update Idea. In the grey wolf optimization algorithm, it can be seen from formula (5) that the position change of grey wolf is mainly to explore the prey position according to the position of three wolves. Then, three wolves (α , β , and δ) led the position update. Because we are now facing the elite population, the information contained in the prey search results of each elite individual may have an impact on the search results of the final location of the prey. Therefore, it is necessary to consider the location information of a single elite individual, so as to maximize the utilization rate of elite individuals and expand the search range of prey around elite individuals.

This article is inspired by the idea of location update of particle swarm optimization algorithm. It introduces the position change idea of a single grey wolf into the grey wolf position update to avoid premature convergence of the algorithm. Adjust the update strategy of formula (7) accordingly as follows.

$$I(n+1) = \frac{I_1 + I_2 + I_3}{3} + q(n), \quad (12)$$

where the position change $q(n)$ of a single grey wolf is expressed as the following formula:

$$q(n+1) = m(q(n) + C_1 \cdot r_7(I_1 - I(n)) + C_2 \cdot r_8(I_2 - I(n)) + C_3 \cdot r_9(I_3 - I(n))), \quad (13)$$

where m is the random number between $[0, 1]$. After many simulation experiments, when the value of m is between $[0.6, 1]$, the algorithm has better search performance and more accurate optimization results. When m is large, it has better global search ability. When w is small, the local search ability is strong, which can effectively avoid premature convergence. r_6, r_7, r_8 are random numbers between $[0, 1]$. C_1, C_2, C_3 are obtained from formula (4). I_1, I_2, I_3 are obtained from formula (6). $I(n)$ represents the current position of the grey wolf.

2.4. Clustering Algorithm Based on GWO and K-Means. Because the input data of agricultural financial risk is usually in the format of text document, and the text document belongs to unstructured data, this article needs to preprocess the text document before text clustering. Convert text data type to data that can be input by GWO K-means algorithm. The basic steps of text preprocessing are text word segmentation, removal of stop words, text feature selection, and text vectorization.

This article uses Jieba word segmentation in Python to segment text documents and remove stop words. Commonly used text representation models mainly include Boolean space model (BM) [12], suffix tree model (STM) [13], vector space model (VSM) [14], and probabilistic retrieval model (PM) [15]. In this article, the most classical text vector model (VSM) is used for text vectorization. For document D , it is represented by $(nf - xdf_1, nf - xdf_2, \dots, nf - xdf_t)$. The calculation formula of $nf - xdf$ is as follows:

$$nf - xdf(n, D) = \frac{lh(nf(n, D) + 0.1) \times lh(T/t_x)}{\sqrt{\sum_{x=1}^t [lh(nf(n, D) + 0.1) \times lh(T/t_x)]^2}}, \quad (14)$$

where t_x is the number of documents containing the word n_x , T is the total number of documents, and $nf(n, d)$ indicates the number of times the word n_x appears in document D .

In this article, GWO and K-means algorithm are used for text clustering analysis. Specifically, GWO algorithm is used to find a group of optimal clustering centres to minimize the distance from all texts in each category to the group of clustering centres, that is, the similarity of each document is the largest. In GWO algorithm, the fitness function is the goal of grey wolf to find the optimal solution. In K-means algorithm, the sum of intraclass distances is an important index to measure the advantages and disadvantages of clustering algorithm. The smaller the value, the better the clustering performance. The purpose of the combination of GWO and K-means algorithm is to use the powerful optimization ability of GWO algorithm to accurately find the optimal clustering centre. Text documents are classified by the clustering centre. This article selects the sum of the intraclass distances between text documents as the fitness evaluation function of GWO algorithm. As shown in formula:

$$Y = \sum_{y=1}^z \sum_{s_x \in c_y} d(s_x, c_y)^2, \quad (15)$$

where z stands for clustering category.

3. System Design of This Article

The structure of agricultural industry financial risk early warning system should be considered from three aspects. It includes acquisition system, which is used for data acquisition and input. It specifically includes macro financial data, risk event data, etc. There is also a database system for inductive analysis of data. And the most critical early warning system for risk prediction and monitoring (as shown in Figure 1).

In the process of information collection, it is necessary to adopt the means of data mining, assisted by manual intervention to collect and sort out the data. The direction of data mining can include crawler technology and scanning monitoring technology, such as risk event crawler and scanning monitoring of risk indicators. The architecture of the acquisition system is shown in Figure 2.

The amount of data collected is huge and comes from a wide range of sources, which cannot be measured by a unified standard. Therefore, further induction and analysis are needed, and the rationality and function of the database system are highlighted. By means of improved K-means clustering algorithm, the early warning model is trained through a large number of data to improve the early warning of agricultural financial risk early warning system. The system architecture is shown in Figure 3.

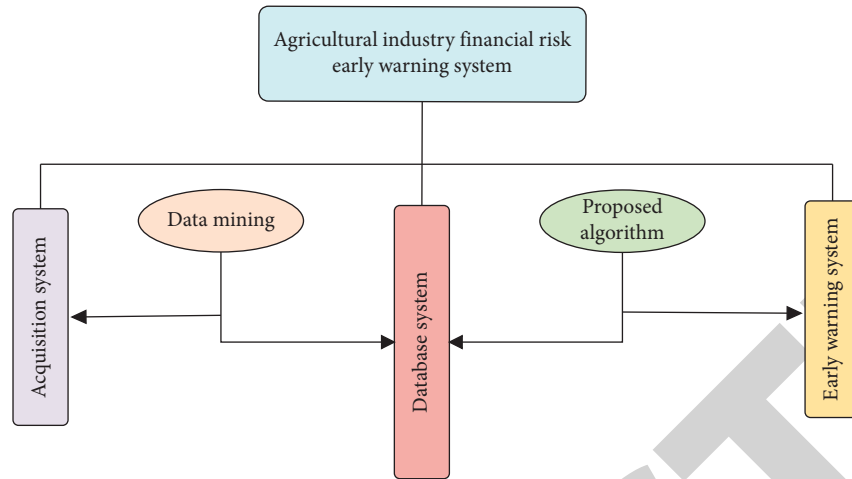


FIGURE 1: The structure diagram of agricultural industry financial risk early warning system.

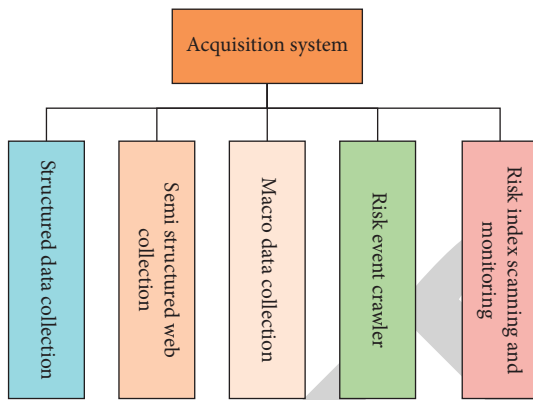


FIGURE 2: Acquisition system architecture diagram.

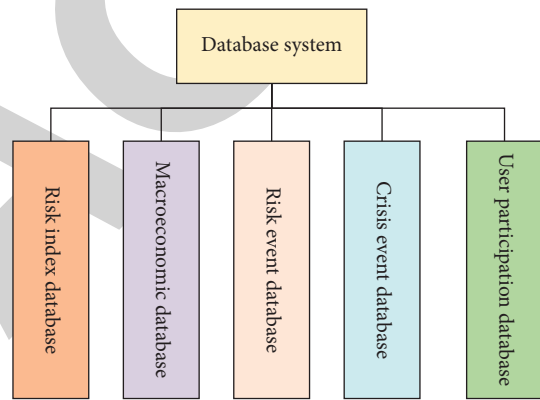


FIGURE 3: The database system architecture.

Early warning system is the explicit output part of the whole agricultural industry financial risk early warning system. Its coverage include risk analysis, prediction analysis, and early warning tracking. Considering that there may be some deviation between the predicted results and the actual results, it is necessary to introduce technical means such as improved K-means clustering algorithm in this article. The predicted results and the actual results are constantly compared and analyzed, and the results are constantly optimized through a large amount of data training. Finally, the prediction results are constantly improved, and the scientificity and accuracy of the results are also improved. Considering that there are many sudden risk factors in financial activities, it is necessary to rely on human intervention. Moreover, the changes of financial activities are very fast and are constantly updated and iterated. Therefore, the early warning system should also be accompanied by a correction algorithm. Continuously iterate and upgrade to optimize the system. The architecture of the early warning system is shown in Figure 4. In the training stage, the model is implemented by human participation. In the experimental testing stage, it is learned and corrected by the model itself.

4. Experiment and Analysis

4.1. *Data Collection and Sorting.* A total of 3,812 enterprise data sets of listed companies from 2,000 to 2,020 were extracted from Cathay Pacific and other databases. After data cleaning, enterprise data sets are divided according to whether they are in ST (special treatment) state, and 472 ST enterprise data sets and 3,167 normal enterprise data sets are obtained. In addition, according to the company information recorded in the data set, we sorted out the high-dimensional feature collection based on the financial characteristics of agricultural industry and nonagricultural industry. This study divides the characteristics of enterprises by agricultural industrial financial indicators and nonagricultural industrial financial indicators. The training set and test set are obtained by multilevel division based on business ability, profitability, growth ability, and management structure.

4.2. *Characteristic Causal Analysis.* In the ST enterprise data set, if the company is repeatedly in ST within the time range studied in this article, the timeline of being in ST for the first

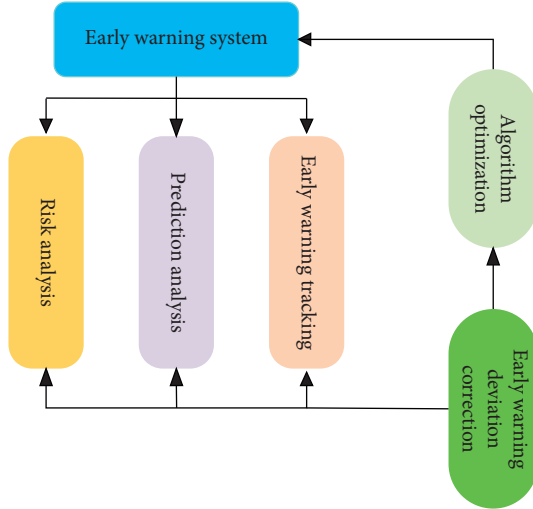


FIGURE 4: Architecture of early warning system.

time shall prevail. According to the timeline of ST for the first time or the timeline recorded in the normal data set, the stability of ST data set and normal data set is tested. After the stationarity test is completed, the code is constructed according to the dimensionality reduction steps of the model in this study, and the low-dimensional feature collection corresponding to the high-dimensional feature table is obtained through iterative training.

4.3. Integrated Classifier Training. As shown in Figure 2, the low-dimensional feature collection (including nonagricultural industrial financial low-dimensional feature collection and agricultural industry financial low-dimensional feature collection) is used as the first training data set to train the integrated classifier. By monitoring the function value object^K of the objective function, the curve between the function value and the training times is constructed. At the same time, in order to compare the excellence of low-dimensional feature collection with high-dimensional feature collection in the training process of integrated classifier, high-dimensional feature collection (including nonagricultural industrial finance high-dimensional feature collection and agricultural industrial finance high-dimensional feature collection) is taken as the second training data set. The integrated classifier is trained to get the training comparison diagram of low-dimensional feature collection and high-dimensional feature collection in the integrated classifier (see Figure 5).

According to Figure 3, when using low-dimensional feature collection to train the integrated classifier, the training times reach 700, the change range of the target value tends to be stable. The average value of the target value is 0.0519, and the training of the integrated classifier is completed. When using high-dimensional feature collection to train the integrated classifier, the change range of the target value tends to be stable after 900 times of training. And the average value of the target value is 0.0636. It can be seen that the low-dimensional feature collection obtained by using the feature causal analysis in this model is more

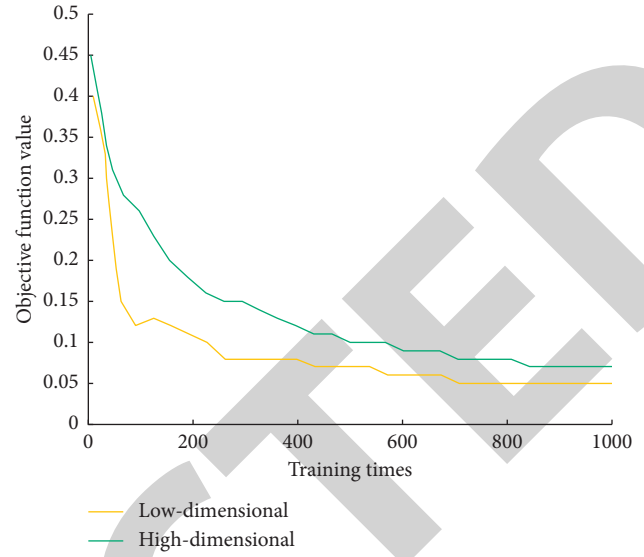


FIGURE 5: Training of high-dimensional and low-dimensional integrated classifier.

conducive to the training of integrated classifiers in terms of training times, target value, and stability.

Literature [16] and literature [17] are relatively typical models with feature analysis and financial early warning function of agricultural industry. In order to explore the comparative results of this model with literature [16] and [17] model, data enrichment was performed on the second training data set above to obtain the third training data set. Proposed model, literature [16] model, and literature [17] model were successively trained to obtain the comparison figure (see Figure 6).

According to Figure 6, in the training cycle, the model cycle in this study is the longest, and the objective function value needs to reach about 1,500 times before it tends to be stable. Literature [17] takes the second place, and it takes about 1,000 training times for the value of the objective function to stabilize. Literature [16] is at least 1,000 times. In terms of objective function value, when the training of each model is completed, the objective function value of the model in this study is the smallest, with an average value of 0.055. In literature [17], the mean value of objective function is 0.081 and in literature [16] it is 0.067. The model in this study has the best adaptability to the third training data set and the highest early warning accuracy.

4.4. Comparative Analysis of Financial Risk Models. After the model training is completed, this study selects the comparison model based on machine learning and deep learning. Among them, the machine learning model selects the literature [18] model with simple model structure and fast training speed and adds the literature [16] model with stationarity test to improve the robustness of the model. The literature [17] model of dimension reduction operation is added under the low-dimensional feature set. Literature [19] model with excellent performance is added under high-dimensional feature collection. For the in-depth learning

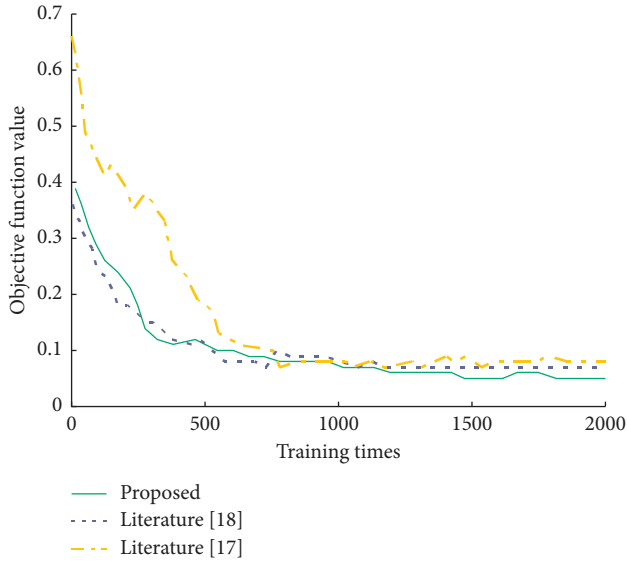


FIGURE 6: Training comparison of proposed model.

model, the typical literature [20] model is selected. Through comparison, the actual early warning performance of this model is further evaluated.

First, the test data set including 83 dimensions is divided into three groups, and the accuracy of each group is tested by each model (as shown in Table 1).

Combined with the performance of each model in Table 2, we can see the difference between machine learning and deep learning. Compared with machine learning, deep learning has obvious advantages in image and natural language [39, 40]. However, in the direction of corporate agricultural industry financial early warning, the early warning performance of the early warning model based on machine learning is not necessarily worse than that of the deep learning model. For example, in literature [17], its training accuracy and test accuracy are obviously superior to literature [20]. The early warning model constructed by a single improved K-means clustering algorithm generally does not have the function of feature screening, such as literature [18]. However, the shortcomings of not having feature screening can be overcome by combining models, such as literature [16], literature [17], etc. Moreover, the early warning performance of the combined model can generally be further improved.

Combined with (1) and (2), in terms of corporate agricultural industry financial early warning, compared with the deep learning model, the clustering algorithm proposed in this article has a better early warning performance. This further verifies the feasibility of combining the grey Wolf optimization algorithm with K-means algorithm for cluster analysis.

From the perspective of the specific performance of each model, the six groups of models can realize the role of agricultural industry financial early warning. However, compared with the other five groups of models, the accuracy of this model is higher. On the premise that each group of models is trained 2000 times, the training accuracy of this model reaches the highest 96.59%, which is superior to the

second best performance literature [17]. In addition, this model also has better feature screening ability. Compared with other models that also have feature screening, such as literature [20], literature [17], literature [16], etc., the model test accuracy of this study is ahead of other groups of models, reaching 86.77%. Therefore, generally speaking, proposed model has strong feature screening ability and effectively improve the accuracy of the company's agricultural industry financial early warning.

The impact of changes in feature dimensions on the early warning performance of the model in the enterprise dataset will be further explored. Based on the principle of removing features in proportion, 83 groups of features in the above test data set are removed. If 25%, 50%, and 75% features in the test data set are removed in turn, 25%, 50%, and 75% test data sets are obtained respectively. The corresponding original test data set is abbreviated as 0% test data set, indicating that feature removal is not performed. Thereafter, four groups of test data sets of 0%, 25%, 50%, and 75% were used successively. The early warning performance of the above six groups of models is evaluated and shown in Table 2.

With the continuous reduction of feature dimensions in the test data set, the accuracy of different models also varies greatly. Among them, the change of feature dimension has the least impact on the test accuracy of literature [17] and literature [16], and the change range is no more than 2.5%. The change of feature dimension has the greatest impact on the accuracy of proposed model and literature [20]. With the decreasing of feature dimension, the accuracy of convolution neural network and long-term and short-term memory network decreases.

In order to further explore the impact of the change of feature dimension on the early warning performance of the model, this study adopts the method of reducing each group of test data sets by 5%. Then, a total of 18 test data sets can be obtained, namely 0%, 5%, 10%, ..., 80%, 85%. Among them, 0% of the test data sets still represent the features not removed and 85% of the test data sets represent the data sets after 85% of the features have been removed.

The accuracy of each model in the above 18 groups of test data sets is summarized, and a dotted line diagram of the accuracy and feature removal ratio is constructed (see Figures 7 and 8).

Further, the fourth-order polynomial is used to fit each point line in Figure 7 to obtain the fitting curve of accuracy and feature removal ratio (see Figure 8).

As can be seen from the overall performance of each fitting curve in Figure 8, the accuracy of proposed model and machine learning models such as literature [18] and literature [17] will fluctuate in a small range and tend to decline gently as the enterprise feature dimension decreases. The accuracy of literature [20] represented by deep learning will drop sharply without fluctuation. Therefore, the model and machine model in this study have stronger model stability and robustness than literature [20].

Importantly, each set of models corresponds to the feature optimal dimension. That is, compared with the performance of the model in the nonoptimal dimension, the

TABLE 1: The accuracy of each model.

Model	Training accuracy (%)	Test accuracy (%)			Average accuracy (%)
		Group 1	Group 2	Group 3	
Literature [18]	80.02	71.28	70.28	69.35	70.30
Literature [16]	87.02	77.44	78.94	80.14	78.84
Literature [19]	91.92	81.27	80.67	83.57	81.84
Literature [20]	89.21	79.95	81.02	81.45	80.81
Literature [17]	94.71	84.2	85.2	85.52	84.97
Proposed	96.59	86.38	86.64	87.29	86.77

TABLE 2: Test accuracy of each model under ratio removal characteristics.

Model	0% test accuracy	25% test accuracy	50% test accuracy	75% test accuracy	Range of accuracy change
Literature [18]	70.12	72.15	73.4	73.2	3.96
Literature [16]	78.66	79.08	78.37	76.91	2.85
Literature [19]	81.66	83.64	83.12	79.85	4.47
Literature [20]	80.63	78.04	72.74	71.72	9.59
Literature [17]	84.79	86.87	85.95	82.94	2.76
Proposed	86.59	87.96	83.99	80.95	7.69

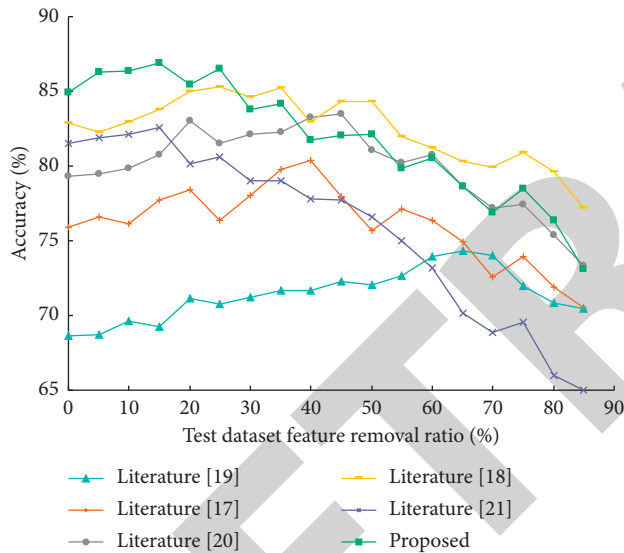


FIGURE 7: Point line relationship between test accuracy of each model and feature removal ratio.

early warning performance under the optimal dimension is often the best. Among them, the feature optimal dimension of the model in this study is about [63, 82] (corresponding to the x -axis [0, 23] in Figure 8). At this point, the test accuracy is as high as 87.26%, which is obviously better than other models. When the feature number removal ratio is greater than 23%, the early warning performance of this model gradually declines. In the interval [47, 63] (corresponding to the x -axis [23, 42] in Figure 8), the early warning performance of the model in this study is reversed by literature [17], that is [47, 63] becomes the feature optimal dimension of literature [17]. In other words, compared with other models, the model in this study has a larger feature optimal dimension (e.g., the test data set given in this study, the optimal dimension of agricultural industry financial features is [63, 82]). Then, it is concluded that the model in this study has the same performance for agricultural industry financial

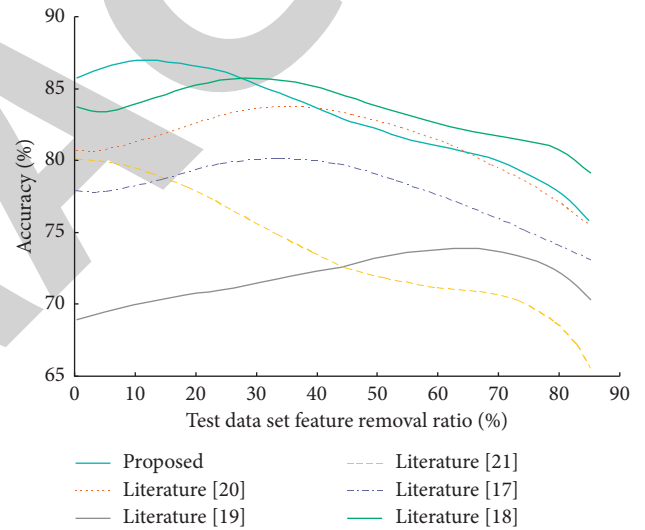


FIGURE 8: Fitting curve of test accuracy and feature removal ratio of each model.

early warning with low-dimensional features. But it has obvious advantages in high-dimensional features, and the accuracy of early warning is much higher than that of other models.

5. Conclusion

Agricultural industry finance is a weak link in China's financial system. In order to improve the risk warning ability of agricultural industry, this article proposes an agricultural industry financial risk warning system based on improved K-means clustering algorithm. In order to solve the problem that K-means algorithm cannot jump out of the local optimal solution in the clustering process, this study combines the grey Wolf optimization algorithm with K-means algorithm for clustering analysis. In the iterative process of grey Wolf optimization algorithm, the elite individuals in grey

Wolf population were cloned and mutated. In this way, elite individuals can be deeply mined, the depth exploration ability of grey Wolf optimization algorithm can be improved, and the local extremum of grey Wolf optimization algorithm can be avoided. At the same time, the algorithm is applied to cluster analysis and cosine distance similarity is used to calculate the similarity between data samples. The effectiveness of the proposed algorithm is verified by comparing the prediction results of different algorithms on agricultural financial risks. The next research direction is to systematically analyze the performance consumption of the algorithm and optimize the performance of the algorithm on the premise of ensuring the accuracy of financial risk prediction.

Data Availability

The labeled data set used to support the findings of this study is available from the corresponding author upon request.

Conflicts of Interest

The authors declare no competing interest.

Acknowledgments

This study was sponsored by Xi'an Kedagaoxin University.

References

- [1] C. Chen and J. Pan, "The effect of the health poverty alleviation project on financial risk protection for rural residents: evidence from Chishui City, China," *International Journal for Equity in Health*, vol. 18, no. 1, pp. 1–16, 2019.
- [2] V. Stehel, J. Horák, and M. Vochozka, "Prediction of institutional sector development and analysis of enterprises active in agriculture," *E+M Ekonomie a Management*, vol. 22, no. 4, pp. 103–118, 2019.
- [3] A. Deng and B. Yu, "Research overview of risk management of SMEs accounts receivable financing based on supply chain finance[J]," *World Journal of Research and Review*, vol. 4, no. 1, Article ID 262858, 2017.
- [4] B. Marchi, S. Zanoni, and M. Y. Jaber, "Improving supply chain profit through reverse factoring: a new multi-suppliers single-vendor joint economic lot size model," *International Journal of Financial Studies*, vol. 8, no. 2, p. 23, 2020.
- [5] A. Deng and X. Lu, "A literature review on the study of chattel pledge supervision in logistics and supply chain finance of China," *Journal of Financial Risk Management*, vol. 06, no. 02, pp. 93–100, 2017.
- [6] H. Hu, Y. Li, M. Tian, and X. Cai, "Evolutionary game of small and medium-sized enterprises' accounts-receivable pledge financing in the supply chain," *Systems*, vol. 10, no. 1, p. 21, 2022.
- [7] M. Abdel-Basset, M. Gunasekaran, M. Mohamed, and N. Chilamkurti, "RETRACTED: a framework for risk assessment, management and evaluation: economic tool for quantifying risks in supply chain," *Future Generation Computer Systems*, vol. 90, no. 1, pp. 489–502, 2019.
- [8] M. Tyagi, P. Kumar, and D. Kumar, "Assessment of CSR based supply chain performance system using an integrated fuzzy AHP-TOPSIS approach," *International Journal of Logistics Research and Applications*, vol. 21, no. 4, pp. 378–406, 2018.
- [9] Z. Bin, "Regional enterprise economic development dimensions based on k-means cluster analysis and nearest neighbor discriminant," *Journal of Intelligent and Fuzzy Systems*, vol. 38, no. 6, pp. 7365–7375, 2020.
- [10] Z. j Teng, J. l Lv, and L. w Guo, "An improved hybrid grey wolf optimization algorithm," *Soft Computing*, vol. 23, no. 15, pp. 6617–6631, 2019.
- [11] C. Y. Liu and Y. Wang, "Grey Wolf algorithm based on S-function and particle swarm optimization," *Journal of Physics: Conference Series*, vol. 1453, no. 1, Article ID 012021, 2020.
- [12] B. Yulianto, W. Budiharto, and I. H. Kartowisastro, "The performance of Boolean retrieval and vector space model in textual information retrieval," *CommIT (Communication and Information Technology) Journal*, vol. 11, no. 1, pp. 33–39, 2017.
- [13] N. Liu, F. Xie, and X. Wu, "Multi-pattern matching with variable-length wildcards using suffix tree," *Pattern Analysis & Applications*, vol. 21, no. 4, pp. 1151–1165, 2018.
- [14] C. Slamet, A. R. Atmadja, D. S. Maylawati, R. S. Lestari, W. Darmalaksana, and M. A. Ramdhani, "Automated text summarization for Indonesian article using vector space model," *IOP Conference Series: Materials Science and Engineering*, vol. 288, no. 1, Article ID 012037, 2018.
- [15] J. Krebs, H. Delingette, B. Mailhé, N. Ayache, and T. Mansi, "Learning a probabilistic model for diffeomorphic registration," *IEEE Transactions on Medical Imaging*, vol. 38, no. 9, pp. 2165–2176, 2019.
- [16] Y. Zizi, A. Jamali-Alaoui, B. El Goumi, M. Oudgou, and A. El Moudden, "An optimal model of financial distress prediction: a comparative study between neural networks and logistic regression," *Risks*, vol. 9, no. 11, p. 200, 2021.
- [17] G. Teles, J. J. P. C. Rodrigues, R. A. L. Rabêlo, and S. A. Kozlov, "Comparative study of support vector machines and random forests machine learning algorithms on credit operation," *Software: Practice and Experience*, vol. 51, no. 12, pp. 2492–2500, 2021.
- [18] X. Ma and S. Lv, "Financial credit risk prediction in internet finance driven by machine learning," *Neural Computing & Applications*, vol. 31, no. 12, pp. 8359–8367, 2019.
- [19] U. Grömping, "Variable importance assessment in regression: linear regression versus random forest," *The American Statistician*, vol. 63, no. 4, pp. 308–319, 2009.
- [20] R. Becerra-Vicario, D. Alaminos, E. Aranda, and M. A. Fernandez-Gamez, "Deep recurrent convolutional neural network for bankruptcy prediction: a case of the restaurant industry," *Sustainability*, vol. 12, no. 12, p. 5180, 2020.