

Decision Scheme for Predicting the Quality Control of Ore Processing

Sha Yuan

School of Mathematics and Computer Science, Yan'an University, Yan'an, 716099, China

Keywords: Linear regression, residual analysis, BP neural network model, decision tree model optimization

Abstract: China is a large country of mineral resources, but its resource endowment is poor and its per capita share is not high. In order to increase the output of ores, respond to national policies, directly or indirectly save non renewable mineral resources and energy required for processing, so as to promote energy conservation and emission reduction, and help achieve the goal of "double carbon". This paper studies the quality control of ore processing, builds a model, and optimizes the model, so as to effectively improve the quality of ore processing and improve the utilization rate of ore. This paper studies the production and processing data of the workshop in the past 10 days. Without considering the influence of voltage, water pressure and other conditions, the product quality results are predicted from the known data. Through the correlation between system I and system II temperature, raw ore parameters 1, 2, 3, 4 and product quality, a BP neural network model is established, and the model analysis, inspection and improvement are carried out to obtain the product quality prediction results. Without considering other conditions, predict the temperature of System I and System II according to the known data, observe and analyze the correlation between the product quality and raw ore parameters and the temperature of System I and System II, establish an inverse model, which is also a BP Shenjing network model, and solve the model, analyze and test it to obtain the most possible temperature prediction results of System I and System II. According to the correlation between system I and system II temperature, raw ore parameters, process data 1, 2, 3, 4 and product quality index ABCD, a decision tree model is established, evaluate and optimize, and predict the product quality qualification rate with the greatest possibility.

1. Introduction

In order to improve the quality of ore processing, non renewable mineral resources and energy required for processing can be saved directly or indirectly, so as to promote energy conservation and emission reduction and achieve the goal of "double carbon". Ore processing is a complex process. During the processing, voltage, water pressure, temperature, etc. are important factors affecting ore processing, which directly affect the quality of ore products^[1]. A production workshop processes a batch of raw ores. The ore processing process is as follows: the processing process needs to go through two links, namely temperature system I and system II. The two links are in no order. When other conditions remain unchanged, production technicians change the product quality by passing in

temperature regulation instructions^[2]. The ore processing process lasts for 2 hours, and no new temperature regulating indicator will be introduced within 2 hours after each temperature regulating instruction is issued. After 2 hours of temperature adjustment, the evaluation index of ore product quality corresponding to the temperature adjustment can be detected (A,B,C,D). Through data processing, four binary linear regression models are established based on the correlation between system I and system II temperatures and the product quality index ABCD^[3]. The parameters are estimated by the least square method to obtain the output quality prediction results, and residual analysis is carried out on the results. If large errors are found, the model is optimized, the data is normalized, a BP neural network model is established, the excitation function is selected, and the number of iterations is set. Carry out analysis and inspection, and finally predict the most likely product quality index value.

2. Model establishment and solution

2.1 data processing

Data analysis and processing are carried out according to the distribution scatter diagram of system I and system II temperature and product quality index ABCD, as shown in Fig. 1, 2 and 3:

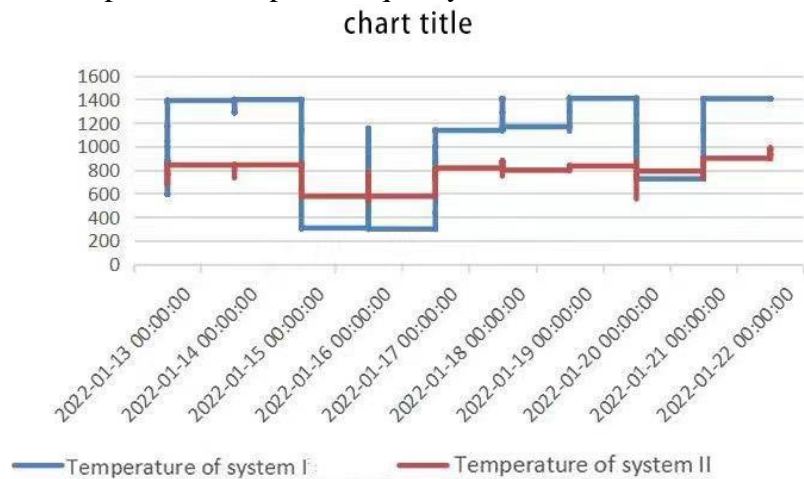


Figure 1: Temperature Distribution Diagram of System I and System II

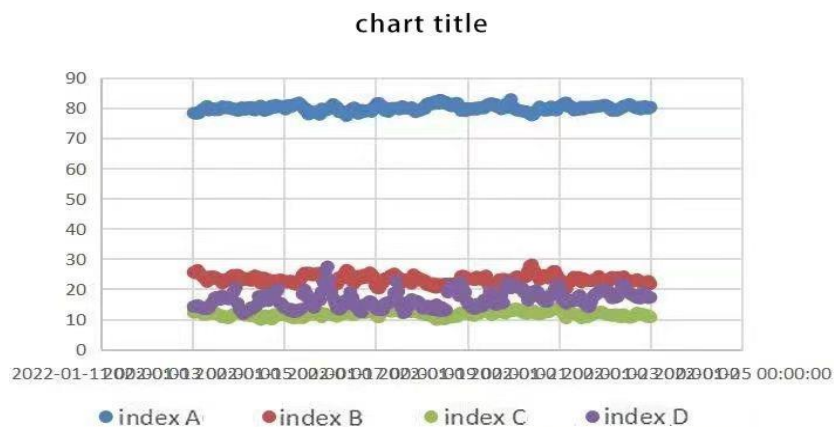


Figure 2: Scatter Distribution of Product Quality Index ABCD

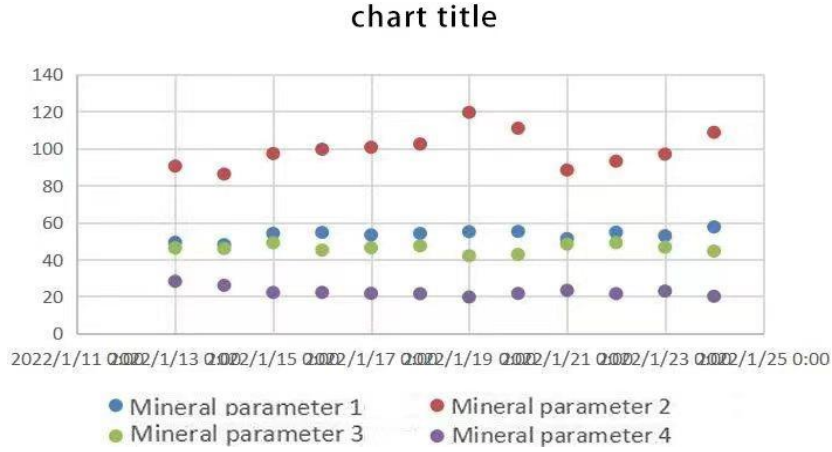


Figure 3: Scattered distribution of raw ore parameters 1, 2, 3 and 4

The system temperatures I and II are processed to correspond with the product quality indicators, and the system temperatures I and II are averaged. The first group of data of system temperatures I and II are averaged x_{i1} and x_{i2} 50 minutes ago[4]. From the 51st group, the average values x_{i1} and x_{i2} are taken at an interval of 1 hour, where $i=2,3,\dots, 235$, corresponding to the product quality indicators, and the model is established.

2.2 Establishment of model

Regression models are generally suitable for quantifying the strength of the correlation between y and x , and giving a relationship function that conforms to y and x . With the temperature of System I and System II as independent variables and the product quality (index ABCD) as dependent variables, plot a point graph[5]. After observation, a linear regression model can be established as follows:

$$Y = b_0 + b_1x_1 + b_2x_2 + \varepsilon \quad (1)$$

b_0, b_1, b_2 is the regression coefficient, x_1 and x_2 are the temperatures of system I and system II corresponding to 235 groups of product quality data, and they are independent of each other. Y is the 235 groups of data left after removing the unavailable data from the product quality. The error term ε is a mutually independent random variable with normal distribution $N(0, \sigma^2)$.

Regression coefficient is processed by least square method, and the process is as follows:

Let X be a matrix of 235 rows and 3 columns composed of $1, x_{i1}, x_{i2}$, and Y be a matrix of 235 rows and 1 column composed of y Matrix (where $i=1,2,\dots, 235$), B is a matrix consisting of b_0, b_1, b_2 with 3 rows and 1 column.

$$X = \begin{bmatrix} 1 & x_{11} & x_{12} \\ \vdots & \vdots & \vdots \\ 1 & x_{235,1} & x_{235,2} \end{bmatrix} \quad (2)$$

$$Y = \begin{bmatrix} y_{11} \\ \vdots \\ y_{235,1} \end{bmatrix} \quad (3)$$

$$B = \begin{bmatrix} b_0 \\ b_1 \\ b_2 \end{bmatrix} \quad (4)$$

$$B = (X^T * X)^{-1} * X^T Y \quad (5)$$

According to the data in the article, draw a scatter chart, fit the data, and obtain the optimal product quality (index ABCD), as shown in Fig 4.

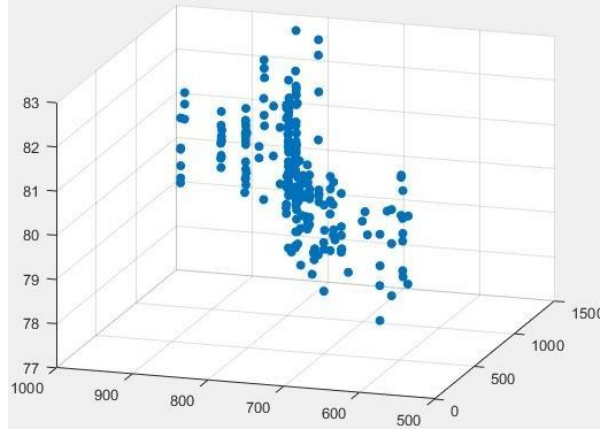


Figure 4: Scatter diagram of product quality

The regression coefficient can be obtained by fitting the temperature of system I and system II with the product quality using MATLAB software[6].

From this, we can get the function of the relationship between the index $A(Y_A)$ and the temperatures of system $I(x_1)$ and system $II(x_2)$ as shown in Table 1,2 and Fig 5,6:

$$Y_A = 79.0658 + 0.005x_1 + 0.0004x_2 \quad (6)$$

Table 1: Regression coefficient relationship

regression coefficient	Coefficient estimate	Coefficient confidence interval
b_0	79.0658	[77.4798,80.6518]
b_1	0.0005	[-0.0003,0.0013]
b_2	0.0004	[-0.0025,0.0033]
$R^2=0.0429 \quad F=5.2054 \quad p<0.0061 \quad S^2=0.8096$		

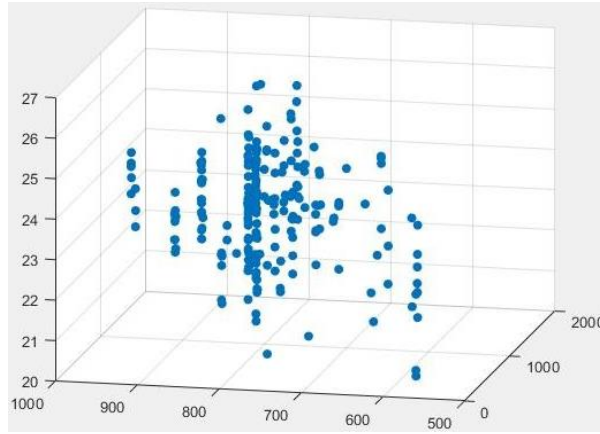


Figure 5: Relationship between index B and temperature

$$Y_B = 23.2761 - 0.0006x_1 + 0.0007x_2 \quad (7)$$

Table 2: Regression coefficient relationship

regression coefficient	Coefficient estimate	Coefficient confidence interval
b_0	23.2761	[21.0937,25.4585]
b_1	-0.0006	[-0.0017,0.0005]
b_2	0.0007	[-0.0033,0.0047]

$R^2=0.0143 \quad F=1.6813 \quad p<0.1884 \quad S^2=1.5329$

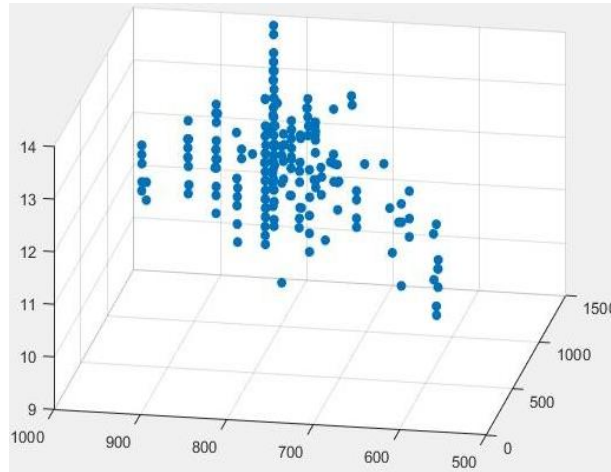


Figure 6: Relationship between index C and temperature

Get the function of the relationship between the index $C(Y_c)$ and the temperatures of system $I(x_c)$ and system $II(x_2)$ as shown in Table 3 and Fig 7:

$$Y_C = 13.4219 - 0.0002x_1 + 0.0018x_2 \quad (8)$$

Table 3: Regression coefficient relationship

regression coefficient	Coefficient estimate	Coefficient confidence interval
b_0	$\hat{13.4219}$	[11.9121,14.9316]
b_1	-0.0002	[-0.0010,0.0006]
b_2	-0.0018	[-0.0046,0.0010]

$R^2 = 0.0603$ $F = 7.4490$ $p < 0.0007$ $S^2 = 0.7336$

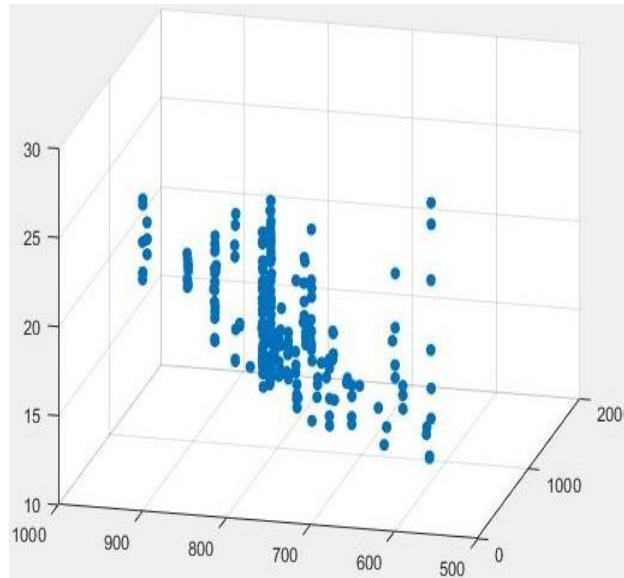


Figure 7: Relationship between index D and temperature

Get the function of the relationship between index $D(Y_D)$ and system $I(x_1)$ and system $II(x_2)$ temperature, as shown in Table 4,5:

$$Y_D = 11.4101 - 0.0012x_1 + 0.0081x_2 \quad (9)$$

Table 4: Regression coefficient relationship

regression coefficient	Coefficient estimate	Coefficient confidence interval
b_0	$\hat{11.4101}$	[6.6606,16.1596]
b_1	-0.0012	[-0.0036,0.0012]
b_2	0.0081	[-0.0007,0.0169]

$R^2 = 0.0230$ $F = 2.7281$ $p < 0.0674$ $S^2 = 7.2604$

Table 1 is obtained through four sets of regression equations and surrogate calculation

Table 5: Relationship between time and system temperature

time	Set temperature of system I	System II set temperature	Indicator A	Indicator B	Indicator C	Indicator D
2022-01-23	1404.89	859.77	80.1122	23.0350	14.6885	16.6887
2022-01-23	1151.75	859.77	79.9856	23.1832	14.7380	16.9924

2.3 Analysis and evaluation

According to the test of model A in the table, $F=1.6813$ is far greater than $F_{(1,n-2),1-\alpha} (\alpha = 0.05)$, indicating that the hypothetical model of rejecting $H_0 : b_1 = 0, b_2 = 0$ is valid, but the determination coefficient R^2 is small and the residual variance S^2 is large, indicating that the model accuracy is not high. Here, the influence of raw ore parameters on it is not considered, which is also the low accuracy of the model.

Similarly, the accuracy of BCD model is not high.

The difference between the actual value of y and the predicted value $y - \hat{y}$ is the residual error of the model, which is regarded as a random error ε . The estimated value of should obey the normal distribution with the mean value of 0. The residual graph can be obtained by the program. It is found that the confidence interval of the residual of some data does not contain zero points. It can be considered that these points deviate from the change trend of the overall data, and these abnormal points should be eliminated.

After removing the abnormal points, the recalculation results are as follows in Table 6:

Table 6: Regression coefficient relationship

regression coefficient	Coefficient estimate	Coefficient confidence interval
b_0	78.9998	[78.1098,79.8898]
b_1	0.0005	[-0.0003,0.0010]
b_2	0.0005	[-0.0020,0.0030]
$R^2=0.0501 \quad F=5.9054 \quad p<0.0073 \quad S^2=0.6054$		

By analyzing the data in Table 5 and Table 6, the estimated value has little change, but the confidence interval becomes shorter, R^2 and F become larger, and S^2 decreases, indicating that the accuracy of the model has been improved.

2.4 Model optimization

235 groups of factors in raw ore processing (i.e. system temperature I, system temperature II, raw ore parameter 1, raw ore parameter 2, raw ore parameter 3, raw ore parameter 4) are taken as input, and the corresponding product quality is taken as output, and these data are normalized with the premnmx function of matlab. The image is as follows in Fig. 8:

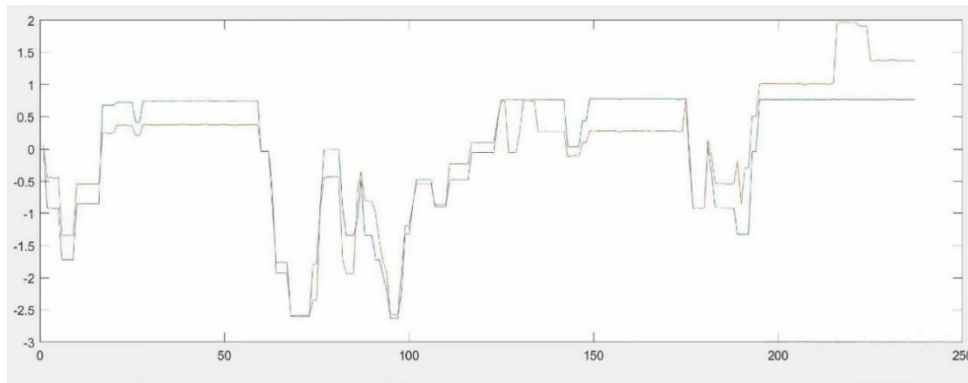


Figure 8: Data normalization processing diagram

Data set: (Note: each column is a set of input factor sets, the number of rows represents the number of neurons in the input layer, and the number of columns represents the number of input factor sets)

Matrix P with 6 rows and 235 columns Matrix T with 235 rows and 1 column

3. Establish BP network model

BP network (Back Propagation Network) is also known as back-propagation neural network. Through the training of sample data, network weights and thresholds are constantly revised to make the error function decline along the negative gradient direction and approach the expected output. It is a widely used neural network model, which is mainly used for function approximation, model recognition and classification, data compression and time series prediction.

BP network has high nonlinearity and strong generalization ability, but it also has some shortcomings, such as slow convergence speed, many iterative steps, easy to fall into local minima and poor global search ability. Genetic algorithm can be used to optimize the "BP network" to find a better search space in the analytic space, and then BP network can search for the optimal solution in a smaller search space.

3.1 Model Solution

(1) Design of input and output layer

The model takes each group of data (system temperature I, system temperature II, raw ore parameter 1, raw ore parameter 2, raw ore parameter 3, raw ore parameter 4) as input and quality (indicator ABCD) as output, so the number of nodes in the input layer is 6 and the number of nodes in the output layer is 4.

(2) Hidden layer design

Relevant research shows that a neural network with a hidden layer can approximate a nonlinear function with arbitrary accuracy as long as there are enough hidden nodes. Therefore, this paper uses a three-layer multi input single output BP network with a hidden layer to establish the prediction model. In the process of network design, it is very important to determine the number of hidden layer neurons. If the number of hidden layer neurons is too large, it will increase the amount of network calculation and easily lead to the problem of over fitting; If the number of neurons is too small, the network performance will be affected and the expected effect will not be achieved. The number of hidden layer neurons in the network is directly related to the complexity of practical problems, the number of input and output layer neurons and the setting of expected error. At present, there is no clear formula for determining the number of neurons in the hidden layer, only some empirical formulas. The final determination of the number of neurons still needs to be based on experience and

multiple experiments. In this paper, the number of hidden layer neurons is determined to be 7 according to previous experiments, as shown in Fig 9.

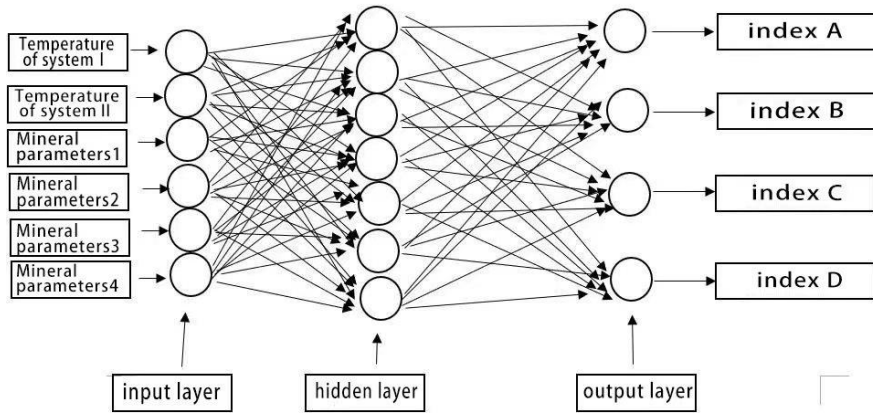


Figure 9: Hidden layer neural network

3.2 Selection of excitation function

BP neural network usually uses sigmoid differentiable function and linear function as the excitation function of the network. In this paper, the S-type tangent function tansig is selected as the excitation function of hidden layer neurons. Because the output of the network is normalized to the range of $[-1,1]$, the prediction model selects the S-type logarithmic function logsig as the excitation function of the output layer neurons.

3.3 Model implementation

This prediction uses the neural network toolbox in MATLAB to train the network. The specific implementation steps of the prediction model are as follows:

After the training sample data is normalized, it is input into the network, and the excitation functions of the hidden layer and output layer of the network are set as tansig and logsig functions respectively. The network training function is traindxd , the network performance function is mse , and the number of hidden layer neurons is initially set as 7. Set network parameters. The number of network iterations epochs is 1000, the expected error goal is 0.00,001, and the learning rate lr is 0.01. After setting the parameters, start training the network, as shown in Fig 10.

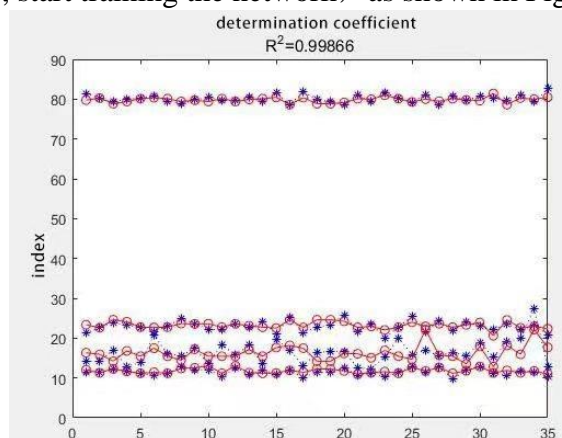


Figure 10: Training Network Diagram

The network completes the learning after reaching the expected error through 14 repeated

learning. After the network training is completed, you only need to input various quality indicators into the network to get the prediction data.

The prediction results are in Table 7:

Table 7: Relationship between time and system temperature

time	Set temperature of system I	System II set temperature	Indicator A	Indicator B	Indicator C	Indicator D
2022-01-23	1404.89	859.77	80.1468	22.9053	11.3363	16.6767
2022-01-23	1151.75	859.77	79.5164	23.7197	12.0251	15.4724

4. Conclusion

By comparing the two models, it is found that the R^2 of BP network model is larger, close to 1, the correlation between independent variables and dependent variables is more significant, and the fitting performance is better. To sum up, both methods have good fitting degree and prediction effect, but BP neural network has more significant fitting effect, smaller relative error and higher fitting accuracy, which can predict the product index ABCD as much as possible. Through data processing, four binary linear regression models are established based on the correlation between system I and system II temperatures and the product quality index ABCD. The parameters are estimated by the least square method to obtain the output quality prediction results, and residual analysis is carried out on the results. If large errors are found, the model is optimized, the data is normalized, a BP neural network model is established, the excitation function is selected, and the number of iterations is set. Carry out analysis and inspection, and finally predict the most likely product quality index value.

References

- [1] Luo Xuehui, Zhang Yong, et al. *Process control of gold ore processing and testing quality. Environmental Protection and Analysis No. 3/Volume 33, 2012 June 1~64, 2012*
- [2] Wang D, Liu S J, Mao Y C, et al. *A Method Based on Thermal Infrared Spectrum for Analysis of SiO₂ Content in Anshan-Type Iron[J]. Spectroscopy and Spectral Analysis, 2018.*
- [3] Jiang Qiyuan, Xie Jinxing, Ye Jun, et al. *Mathematical Models. Higher Education Press, May 2018 (reprinted on November 2020) 2020*
- [4] Qiao Kelin, Ren Fangling. *Probability Statistics and Application. China Business Press December 2017*
- [5] Chen S, Luk B L, Harris C J, et al. *Fuzzy-logic tuned constant modulus algorithm and soft decision-directed scheme for blind equalisation[J]. Digital Signal Processing, 2010, 20(3):846-859.*
- [6] Abdul M, Tae-Sun C. *A New Ensemble Scheme for Predicting Human Proteins Subcellular Locations[J]. Communications in Computer and Information Science, 2010, 61(1):185-192.*