Modern Economy, 2022, 13, 860-884
https://www.scirp.org/journal/me
ISSN Online: 2152-7261
ISSN Print: 2152-7245

Agricultural Credit Risk Assessment in China Based on the BP and GA-BP Neural Network

Fubing Sun

Zhejiang University of Finance & Economics Dongfang College, Jiaxing, China
Email: sfb007@zufe.edu.cn, 29231936@qq.com

How to cite this paper: Sun, F. B. (2022). Agricultural Credit Risk Assessment in China Based on the BP and GA-BP Neural Network. *Modern Economy, 13,* 860-884.
https://doi.org/10.4236/me.2022.136047

Received: April 16, 2022
Accepted: June 26, 2022
Published: June 29, 2022

Abstract

The credit constraint caused by difficult and expensive loans is a crucial obstacle to agricultural modernization in China. This is due to the high risk and uncertainty of agricultural production and operation activities and the high transaction cost and asymmetric information of agricultural credit activities, which lead to ineffective risk assessment. In this study, comprehensive information on agricultural credit business reports, customer questionnaires, and loan application forms of Chinese banks are combined with the characteristics of the agricultural industry and credit scenarios to develop an innovative agricultural credit risk assessment index system. The index system is constructed mainly based on the first repayment source and risk process. Further, a genetic algorithm optimizes the BP neural network. The sample data of 1165 agricultural credits collected from Zhejiang, Jiangsu, Shandong, and Henan provinces are analyzed. The results of the classification prediction simulation show that this method effectively reduces the problem of the BP neural network converging to a local minimum and increases the accuracy and sensitivity correction of data prediction. This overcomes the problem of difficult risk assessment due to nonstandard and inaccurate agricultural credit data, thus providing theoretical and practical solutions for improving the efficiency of agricultural credit risk assessment and control.

Keywords

Agricultural Credit, Risk Assessment, Genetic Algorithm, BP Neural Network

1. Introduction

The African swine fever and Sino-US trade war caused tight supply and continuous price increases of Chinese agricultural products, such as meat, poultry, and feed, for a sustained period. Thereafter, the COVID-19 and locust plague in Africa broke out in 2020, triggering volatility and panic in the international food market. Consequently, the United Nations World Food Program and FAO jointly issued the "Early Warning of Hot Spots with Serious Food Insecurity," according to which, "at least 25 countries will face a severe risk of famine in 2020, and the world is on the verge of the worst food crisis in 50 years". Thus, issues concerning agriculture, rural areas, and food security have recently become a topic of great interest to the Chinese people.

The structure of the article is as follows: the first part is a brief introduction to the full text, the second part is the system and data collection of the new agricultural business entity, the third part uses the establishment of the neural network model, the fourth and fifth parts are to substitute the data into the model for calculation, and the seventh part is a summary.

Meanwhile, the investment of a unit of land capital and the intensity of agricultural capital are constantly increasing. Therefore, it is necessary to establish a rural financial support system and adequate rural financial supply to ensure sustainable development and establish a modern management system for the agriculture sector. However, according to the statistical data provided by the "China Rural Financial Services Report", the balance of domestic and foreign currency rural loans of Chinese financial institutions has been around 20% of the balance of various loans, and the country's coverage rate of rural banks is only about 50%. Therefore, many rural areas lack basic financial services, and agricultural production faces severe credit constraints.

The reason for China's agricultural credit constraints lies in the high degree of fragmentation and scale of agricultural production land, limited ability to resist natural disasters, weak rural credit infrastructure, lack of adequate information on credit transactions, and high transaction costs. These reasons cause ineffective risk identification and assessment in rural financial institutions. Meanwhile, agricultural business entities have problems, such as low income, limited assets, ineffective mortgages, and inadequate guarantees. In addition, poor credit records, irregular land circulation, and insufficient insurance guarantee constrain businesses. Furthermore, there is concern that rural financial institutions lack innovation and effective methods to prevent and control risks.

Therefore, to address this dilemma, we explore the effective index system and model method of risk assessment. Due to information asymmetry, we also rely on financial science and technology innovation to help China's rural financial institutions avoid adverse selection and moral hazards after trading. Financial institutions reduce transaction costs through the scale effect, providing technical support for improving the efficiency of credit risk management and relieving credit constraints.

2. Literature Review

The existing theory and practice of risk assessment of agricultural credit business, especially the vast rural financial institutions in China, remain based on mortgage, guarantees, and other secondary repayment sources. On the other hand, they share a system with traditional credit businesses and do not develop a unique index system for risk assessment of agriculture-related credit. This will inevitably lead to significant defects in the pertinence, scientificity, and accuracy of the risk assessment. For example, in constructing a risk assessment index system, Micha et al. (2015) comprehensively evaluated the impact of obtaining convenient financing from indicators, such as the age of the agricultural operators, education level, industry prosperity, planting area, and government support. Sousa (2015) used a dynamic Bayesian model to construct a discriminant model, focusing on indicators, such as age, occupation, income, and social status of agricultural operators in the risk assessment. Meng and Chi et al. (2015) combined partial correlation analysis and comprehensive discriminant ability to build a credit evaluation index system of 16 indicators for farmers' microfinance. They used support vector machine (SVM) to build a credit evaluation model. Based on farmers' basic information and credit details, a rural credit evaluation system with 2 first-level and 11 second-level indicators was constructed (2019). In applying the risk assessment model, Dutta and Rating (1988) analyzed how the change in network structure and the number of independent variables in the neural network model affect its credit-rating-discrimination-ability, at a prediction accuracy rate of 76% - 82%. Sousa (2015) built a Bayesian dynamic credit risk assessment model based on ordered data, and the empirical results are consistently superior to the static one. Meng and Chi et al. (2015) built a credit risk evaluation model for farmers' microfinance by maximizing the range. The results showed that the evaluation results of different single methods are contradictory. Huang et al. (2019) combined the traditional scorecard with the machine learning model and concluded that the effectiveness was better than a single XGBoost credit scoring model.

The existing research on agricultural credit risk assessment is mainly based on the secondary repayment sources of customers, such as basic credit, financial status, mortgage, and guarantee. These traditional credit risk assessment technical tools are mainly re-applied in agricultural credit risk assessment. However, several Chinese agricultural business entities, especially those applying for operating loans of over 300,000 - 500,000 yuan, find it difficult to pass the risk assessment due to the absence of mortgage guarantees. They are often subject to existing policies and credit constraints. Therefore, this paper begins with the design of the first repayment source, mainly based on the income from production and operation. Then, we combine the machine learning methods with high accuracies, such as BP and GA-BP neural networks, to make the credit risk assessment more targeted, scientific, and accurate.

3. Improvement of the Agricultural Credit Risk Assessment Index System and Collection of Sample Data

3.1. Improvement of the Risk Assessment Index System

First, we review numerous documents and divide agricultural credit risk into credit, operational, and market risk, based on national risk classification standards. Agricultural credit risk has the characteristics of general financial risk, the particularity of the agricultural industry, periodicity, complexity, infectivity, and relevance. Second, considering that agricultural credit activities run through the whole process before, during, and after production, the main risks and their influencing factors are different in different industries and links. These three types of risks are further subdivided into 10 secondary indicators that include solvency, technical risk, and price risk. Then, the basic questionnaire is designed by collecting the information of agriculture-related credit business reports, customer questionnaires, and loan application forms of various banks. The questionnaire is based on analyzing agricultural credit scenarios from the perspective of risk sources and processes. It combines them with many default behaviors and the three-level indicators influencing agricultural credit risk. Finally, combined with field investigation and interviews, the questionnaire is constantly revised and improved by the Delphi method to make the design more scientific, rigorous, and operable. Therefore, relying on the existing bank credit risk evaluation index system, a large number of index parameters based on the first repayment source are further defined so that the overall index is more scientific and precise. Finally, an innovative three-level evaluation index system of agricultural credit risk is designed (**Table 1**).

3.2. Data Collection and Inspection

To obtain real, effective, and representative credit sample data, an empirical investigation was conducted in Zhejiang, Jiangsu, Shandong, and Henan in China, some representative regions with the most significant scale of agricultural credit supply and relatively active rural financial reform in 2018-2019. Since these provinces are the first provinces in China to start the agricultural financial reform and have special agricultural financial products, to a certain extent, they can represent the future direction of China's domestic agricultural reform and future popularization policies, and the data of these provinces can effectively represent the overall universality. Local government agricultural authorities, agriculture-related financial institutions, industry associations, and agricultural business entities[1] were interviewed, such as civil servants, farmers, loan officers, factory directors, local farmers, respectively. In particular, the representative counties and cities under the jurisdiction of China's rural financial reform pilot zone—Lishui City, Zhejiang Province—were investigated. Meanwhile, in-depth and detailed interviews were conducted with the executives and frontline credit personnel of 105 agriculture-related credit banks. As a result, the analysis of the factors influencing credit risk was cross-verified and improved with actual credit business operation experience. Simultaneously, nearly 2000 credit sample data were obtained by combining field investigation with bank internal data collection. The resulted from the phenomenon that Chinese agricultural operators

[1]Due to the wide scope of agricultural credit business, the credit samples investigated in this paper are mainly production-oriented agricultural business subjects.

Table 1. Innovative agricultural credit risk assessment index system.

Level 1 Indexes	Level 2 Indexes	Level 3 Indexes
Credit Risk	Basic situation	Gender (male = 0, female = 1), age, marital status (single = 0, married = 1, divorce = 2), education (middle school and under = 1, high school and technical secondary school = 2, college and above = 3), health (health = 0, sub-health = 1), have management experience (yes = 0, no = 1), has enterprises and institutions work experience (yes = 0, no = 1), years of farm work, multiple occupations/professional (full-time = 0, part-time = 1);
	Debt paying ability	Total household assets, outstanding bank loans, other loan amounts;
	The risk of moral hazard	Whether there is a record of default (yes = 0, no = 1); Is there any introduction of acquaintances in the loan process (yes = 0, no = 1); Local credit environment
Operational Risk	Natural risk	Whether to buy agricultural insurance (yes = 0, no = 1);
	Financial risk	Total asset-liability ratio, sales profit margin, investment in agricultural production facilities, whether there is an external guarantee (yes = 0, no = 1), whether there is an external investment (yes = 0, no = 1);
	Technical risk	Whether there are professional technical personnel (yes = 0, no = 1), and whether there is mechanical automation equipment (yes = 0, no = 1);
	Production management risk	Production scale, with or without simple electronic management (with = 0, without = 1);
	Policy risk	Whether there are policy risks such as environmental protection (yes = 0, no = 1), land transfer life;
Market Risk	Price risk	The understanding of the market price fluctuation (follow the trend = 0, understand = 1, be sure = 2), whether the product is a local excellent specialty (yes = 0, no = 1), whether the product is a registered trademark (yes = 0, no = 1), whether the product is certified (yes = 0, no = 1);
	Supply and demand risk	Whether there are long-term and stable purchase channels (there is = 0, no = 1), whether there are long-term stable sales channels (there is = 0, no = 1).

Data source: calculated and sorted by the author.

generally have low education, irregular management, and non-standardized agricultural credit bank business data and file management system. After further screening, sorting, analysis, and consulting relevant credit personnel, 1165 valid credit sample data were finally obtained.

To further verify the scientificity of the selected evaluation indicators, we construct a basic fixed-effect model for initial pre-inspection and consult Chinese universities and banking experts to validate the constructed evaluation indicators. Simultaneously, it was found in the survey that the credit risks of agricultural business entities in different industries vary. Therefore, the subsequent empirical analysis divides the obtained credit sample data into four categories according to the different industries: cash crops, field crops, animal husbandry, and aquaculture[2].

[2]Due to the relatively small number of aquaculture samples and uneven regional distribution found in the process of empirical investigation, this study first explored the other three types of agricultural operators. Meanwhile, for the convenience of expression, the agricultural operators of the industry were referred to as the agricultural operators of the industry in the subsequent sample analysis.

4. Design of the BP Neural Network and the GA-BP Neural Network Model Structure

In addition to setting up a scientific and reasonable evaluation index system, choosing suitable and efficient models and tools plays a vital role in the accuracy of the evaluation output. In recent years, with the rapid development of artificial intelligence [AI] and information technology [IT], the ideas and methods of credit risk assessment have become more diversified. As a result, advanced technologies, such as neural networks, genetic algorithms, and decision support systems, have been gradually introduced into the field of risk assessment, effectively overcoming the limitation of traditional econometric analysis methods with strict data requirements. LeCun, Bengio, and Hinton (2015) first proposed deep learning discovers intricate structure in large data sets by using the backpropagation algorithm to indicate how a machine should change its internal parameters that are used to compute the representation in each layer from the representation in the previous layer.

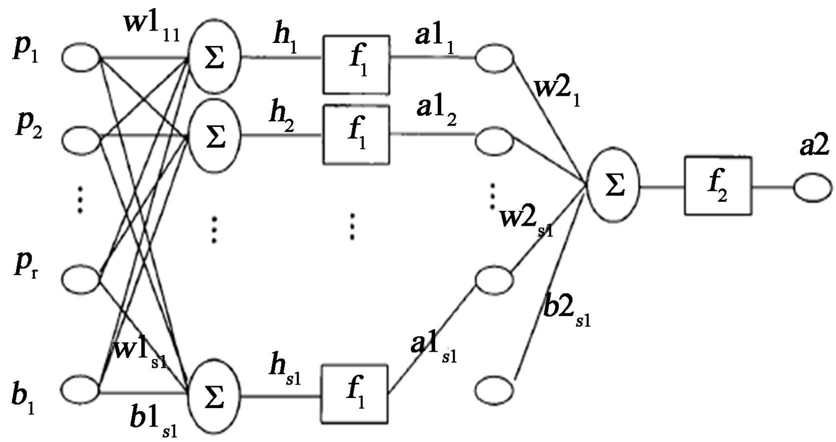
BP neural network is a unidirectional propagation multilayer forward network based on the BP algorithm; it includes the input, hidden, and output layers. It uses the gradient descent method and gradient search technology to minimize the mean square error between the actual and expected output of the network. The calculation process of the BP neural network consists of forward and reverse calculation processes. In the forward calculation process, the foreign information goes through the hidden layer from the input layer, and then goes back to the output layer such that the state of neurons in each layer can only affect the state of neurons in the next layer, but the neurons in each layer do not affect each other (2016). If the calculation process cannot obtain the expected output result at the output layer, it will turn to backpropagation. This returns the error signal layer by layer along the original connection path, makes repeated calculations by modifying the weights of each neuron, and finally minimizes the error signal. Therefore, the BP neural network has unique advantages in the logical and classification judgment of samples. The black box network can effectively and quickly self-learn and achieve the dynamic fitting effect of judgment function. Simultaneously, it can also fit all types of irregular functions. Thus, theoretically, the BP neural network can fit all functions.

Determination of the BP Neural Network

The input layer of the neural network is the initial parameter and the middle layer selects the node number according to $\log_2 A$. A is the number of input layer nodes, as the output layer is the target result, and its structure is shown in **Figure 1**.

The elimination and the formula method are used to determine the number of neurons in the middle layer. The formula method is widely used, and the specific reference formulas include the following three:

First, calculate $n_1 = \sqrt{m+n} + a$, where m and n are the number of neurons in the output layer and a is the input layer. Then $n_1 = \log_2 n$, where n is the



Data source: Machine learning [M], Tsinghua University Press, 2016.

Figure 1. Structure diagram of neural network.

number of neurons in the input layer; Final calculate $\sum_{i=0}^{n} C_n^i > k$, where k is the sample number, n_i is the number of intermediate nerve elements, n is the number of afferent nerve elements, where $i > n_i$, $C_n^i = 0$.

There are multiple internal calculation algorithms in a neural network, and the basic algorithm cannot meet the agricultural credit risk assessment due to low sample data quality. Therefore, all the basic default algorithms are changed and optimized to adapt to the calculation processing. The specific functions involved in the construction and calculation process are as follows:

LM algorithm is similar to the quasi-Newton method, and its purpose is to avoid calculating the Hessian matrix when modified at an approximate second-order training rate. When the error performance function has the form of square sum error (typical error function of training feedforward network), the Hessian matrix can be approximately expressed as

$$H = J^{\mathrm{T}} J \tag{4}$$

Simultaneously, the expression of the gradient is

$$g = J^{\mathrm{T}} e \tag{5}$$

where J refers to the Jacobian matrix containing the network error function, the first derivative of the weights and thresholds, and e is the network error vector. The Jacobian matrix can be calculated by standard feedforward network technology, which is straightforward than the Hessian matrix. Similar to Newton's method, the LM algorithm can be modified by the approximate Hessian matrix as follows:

$$x(k+1) = x(k) - \left[J^{\mathrm{T}} J + \mu I \right]^{-1} J^{\mathrm{T}} e . \tag{6}$$

When the coefficient μ is 0, the above formula is the Newton method. When the value of coefficient μ is large, the above formula becomes the gradient descent method with a smaller step size. Because Newton's method approaches the minimum error faster and more accurately, the algorithm should be as close as possible to Newton's method. With each successful iteration step (the error performance is reduced), the μ is reduced. Only when the error performance increases after the trial iteration will μ increase. In this manner, the error performance of the algorithm is always the smallest after each iteration. LM algorithm is the fastest algorithm (up to hundreds of connection weights) proposed mainly for training medium-sized feedforward neural networks. It is also effective for MATLAB implementation because matrix calculation is realized by functions in MATLAB. The attributes become clear when setting.

Improvement of the Genetic Algorithm

It is easy to fall into the local extremum with the BP neural network, leading to over-fitting and training failure. To better realize its popularization and application value, this study mainly avoids its inherent application defects by adjusting and modifying the relevant parameters of the model and combining the genetic algorithm strategy with the neural network. The study also puts forward a strategy for strengthening network training using a genetic algorithm and compares the advantages and disadvantages of the two. As a result, the improved model is more operable and accurate in an application, and finally, it improves the risk assessment efficiency by giving full play to their advantages.

Introducing a genetic algorithm into the BP neural network and its highly parallel global search algorithm overcomes its shortcomings. It helps avoid the problem of falling into a local minimum and improves the convergence speed of the network. This enhances the learning and generalization ability of the model and its performance. In determining the model structure, the input and output parameters are mainly determined according to the problems to be solved. Contrastingly, the genetic algorithm mainly determines the initial weight and offset value by calculating the individual fitness value and the weight and offset value corresponding to the optimal individual through selection, crossover, and mutation operations. It is only necessary to assign the values of this parameter individual to the weights and thresholds of the BP neural network in the operation process because each parameter individual contains all the weights and thresholds of the neural network. Meanwhile, the BP neural network prediction mainly determines the final structure of the network according to the first two steps to predict the new data and get the results.

Fitness function: Through the initial weight and threshold of BP neural network obtained by the individual, the training data is used to train the BP neural network to predict the system output. At the same time, the absolute value E of the error between the predicted output and the expected output and F as the individual fitness value are calculated as follows:

$$F = k\left(\sum_{i=1}^{n} abs\left(y_i - o_i\right)\right) \tag{7}$$

In the above formula, n is the number of nodes output by the network, and y_i is the expected output value of the ith node of BP neural network, o_i is the predicted output value of the ith node, and k is the coefficient.

Select operation: The selection operation of genetic algorithm belongs to roulette method. It is a selection strategy based on fitness proportion, and the selection probability of each individual *i* is different p_i as follow:

$$f_i = \frac{k}{F_i} \tag{8}$$

$$p_i = \frac{f_i}{\sum_{j=1}^{N} f_i} \tag{9}$$

In the above formula, F_i is the fitness value of individual *I*, because the smaller the fitness value, the better. Therefore, before individual selection, calculate the reciprocal of the fitness value, where *k* is the coefficient and *N* is the number of individuals in the population.

Cross operation: Because individuals mainly use real number coding, the crossover operation method adopts real number crossover method, and the *K* chromosome a_k and chromosome *l*th a_l, their cross operation method at *J* is as follows:

$$a_{kj} = a_{kj}\left(1-b\right) + a_{lj} b \tag{10}$$

$$a_{lj} = a_{lj}\left(1-b\right) + a_{kj} b \tag{11}$$

In the above formula, *b* is a random number between [0, 1];

Cross operation can effectively increase the randomness of data and reduce the specified partial derivative of single regular data to the result.

Mutation operation: Select the *j*-th gene a_{lj} in the *i*-th individual for mutation operation:

$$a_{ij} = \begin{cases} a_{ij} + \left(a_{ij} - a_{\max}\right) * f\left(g\right) r > 0.5 \\ a_{ij} + \left(a_{\min} - a_{ij}\right) * f\left(g\right) r \le 0.5 \end{cases} \tag{12}$$

where $a_{\max}$ is the upper bound of the gene and $a_{\min}$ is the lower bound of the gene; $f\left(g\right) = r_2\left(1 - \frac{g}{G_{\max}}\right)^2$, r_2 is a random number, *G* is the current iteration number, $G_{\max}$ is the maximum evolution number, and *R* is the random number between [0, 1].

BP neural network optimization: After completing a series of operations, the weight and threshold of the optimal solution are substituted into the BP neural network for cyclic iterative calculation, and a better solution is obtained.

5. Demonstration and Simulation

The BP neural network easily falls into over-fitting during credit risk assessment. Thus, the genetic algorithm fits the whole data better and has a higher degree of discrimination against unknown data in forecasting. Therefore, to further explore a more accurate and efficient model and method for credit risk assessment, the demonstration and simulation will begin with the fitness function based on BP neural network. Then, the above 33 index parameters are used to conduct the selection, crossover, and mutation operations on the credit samples of three types of agricultural business entities engaged in cash crops, field crops, and livestock breeding production. Then, the genetic algorithm is substituted for analysis to compare the advantages and disadvantages of the two models regarding risk assessment accuracy.

MATLAB R2019a is used to build BP neural network, and 1165 sample data are classified according to industry. The specific process includes determining training samples, selecting network parameters and related functions, analyzing training results, and warning network risks.

5.1. Neural Network Construction

Taking the above 33 risk factors as the input layer of network training, whether to default as the output layer, that is, there are 33 input layer nodes and 1 output layer node, the number of hidden layers is $\log_2 33$, and after rounding, it is 5.

5.2. Neural Network Training

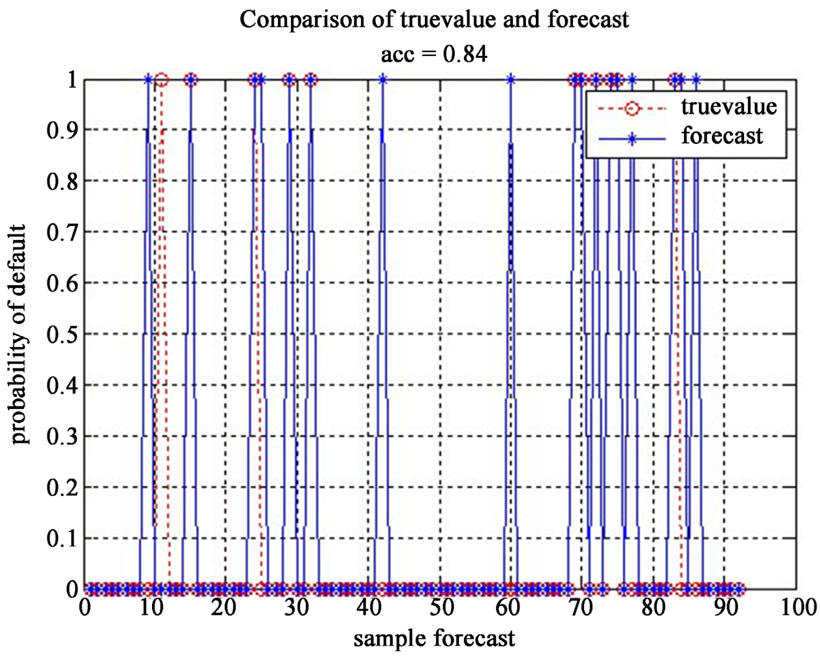
After the neural network structure and training parameters are set, the accuracy of the training set is explained by the confusion matrix. The sum of data in the 2 × 2 matrix represents the total training sample data. The percentage represents the accuracy. Finally, the accuracy of the test simulation set is measured by prediction accuracy, which is equal to the real value of the total number of test sets. According to different industries, the classification simulation test is carried out, and the specific results are discussed in the following sections.

6. BP and GA-BP NEURAL Network Analysis of Cash Crop Samples

The credit risk of cash crop samples is analyzed, including the above 33 parameters, with a total of 458 sample data. A total of 366 sample data are selected as training sets, and the remaining 92 are selected as test simulation sets. According to the model results, the probability of being judged correctly is 97.7% among the 366 training sample data. Furthermore, the prediction accuracy of the BP neural network model for cash crop samples is 73% (**Figure 2**), while the prediction accuracy of the GA-BP neural network model is 84% (**Figure 3**), which is better than that of BP neural network.

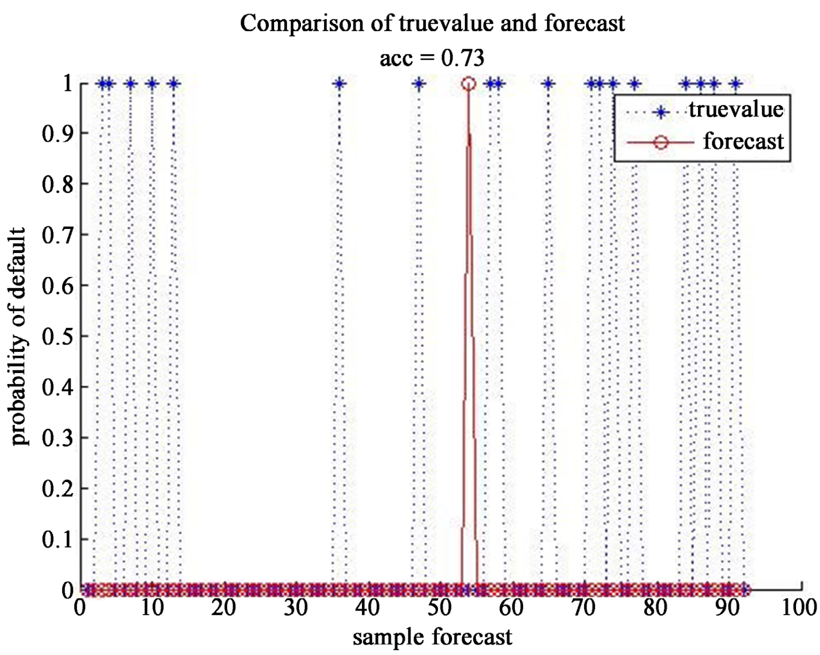
In the calculation process of the genetic algorithm, the optimization of the BP neural network is consistent. The training dataset is divided into three groups to prevent over-fitting. Training is 75% of the total data, validation is 15%, and the test is the remaining 10%. Only the training data participate in the training, and the other data are used for later inspection and explanation. As shown in **Figure 4**, the training set r of the BP neural network is 0.980, the verification set r is 0.854, and the test simulation set r is 0.9415. Thus, the fitting effect is relatively good, and the overall fitting degree is 0.953. In **Figure 5**, the training set r of the GA-BP neural network is 0.897, the verification set r is 0.895, and the test simulation set r is 0.876. The fitting effect is good with an overall fitting degree of 0.894 and no fitting phenomenon.

The variance error of the BP neural network for cash crop samples is concentrated between −0.05 and 0.06, and the error rate is negligible (Figure 6). The variance error of the GA-BP neural network for cash crop samples is concentrated



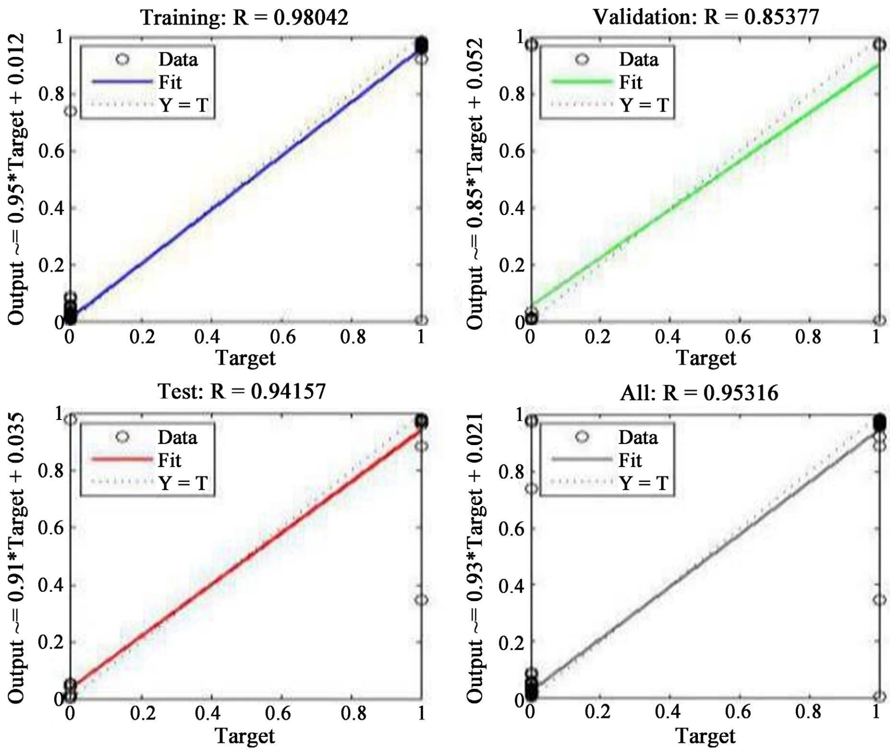
Data source: calculated and sorted by the author.

Figure 2. Prediction accuracy of cash crop samples by the BP neural network.



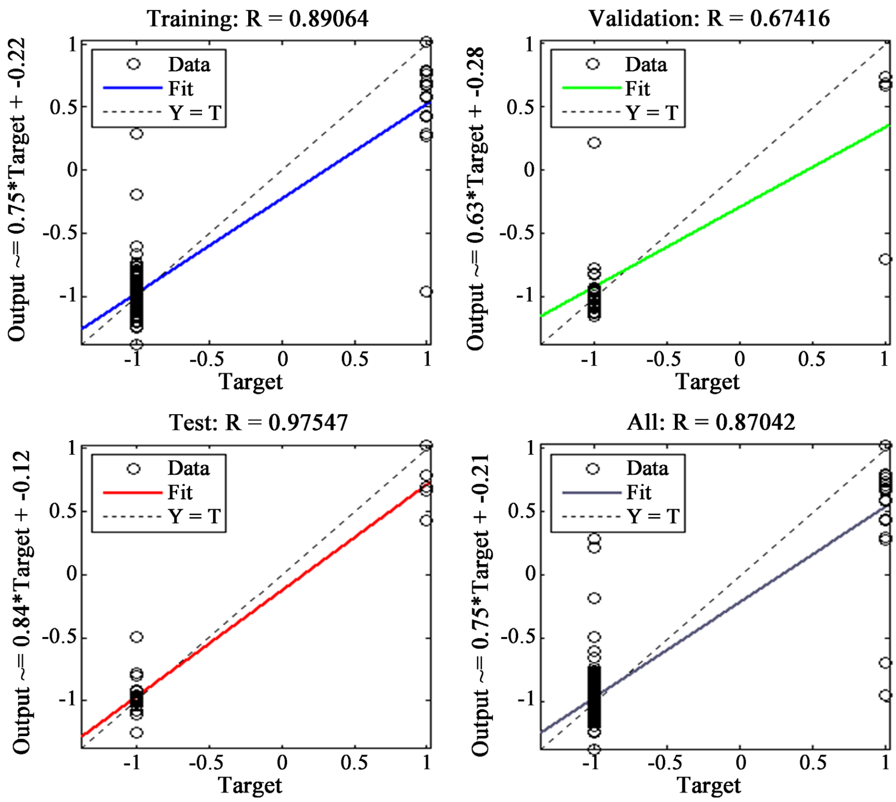
Data source: calculated and sorted by the author.

Figure 3. Prediction accuracy of cash crop samples by the GA-BP neural network.



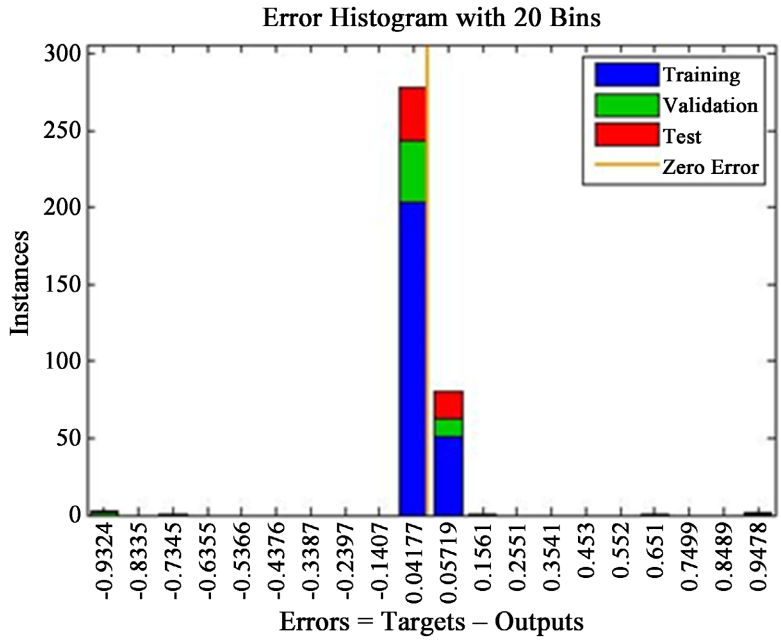
Data source: calculated and sorted by the author.

Figure 4. BP neural network R of cash crop samples.



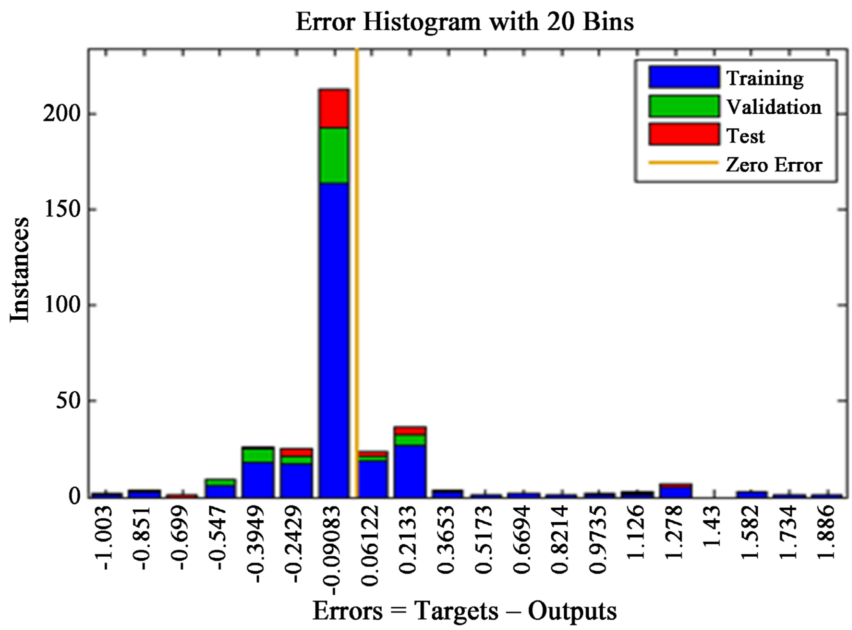
Data source: calculated and sorted by the author.

Figure 5. GA-BP neural network R of cash crop samples.



Data source: calculated and sorted by the author.

Figure 6. BP neural network error histogram of cash crop samples.

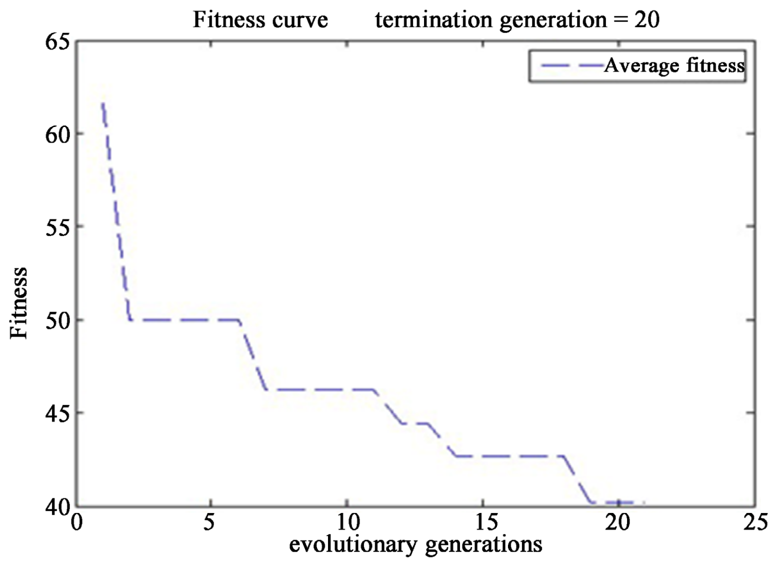


Data source: calculated and sorted by the author.

Figure 7. GA-BP neural network error histogram of cash crop samples.

between −0.09083 and −0.06122, approaching 0, and the error rate is negligible (Figure 7). Thus, there is little difference between the two methods in operation error, but the genetic algorithm is more accurate in judging extreme values.

In the optimization process of genetic algorithm, the number of evolutions, that is, according to the principle, the lower the function value of fitness, the higher the fitness, and the better the individual (Figure 8). When 20 iterations



Data source: calculated and sorted by the author.

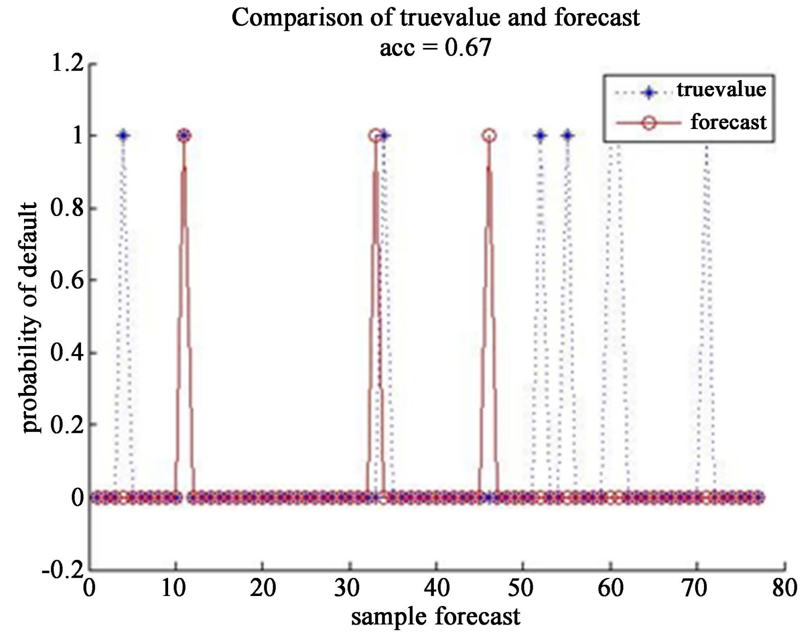
Figure 8. 20-generation adaptive diagram of the GA-BP neural network for cash crop samples.

are reached, the individual fitness reaches the highest, and the optimal weights and thresholds are obtained simultaneously. As the number of iterations and the population size will affect the calculation time and performance, one size is selected as 50. Choosing 20 and 50 can effectively optimize the intermediate operation time when choosing code exchange. The probability of cross genes is selected as 0.3, and the probability of code variation is selected as 0.1. Because RNA cross probability is slightly larger than self-variation in biology and in the process of calculation optimization and performance comparison, according to the principle, the lower the function value of fitness, the higher the fitness, and the better the individual. When 20 iterations are reached, the individual fitness reaches the highest, and the optimal weights and thresholds are obtained simultaneously.

BP and GA-BP Neural Network Analysis of Field Crop Samples

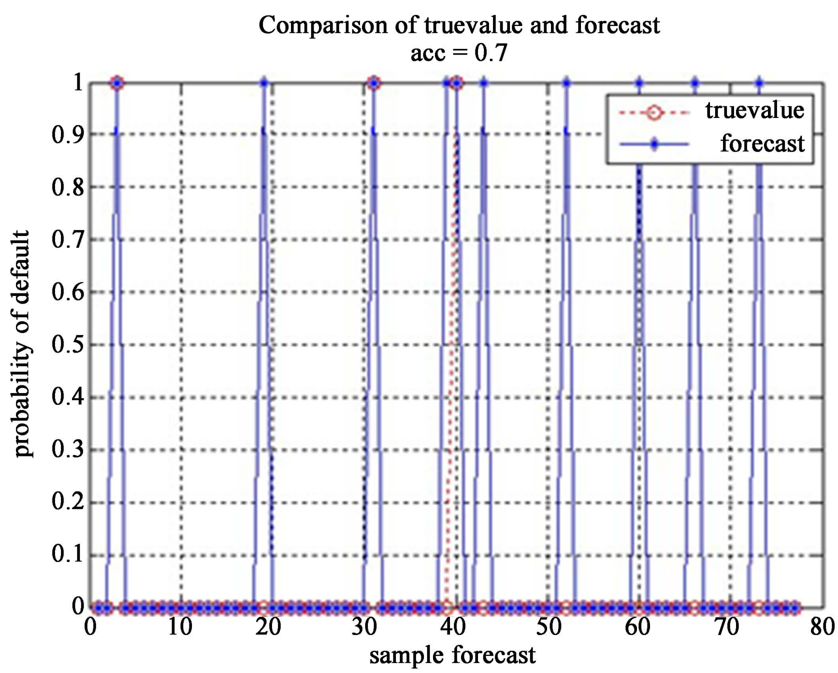
The credit risk of field crop samples is also analyzed, including the above 33 parameters, a total of 387 sample data. Among the sample data, 310 are selected as training sets, and the remaining 77 are selected as test simulation sets. According to the model results, the probability of being judged correctly is 92.9% among 310 training sample data. **Figure 9** shows that the prediction accuracy of the BP neural network model of field crops with 33 parameters is 67%. The prediction accuracy of the GA-BP neural network model is 70%, and the prediction effect is more pronounced (**Figure 10**).

The training dataset is further divided into three groups to prevent over-fitting. Training accounts for 75%, validation 15%, and test accounts for the remaining 10% of the total data. Only the training data participate in the training, and the other data are used for later inspection and explanation. The training set r of the BP neural network is 0.961, the verification set r is 0.995, and the test simulation



Data source: calculated and sorted by the author.

Figure 9. Prediction accuracy of field crop samples by the BP neural network.

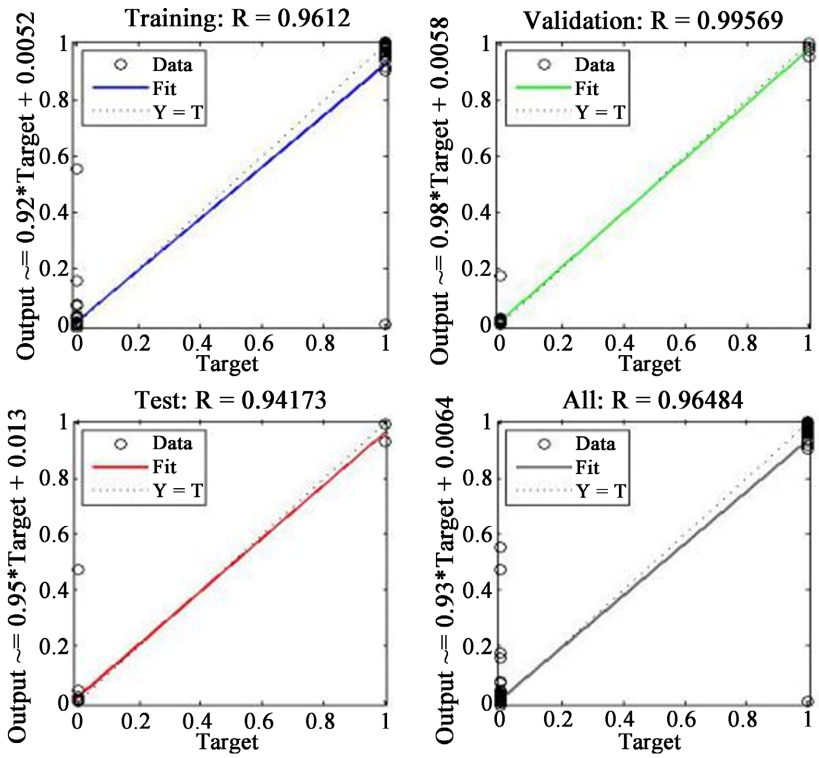


Data source: calculated and sorted by the author.

Figure 10. Prediction accuracy of field crop samples by the GA-BP neural network.

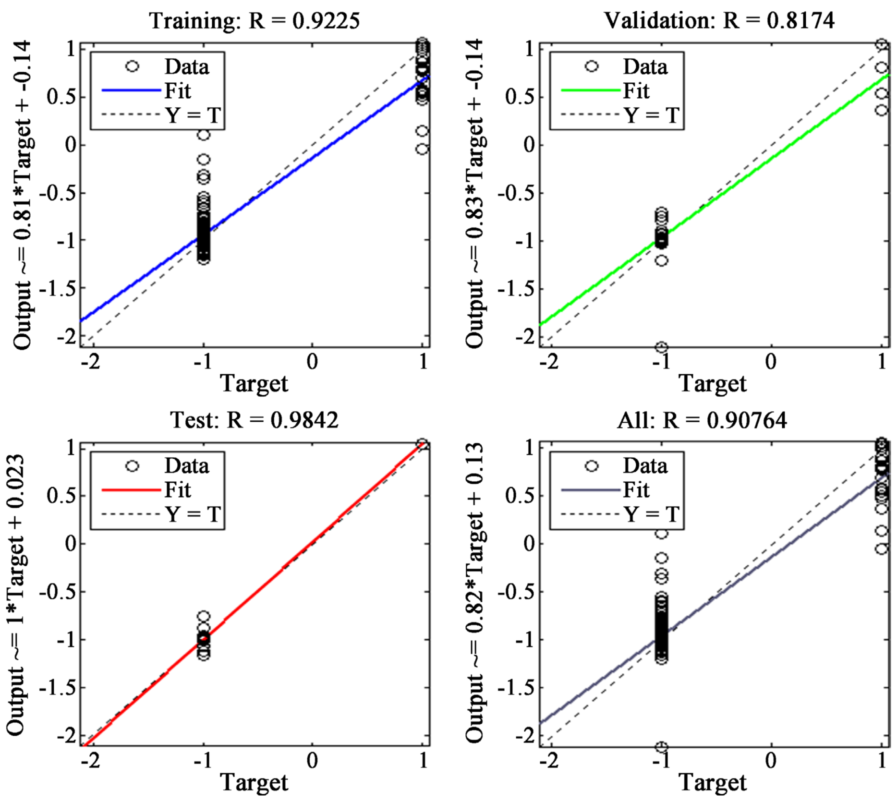
set r is 0.942 (**Figure 11**). Thus, the fitting effect is good, and the overall fitting degree is 0.965. **Figure 12** shows that the training set r of the GA-BP neural network is 0.923, the verification set r is 0.817, and the test simulation set r is 0.984. Again, the fitting effect is good, and the overall fitting degree is 0.908.

The variance error of the BP neural network prediction model is concentrated



Data source: calculated and sorted by the author.

Figure 11. BP neural network R of field crop samples.

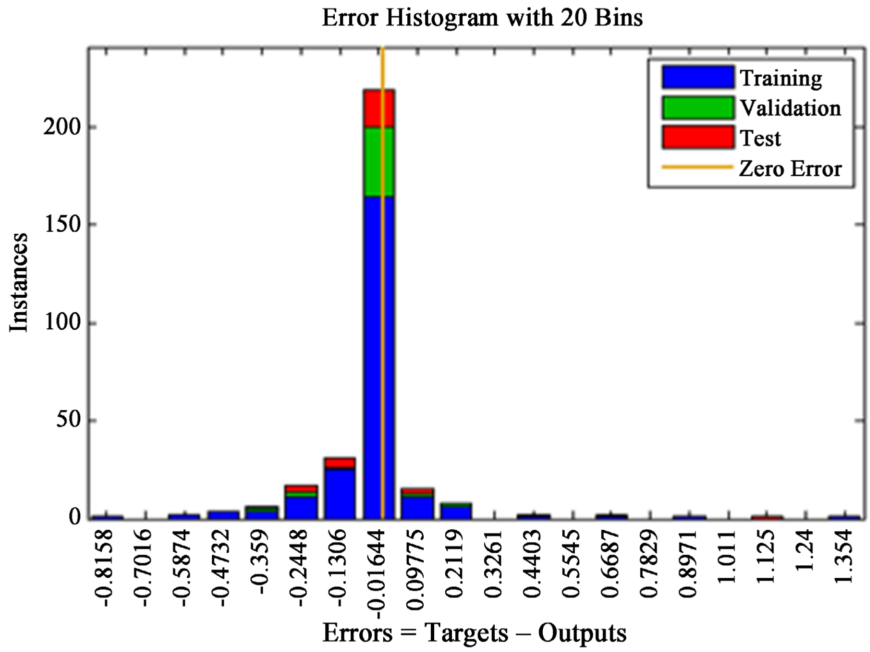


Data source: calculated and sorted by the author.

Figure 12. GA-BP neural network R of field crop samples.

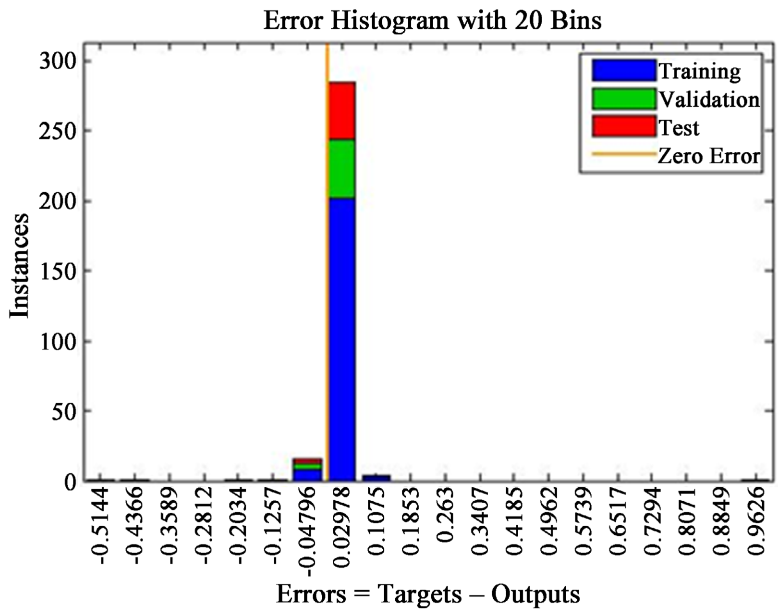
between 0 and 0.03, and the error rate is small and concentrated near 0 (Figure 13). Figure 14 shows that the variance error of the GA-BP neural network prediction model is concentrated between −0.017 and 0, and the error is less than that of the BP neural network.

The evolution times (i.e., iteration times) are selected as 30, and the population size is selected as 50 in the optimization process of the genetic algorithm



Data source: calculated and sorted by the author.

Figure 13. BP neural network error histogram of field crop samples.



Data source: calculated and sorted by the author.

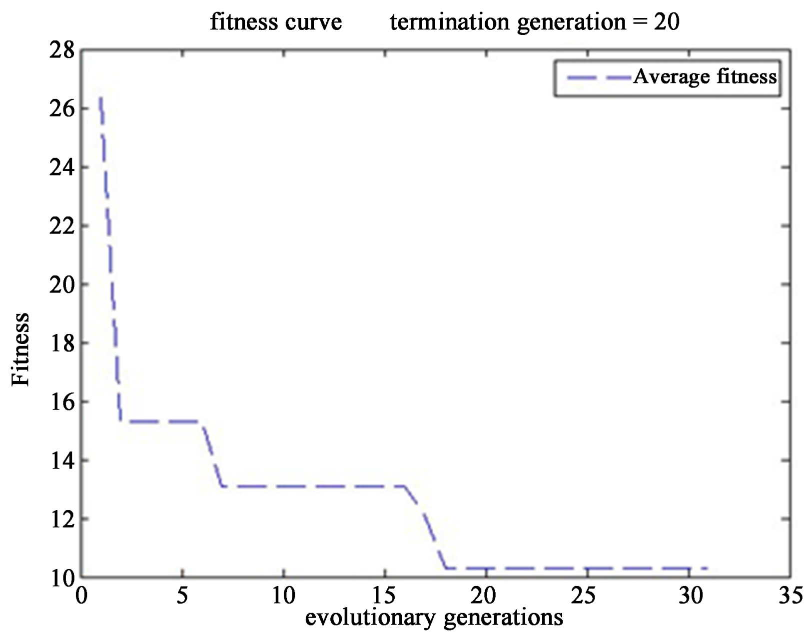
Figure 14. GA-BP neural network error histogram of field crop samples.

(Figure 15). As iteration times and population size will affect the calculation time and performance, choosing 20 and 50 can effectively optimize the intermediate operation time. When 20 iterations are reached, the individual fitness reaches the highest, and the optimal weights and thresholds are obtained simultaneously, consistent with the previous generations of the genetic algorithm for cash crops.

7. BP and GA-BP Neural Network Analysis of Animal Husbandry Samples

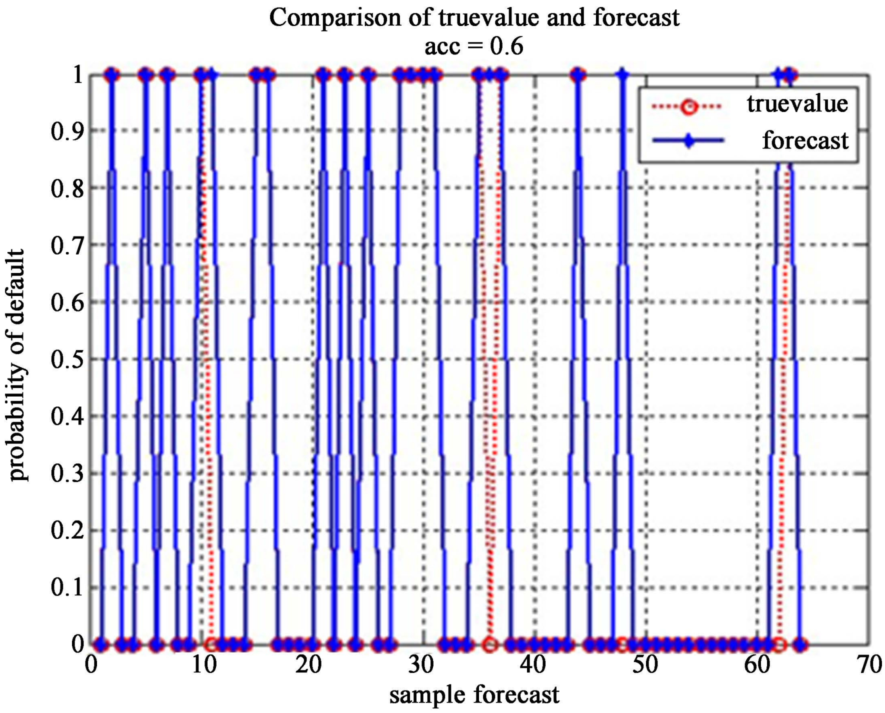
The credit risk of animal husbandry samples is also analyzed using the above 33 parameters and 320 sample size. Among the total data, 80% were selected, with 256 sample data as training sets and the remaining 64 as test simulation sets. According to the model results, the probability of being judged correctly is 88.3% among 256 training samples. The lower accuracy of animal husbandry compared with cash and field crops are due to the unique data of animal husbandry. This is mainly due to the tightening of China's environmental protection policies in recent years and the impact of African swine fever and other events that made collecting samples difficult and fluctuating relevant data. Figure 16 shows the prediction accuracy of the animal husbandry BP neural network model with 33 parameters is only 44%, which does not reach more than 50%. Figure 17 shows that the prediction accuracy of the animal husbandry GA-BP neural network is 60%, which has a better prediction effect. This model prediction accuracy is lower than that of cash and field crops due to the data.

The training dataset is divided into three groups again to prevent over-fitting.



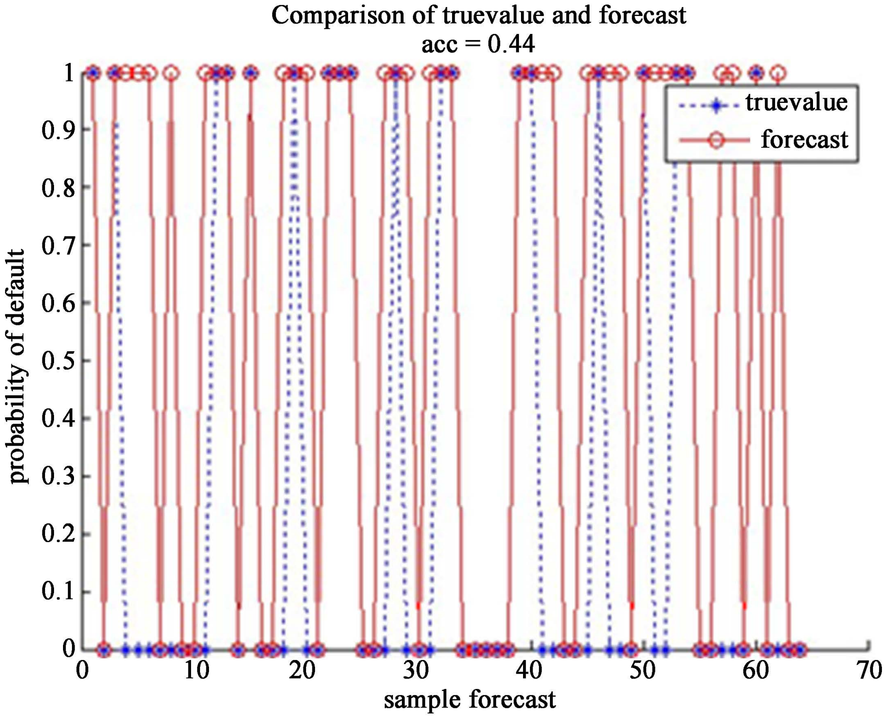
Data source: calculated and sorted by the author.

Figure 15. 20-generation adaptive diagram of GA-BP neural network for field crop samples.



Data source: calculated and sorted by the author.

Figure 16. Accurate prediction of animal husbandry samples by the BP neural network.



Data source: calculated and sorted by the author.

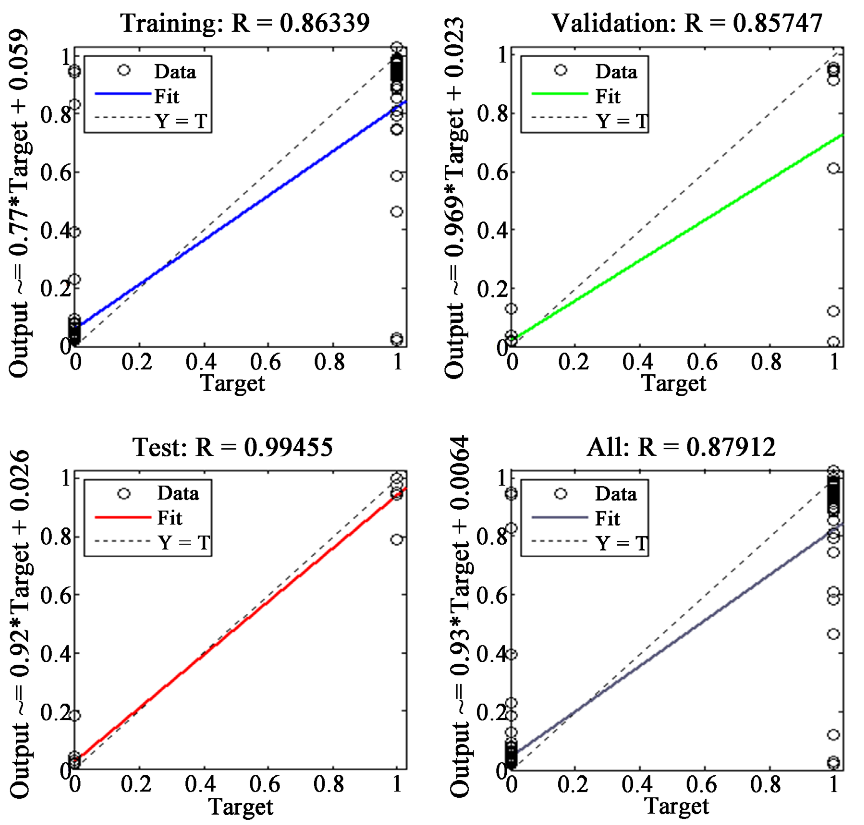
Figure 17. Accuracy of prediction of animal husbandry samples by GA-BP neural network.

Training accounts for 75%, validation 15%, and test accounts for the remaining

10%. Only the training data participated in the training. The other data did not participate in the training but are used for later inspection and explanation. The training set r of the BP neural network model is 0.986, the verification set r is 0.998, and the test simulation set r is 0.999 (Figure 18). The fitting degree is slightly higher, and the overall fitting degree is 0.989. Figure 19 shows that the training set r of the GA-BP neural network model is 0.863, the verification set r is 0.961, and the test simulation set r is 0.955. The fitting effect is better, and the overall fitting degree is 0.889.

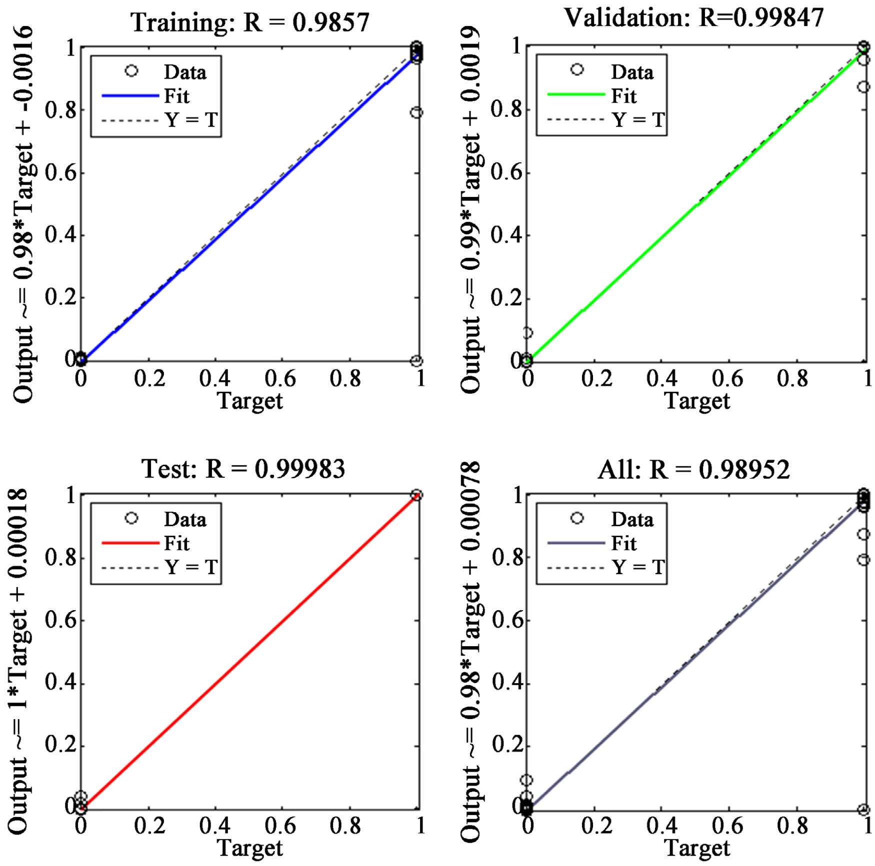
As shown in Figure 20, the variance error of the BP neural network prediction model is concentrated between −0.05 and 0, and the error rate is small, concentrated near 0. Meanwhile, the variance error of the GA-BP neural network prediction model is concentrated between −0.47 and 0.11, distributed at both ends of 0, and the error rate is small (Figure 21).

In the optimization process of the genetic algorithm, the number of evolutions (i.e., the number of iterations) is selected as 30 and the population size as 50 (Figure 22). The probability of crossing genes is selected as 0.3, and the probability of code variation is selected as 0.1. As a principle, the lower the function value of fitness, the higher the fitness, and the better the individual. When 25 iterations are reached, the individual fitness reaches the highest, and the optimal weights and thresholds are obtained simultaneously.



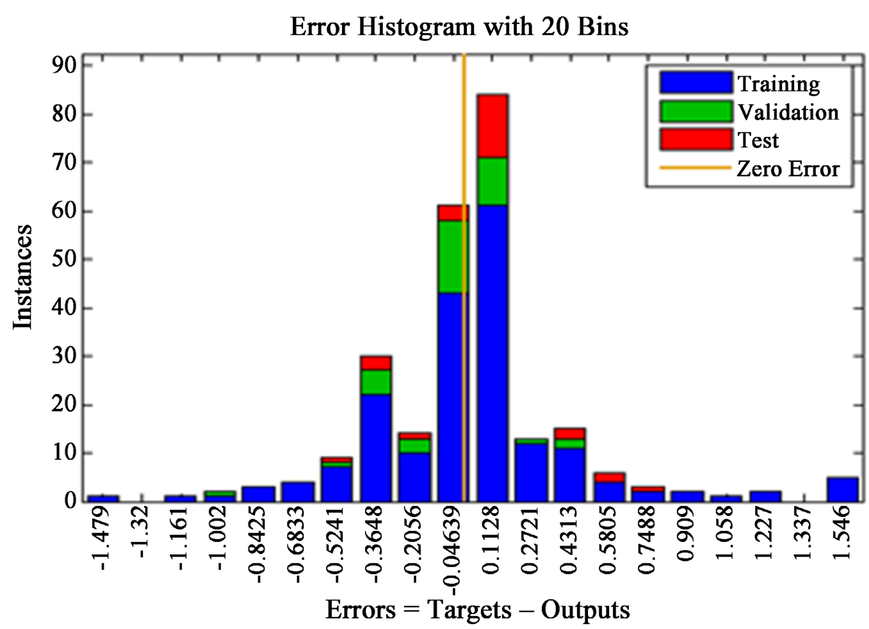
Data source: calculated and sorted by the author

Figure 18. BP neural network R of animal husbandry samples.



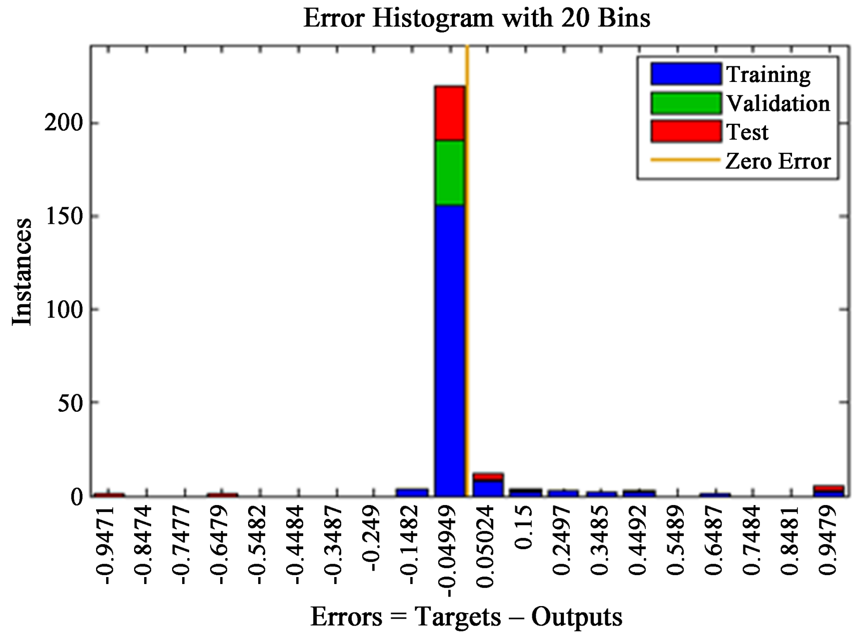
Data source: calculated and sorted by the author.

Figure 19. GA-BP neural network R of animal husbandry samples.



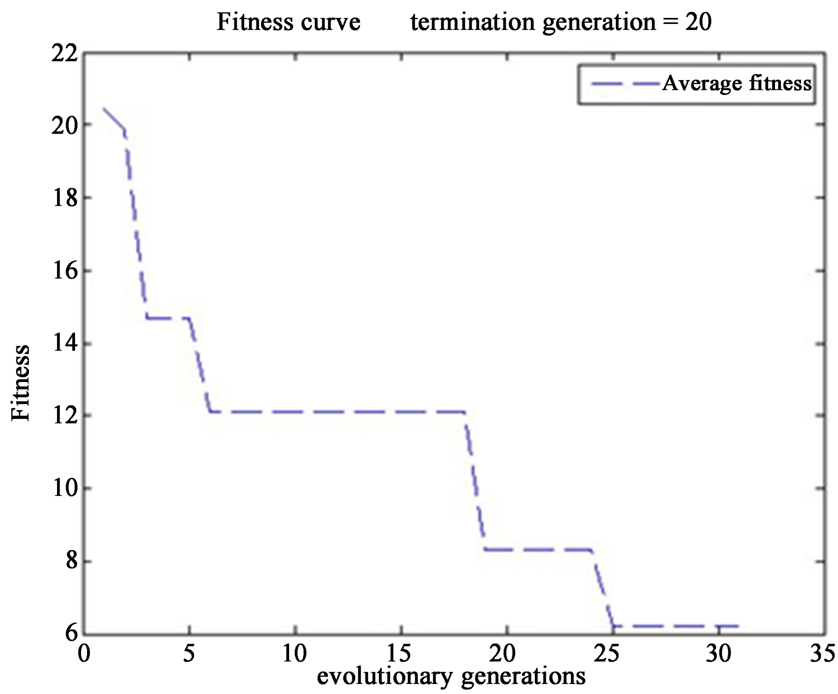
Data source: calculated and sorted by the author.

Figure 20. BP neural network error histogram of animal husbandry samples.



Data source: calculated and sorted by the author.

Figure 21. GA-BP neural network error histogram of animal husbandry samples.



Data source: calculated and sorted by the author.

Figure 22. 30-generation adaptive diagram of GA-BP neural network for animal husbandry samples.

8. Comparison of the Results of the Two Models

Combined with the credit sample data results of the above three types of agricultural business entities, the genetic algorithm is higher than the BP neural network

Table 2. Comparison of the BP neural and GA-BP neural network results of cash crop samples.

Category	Prediction accuracy	Train R	Validation R	Test R	Total R
BP	73%	0.980	0.854	0.942	0.953
GA-BP	84%	0.897	0.895	0.876	0.894

Data source: calculated and sorted by the author.

Table 3. Comparison of the BP neural and GA-BP neural network results of field crop samples.

Category	Prediction accuracy	Train R	Validation R	Test R	Total R
BP	67%	0.961	0.995	0.942	0.965
GA-BP	70%	0.923	0.817	0.984	0.908

Data source: calculated and sorted by the author.

Table 4. Comparison of the BP neural and GA-BP neural network results of animal husbandry samples.

Category	Prediction accuracy	Train R	Validation R	Test R	Total R
BP	44%	0.986	0.998	0.999	0.989
GA-BP	60%	0.863	0.961	0.955	0.889

Data source: calculated and sorted by the author.

in the overall data fitting degree. In the prediction, the unknown data are highly discriminated, and the BP neural easily falls into an over-fitting state. Thus, a genetic algorithm is more suitable for prediction analysis. It is more accurate to select the GA-BP neural network by fitting and judging the existing data. The specific comparison results are shown in Tables 2-4.

The comparison of the above tables shows the following: 1) The BP neural network is slightly faster than the GA-BP neural network in algorithm calculation. However, the latter is higher than the former in agricultural credit risk prediction accuracy. 2) The BP neural network is high enough in ontology fitting degree, but with an over-fitting phenomenon. In contrast, the GA-BP neural network does not have over-fitting when fitting sample data. It has a good fitting effect, which can better reflect the real situation of China's agricultural credit risk assessment.

9. Conclusion

Based on empirical investigation and bank interviews, this study first screened and tested the existing bank credit risk evaluation index system by Delphi method and fixed-effect model and constructed an innovative agricultural credit risk evaluation index system based on the first repayment source. Then, the BP and GA-BP neural networks were used to build the model of China's agricultural credit risk evaluation index system. Finally, the representative agricultural credit sample data collected in Zhejiang, Jiangsu, Shandong, and Henan were divided into three categories according to industries: cash crops, field crops, and animal husbandry. In particular, 80% of them are trained empirically, and the remaining 20% are simulated and predicted. After studying the results and the threshold layers of weights, the comparison table of prediction accuracy of the GA-BP neural network prediction model was obtained, which provides a reference for agricultural credit risk control. This study found the following:

First of all, in addition to credit risk, China needs to pay more attention to operational and market risks. Agricultural business entities in different industries have different risk influencing factors. Therefore, when constructing the follow-up risk assessment index system, it is necessary to screen indicators from the perspective of industry scenes. In addition, analyzing them from the perspective of the first repayment sources, such as revenue and market conditions, is essential.

Secondly, because China's agricultural credit information collection is difficult and the data are missing, it must be optimized and upgraded by combining qualitative and quantitative methods. BP, GA-BP neural network, and other machine learning methods can be introduced to improve the calculation processing, efficiency, and accuracy of risk assessment. The algorithm's discrimination degree, calculation efficiency, and data utilization rate significantly improve compared to logistics-based discrimination methods. The digitalization, automation, and artificial intelligence transformation of financial institutions are also greatly improved.

Finally, although BP neural network is slightly faster than the GA-BP neural network in calculations, the BP neural network is not as accurate in the degree of prediction. Simultaneously, it is difficult to enter the over-fitting state in the calculation, and the error distribution is relatively average, and the error concentration is high. However, if there is sufficient data to predict the overall credit risk in the future, GA-BP neural network will significantly improve the prediction accuracy and reduce the error. Simultaneously, it can effectively classify and judge the target default. This kind of method needs a lot of data and complex parameters to solve this kind of problem.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon reasonable request (e-mail: sfb007@zufe.edu.cn).

Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

References

Dutta, S., & Rating, B. (1988). A Non-Conservative Application of Neural Networks.

IEEE International Conference on Neural Networks, 2, 443-450. https://doi.org/10.1109/ICNN.1988.23958

Huang, Z. G. et al. (2019). Construction and Application of Multi-Source Data Credit Rating Universal Model Stack Framework. *The Journal of Quantitative & Technical Economics*, *4,* 155-168. (In Chinese)

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep Learning. *Nature, 521,* 436-444. https://doi.org/10.1038/nature14539

Meng, B., & Chi, G. T. (2015). Research on Credit Evaluation Model of Farmers' Microfinance. *Modernization of Management, 1,* 106-108. (In Chinese)

Micha, E. et al. (2015). Uptake of Agri-Environmental Scheme in the Less-Favored Areas of Greece: The Role of Corruption and Farmers' Responses to the Financial Crisis. *Land Use Policy*, *48,* 144-157. https://doi.org/10.1016/j.landusepol.2015.05.016

Sousa, M. R. (2015). Uptake of Agri-Environmental Schemes in the Less-Favored Areas of Greece: The Role of Corruption and Farmers' Responses to the Financial Crisis. *Land Use Policy, 17,* 123-142.