Research on the Health State Evaluation Method of Coal Shearer Based on Improved XGBoost

Yanchuan Li¹,², Xiangang Cao¹,², Mengyuan Zhang¹,², Xinyu Duan¹,², Youjun Zhao³

¹Mechanical Engineering college, Xian University of Science and Technology
²Shaanxi key laboratory of mine mechanical and electromechanical equipment Intelligent Monitoring, Xian, China
³Xi'an Coal Mining Machinery Co., Ltd.

Abstract: The health evaluation of shearsers has always been a hot and difficult point in the study of shearer health management. In this paper, aiming at the problems of redundancy of current shearer status evaluation parameters and imbalance of the evaluation data set, this paper proposes a shearer health status evaluation method based on the improved extreme boosting tree algorithm (XGBoost). First of all, in view of the redundancy of the multi-dimensional monitoring parameters of the shearer, the SP correlation coefficient combining the Spearman correlation coefficient and the Person correlation coefficient is proposed. According to the calculation results of the SP correlation coefficient, the shearer health indicator is screened out, and the coal mining machine is constructed. Machine health evaluation index system; In view of the poor interpretability of traditional neural network evaluation methods and the unbalanced data of the shearer data set categories, a shearer health evaluation method based on improved XGBoost is proposed, and the evaluation of the XGBoost model Key parameters such as the maximum depth of the tree max_depth, the minimum leaf node weight and min_child_weight are improved and tuned; finally, the evaluation model verification and analysis are carried out through the shearer instance data set. The experimental results show that the average accuracy of the health status evaluation method based on the improved XGBoost can reach 98.50%, and the average F1 average can reach 97.61%, which can effectively solve the problem of imbalance of the data set and the problem of status evaluation.

Keywords: Shearer; correlation analysis; evaluation index system; XGBoost; status evaluation

1 introduction

As a multi-component complex system, the shearer is affected by various complex factors such as variable working conditions, variable load, and environmental noise when cutting coal and rock. The collected monitoring signals such as vibration and current are often difficult to effectively extract the key information. At present, the research work on the health evaluation of the shearer is mainly focused on the modeling method. Zhai Wenrui[1] modeled the performance degradation process of the shearer on the basis of selecting the monitoring data of the shearer working condition, and used the extreme learning machine to evaluate the operating condition of the shearer cutting part; Lei Yinan[2] passed The main parameters of the shearer health operation are analyzed, combined weighting method and fuzzy comprehensive evaluation method are used to evaluate the current status of the shearer, and the shearer health monitoring system is designed, but the status evaluation model is not applied to the evaluation system. Si Lei[3] built a state evaluation system for the shearer, based on rough set attribute reduction and BP neural network to evaluate the operating status of the shearer. However, the selected index parameters have fewer dimensions and the neural network parameters lack optimization. Aimed at the high-dimensional data set of the shearer.

At present, domestic and foreign scholars' health evaluation methods for other multi-component systems and equipment can be roughly divided into the following three categories according to different evaluation principles: experience-based health evaluation[4], model-based health
evaluation\textsuperscript{5} and Health status assessment based on artificial intelligence\textsuperscript{6}. Among them, the experience-based health evaluation includes: grey theory method, fuzzy comprehensive evaluation method\textsuperscript{7} cloud model theory\textsuperscript{8}, analytic hierarchy process, D-S evidence theory; model-based health evaluation mainly includes failed physical models, faults Tree et al\textsuperscript{9}\textsuperscript{10}; artificial intelligence-based health assessment methods include: support vector machines, Markov theory, Bayesian networks, artificial neural networks\textsuperscript{11}\textsuperscript{12}\textsuperscript{13}, etc. Lai Yuehua\textsuperscript{14} and others established a physical model for each failure mode by considering various variables according to the failure mode and failure mechanism of the equipment. In view of the lack of objectivity in the selection of the membership function of each state of wind turbines, Dong Xinghui\textsuperscript{15} and others studied a method for evaluating the state of wind turbine equipment based on the combination of cloud model theory and combination weights, and used the membership cloud model to calculate wind turbines. The degree of membership of each level is combined with the combined weighting method to evaluate the health status of wind turbines; Zhao Dongming\textsuperscript{16} and others have studied a CNN-based generator status evaluation method. The equipment is divided into information layer and characteristic layer. Establish the generator equipment state evaluation model, and finally select the generator equipment monitoring data as the training data, and apply the CNN model to get the current health status of the generator.

In summary, the model-based evaluation method is difficult to model and solve is extremely complicated, while the experience-based evaluation method is greatly affected by subjective factors, and it is difficult to determine the weight of each component and index. Evaluation methods based on traditional neural networks have poor interpretability and tend to fall into local minimums. In addition, most health assessment methods cannot target redundant, mutated, and unbalanced data sets, resulting in low assessment efficiency. In response to the above problems, this paper proposes a method for screening the state index parameters of the shearer based on the SP correlation coefficient, screening parameter data with lower correlation, constructing a shearer health evaluation index system and selecting health evaluation indicators; For problems such as unbalanced samples of the shearer, the improved XGBoost algorithm is used to evaluate the health status of the shearer, and two key parameters are tuned. The evaluation result of the XGBoost algorithm and the confusion matrix are obtained through experimental verification to be the health status of the shearer. The evaluation work provides a basis, which has a certain significance for the health management of the shearer.

2 Construction of Evaluation Indexes for Shearer's Health Status

2.1 Analysis of the basic structure and monitoring parameters of the shearer

According to the form of working structure, the shearer can be classified into four types: drum type, drilling type, vertical drum and frame type. This article mainly studies the electric traction double drum shearer. The double-drum shearer can be divided into traction part, cutting part, electrical system and auxiliary device. Among them, the cutting part is composed of a cutting drum 1, a rocker gear box 2, a cutting motor 3, etc., and its main function is coal mining and coal falling. The traction part is mainly composed of a traction motor 4, a traction reduction box 11, etc., and is a traction device for the shearer to travel. The electrical system mainly includes the electrical control box 7, the frequency conversion box 8, the solenoid valve cabinet 9, the transformer box 10 and other functions such as operation, control, and speed regulation. The auxiliary devices include the height adjustment device 12, the crushing mechanism 13, and the crushing motor 14. Wait. Therefore, the names of the parts of the double-drum electric traction shearer are shown in Fig.1.

![Fig.1 Basic structure of electric traction shearer](image)

1. Cutting drum 2. Rocker arm 3. Cutting motor 4. Traction motor 5. Oil pump motor 6. Hydraulic unit 7. Electrical control box 8. Frequency conversion box 9. Solenoid valve cabinet 10. Transformer box 11. Traction reduction box 12. Height adjustment device 13. Crushing mechanism 14. Crushing motor

2.2 Selection of Shearer State Quantity and Data Collection

According to the structure and working method of the shearer, and considering the actual installable parts of the shearer, the main operating status monitoring data types of the shearer are analyzed and obtained, including various motor temperature data, various part temperature data, various water pressure data, oil Several items such as oil
pressure data, various vibration data, various motor current and voltage data. Among them:

The motor temperature data mainly includes the temperature of the left and right cutting motors, the temperature of the left and right traction motors, and the temperature of the crushing motor. The temperature data of other parts of the equipment mainly includes the temperature of the traction transformer, the temperature of the gear box, and the temperature of the inverter floor.

The various pressure data, oil level and oil pressure data of the shearer mainly include the water pressure of the cooling pump, the oil pressure of the hydraulic system pipeline, and the oil pressure of the oil tank of the height adjustment section. The various vibration data of the shearer mainly include the vibration of the traction motor and the vibration of the cutting motor.

Various voltage and current data include current and voltage of traction motor, current and voltage of cutting motor, voltage and current of transformer, and current and voltage of crushing motor. Corresponding sensors are installed on each key component or part of the shearer to monitor, monitor and ensure the health of the shearer in real time.

2.3 Correlation analysis of state parameters of shearer

The shearer is a highly coupled system that integrates electromechanical and hydraulics with multi-component collaborative work. During the operation of the shearer and the coal mining process, the components and parts are closely connected, and the status parameters monitored by each component during the operation and mining process. There are complex association relationships, so it is necessary to find out these relationship data, remove data redundancy, and prepare for the next step of the shearer health evaluation work. At present, the correlation coefficients of the two variables are calculated as:

(1) Spearman Correlation coefficient

The Spearman correlation coefficient is generally called the rank correlation coefficient. It analyzes the correlation between the two parameter variables through the rank size of the two parameter variables. The calculation formula is as follows:

\[ \rho_s = 1 - \frac{6 \sum d_i^2}{N(N^2 - 1)} \]  

\[ \rho_{1} \] is the Spearman correlation coefficient between the two variables; \( N \) is the sample size; \( d_i \) is the rank difference between the variables.

(2) Person Correlation coefficient

Person correlation coefficient generally expresses the correlation between two variables with a linear relationship, generally taking a value between -1 and 1. A positive number indicates a positive correlation between the two parameter variables, and a negative number indicates a negative correlation between the two parameter variables. When the preconditions are certain, if the absolute value is used to express the correlation between two parameter variables, the stronger the correlation between the two parameter variables, the greater the absolute value of the Person correlation coefficient, and vice versa. The calculation formula of the Person correlation coefficient between data variables \( X \) and \( Y \) is as follows:

\[ \rho_p = \frac{\text{cov}(X,Y)}{\sigma_X \sigma_Y} = \frac{E((X - \mu_X)(Y - \mu_Y))}{\sigma_X \sigma_Y} \]  

\[ \text{cov}(X,Y) \] represents the covariance of \( X \) and \( Y \); \( \sigma_X, \sigma_Y \) represents the standard deviation of \( X \) and \( Y \); \( \mu_X \), \( \mu_Y \) is the average of the variables \( X \) and \( Y \).

(3) SP Correlation coefficient

A single correlation coefficient cannot objectively characterize the close correlation between data variables, so this paper chooses a comprehensive correlation coefficient, that is, the Spearman correlation coefficient and the Person correlation coefficient are used to characterize the correlation between the shearer state data variables. The SP correlation coefficient between the state parameters is represented by \( r_{sp} \), and the comprehensive coefficient SP is:

\[ r_{sp} = \frac{1}{2} (\rho_s + \rho_p) \]

By calculating the value of the SP correlation coefficient \( r_{sp} \) between the status parameters, the status evaluation indicators are screened out, and the shearer health status evaluation system is constructed. \( r_{sp} \) is used to represent the comprehensive correlation coefficient between the state monitoring parameters \( x \) and \( y \), so \( r_{sp}[-1,1] \), the correlation between the value of the correlation coefficient \( r_{xy} \) between the two state parameters and the corresponding two variables is described in the table 2.

| Table 2 | The relationship between the absolute value of the correlation coefficient and the state monitoring parameters |
Through the calculation of the comprehensive correlation coefficient of the monitoring parameters of the shearer in the same environment, the multi-dimensional data with no obvious correlation is selected as the index parameter, which reduces the complexity of the input characteristics of the evaluation model and ensures the further health evaluation of the shearer.

This paper takes the monitoring parameters of the shearer traction unit as an example, completes the correlation analysis between the monitoring parameters, selects the parameters with strong correlation between the status parameters as the index data, and then completes the screening of the shearer status evaluation indicators. In order to further analyze the correlation between the state monitoring parameters and indicators of the shearer, the 8 state parameter sequence data of the traction part under normal operating conditions are selected. All the monitoring data of the traction part of the shearer include the temperature of the traction motor, the speed of the traction motor, and the traction motor. Eight parameters such as vibration, traction motor current, cooling water pressure, traction motor torque, oil pressure in the cylinder, and temperature of the traction reduction gear box are calculated. The comprehensive correlation coefficients between these eight parameters are calculated respectively, and the correlation degree is higher than the threshold value of 0.6. Several parameters are replaced with the same index parameter to reduce redundant attributes between monitoring parameters. The correlation heat map results obtained through correlation analysis are shown in Fig.2:

It can be seen from Fig.2 that the temperature of the C1 traction motor is closely related to the voltage of the C2 traction motor. Therefore, the temperature of the traction motor can be used to replace these two indicators. The monitoring parameters are in order: traction motor vibration, traction motor temperature, traction motor current, and traction motor speed. By analogy, the correlation analysis results of the state parameters of other components are obtained, the state parameters with low correlation are selected as the health evaluation indicators of the shearer, and the data with strong correlation is eliminated.

2.4 Construction of evaluation index system for shearer health status

Since the shearer integrates mechanical, electrical and hydraulic systems, there are many monitoring parameters that affect the state of the shearer. However, if all of them are used as indicators for the evaluation of the state of the shearer, it will increase unnecessary workload and impact assessment. Work efficiency. Therefore, on the basis of coal mine investigations, this paper combines the structure of the shearer, the location of the prone to failure part and the correlation analysis of the monitoring parameters of the shearer in Section 3.2.2, and screens out the most important multi-dimensional indicators that affect the health of the shearer. On this basis, determine the composition of the shearer health evaluation index system, which is divided into layers from the index layer, the component layer to the whole machine layer, as shown in Fig.3.
3 The health evaluation model of shearer based on improved XGBoost

3.1 XGBoost

XGBoost is an excellent integrated learning method in Boosting. It was proposed by Professor Chen Tianqi and others in 2016. Once the XGBoost algorithm appeared, it has achieved good results in various data mining competitions. XGBoost, as an ensemble learning method, uses multithreading to accelerate the construction of trees, uses the tree model as a basic classifier to form a powerful classifier, and integrates multiple basic classifiers together, which is highly efficient and effective in classification tasks. The advantages of accuracy and good interpretability.

Aiming at the problem of sample imbalance, the XGBoost algorithm has a parameter scale_pos_weight that adjusts the sample imbalance. Usually, the ratio of the negative sample to the positive sample is input as this parameter, which can well solve the problem of imbalance in the state data set of the shearer. In addition, the objective function of XGBoost directly uses Taylor expansion to fit the tree model, so the fitting effect is better. The basic concepts and theories of the XGBoost algorithm are as follows:

(1) Base learner

The extreme gradient boosting tree is composed of two basic components: regression tree and classification tree. XGBoost (extreme gradient boosting) uses classification and regression tree (classification and regression tree, CART) as the basic learner, and uses XGBoost to train the evaluation model. The attribute of the feature is transferred to each leaf node, corresponding to the score of each leaf.

(2) Tree complexity

Each regression tree can be divided into a structure part and a leaf node weight part, then the t-th tree model

\[ f_t(x) = w_{q(x)}, w \in \mathbb{R}^T \]

where \( w \) is the score of the leaf node; \( q(x) \) is the number of the leaf node corresponding to the sample \( x \); \( T \) is the number of leaves, and \( \mathbb{R}^T \) is the T-dimensional real number, which represents the set of leaf weights. The complexity includes the number of nodes in a tree and the modulus square of the output score on each leaf node. Therefore, the complexity of the tree is:

\[ \Omega(f) = \gamma T + \frac{1}{2} \sum_{j=1}^{T} W_j \]

In the formula, \( \gamma \) is the complexity, \( \gamma \) is the penalty coefficient of the number of leaf nodes, \( \lambda \) is the regular term coefficient, \( W_j \) is the score corresponding to the leaf node \( j \).

(3) The objective function is:

\[ O^c_t = \sum_{i=1}^{n}(G_i f_i + \frac{1}{2} \theta_f f_i^2(x_i) + \Omega(f_i)) \]

The objective function formula (3) is the Taylor expansion, \( n \) is the order, \( X_i \) is the i-th training sample, \( O^c_t \) is the objective function, and \( c \) is the number of training iterations.

\[ O^c_t = \sum_{i=1}^{n}(G_i w_j + \frac{1}{2} (H_i + \lambda) w_j^2) + \gamma T \]

In the formula, \( I = \{ i | q(x) = j \} \), \( G_i = \sum_{n \in I} \beta_i \), \( H_j = \sum_{n \in I} \beta_j \).

(4) Scoring function
Assuming the structure of a known number, the optimal value of \( w \) and the corresponding objective function \( O^* \) can be obtained \( w^* \) from (9) as:

\[
w^* = -\frac{1}{H} - \frac{1}{\lambda + \lambda}
\]

(10)

\[
O^* = -\frac{1}{2} \sum_{j=1}^{J} \frac{G_j^2}{H_j + \lambda} + \lambda T
\]

(11)

(5) Gain

When creating the tree model, Greedy algorithm can be used to add segmentation to existing leaves each time. For a specific segmentation scheme, the gain obtained is:

\[
G_{ns} = \frac{1}{2} \left( \frac{G_1^2}{H_1 + \lambda} + \frac{G_2^2}{H_2 + \lambda} \right) - \lambda
\]

(12)

The state data of the shearer is used as the feature input, and the four health states of the shearer are used as the classification output result. Through model training, through the optimization and comparison of various model parameters such as the maximum depth of the tree, the best parameters are obtained value.

### 3.2 Improved XGBoost's shearer health evaluation model

(1) XGBoost evaluation model training

In this section, XGBoost integrated learning is used to establish a shearer health evaluation model. By dividing the health status of the shearer, training samples are established, and the key parameters of the XGBoost model are optimized. Finally, the health status of the shearer is evaluated, and a conclusion is drawn. The main process of applying XGBoost to assess the health status of the shearer is as follows:

The first step of shearer health evaluation is to use the index data after relevant analysis and feature selection as the input features of XGBoost, and the different health status levels of the shearer as the category labels of the evaluation algorithm.

The second step is to divide the shearer state data set. Through the selection of the state quantity and the construction of the index system, the state data set of the shearer is divided into a training set and a test set, which are divided according to a certain ratio.

The third step is to initially set the main parameters of the XGBoost classification model. After the model is established, various parameters of the shearer state evaluation model are set, such as the maximum depth of the tree, the learning rate of the model, the minimum leaf weight and so on.

The fourth step is to use the training set data to train the shearer XGBoost state evaluation model, and use the shearer data test set to test the model. By constructing a CART decision tree, and then adding state classification nodes in turn to fit the previous evaluation results respectively, the goal in the training process is to minimize the loss function, and the feature with the smallest loss function is obtained as the feature of the bifurcation tree. On this basis, the prediction score of each leaf node, that is, the state is calculated, and the prediction score of each evaluation result of each tree is used as the probability value, and the state classification and evaluation are completed according to the maximum probability value.

The last step is to continuously adjust the XGBoost model parameters. Check the classification effect of the evaluation model by changing the values of various parameters, and use the XGBoost parameters with the best comprehensive evaluation effect as the final evaluation model parameters. The specific evaluation process is shown in Fig.4.
To ensure the accuracy of the evaluation results, this section selects 1000 pieces of the same state index data of the shearer as in section 4.2 as the experimental data, and imports the normalized 15-dimensional data of the shearer index parameters such as the temperature of the traction motor into XGBoost Evaluation model. Use 80% of the data set as the training set and 20% as the test set. Through the initial parameter settings of the XGBoost model, the Python language is used for programming simulation experiments. After constructing the shearer health evaluation model of the XGBoost algorithm, various parameters of XGBoost need to be optimized through a reasonable parameter adjustment method, and the results of the parameter adjustment are constantly compared to obtain the best model parameters.

In this chapter, the classification error rate of the training set and the test set is used as the evaluation index of the model, and the optimal parameters of the shearer state evaluation model are obtained through multiple adjustments. The average accuracy rate and average recall rate of the state evaluation model are used as evaluation indicators to judge the classification effect of the evaluation model. The main initial parameter settings of the XGBoost model are shown in Table 2, and the remaining parameters are set according to default values.

### Table 2 Key parameters of XGBoost evaluation model

| Parameter       | Initial Value |
|-----------------|---------------|
| max_depth       | 2             |
| num_boost_round | 300           |
| min_child_weight| 4             |
| subsample       | 0.2           |
| colsample_bytree| 1             |
| gamma           | 0             |
| num_class       | 4             |

Before optimizing the parameters of the XGBoost shearer evaluation model, it is first necessary to analyze the key parameters that affect the evaluation efficiency of the XGBoost model. The first type of parameters are the parameters that adjust the over-fitting, that is, the maximum depth of the tree `max_depth`, the minimum leaf node weight, and `min_child_weight`. Generally speaking, the deeper the `max_depth`, the more detailed and specific data sample information can be learned by the evaluation model, but when the depth of the tree is too deep, over-fitting may occur. At this time, the test set data classification error rate is higher, and the training set classification error rate is lower; the larger the value of `min_child_weight`, the more comprehensive the characteristics of the sample can be learned, but when the value of `min_child_weight` is too large, the model learns more useless information, so over-fitting occurs. Therefore, this chapter mainly optimizes these two key parameters.

In this paper, the parameters of the XGBoost evaluation model are tuned through cross-validation. When selecting the optimal depth of the tree, by changing the depth of the tree, that is, when the maximum tree depth `max_depth` is set to 2, 4, 6, and 8, respectively, compare the multi-class error rates of the training set and test set data, as shown in Fig.5.

![Graph showing classification error rates](image)
From the comparison of the classification error rate of the XGBoost model with different tree heights in Fig. 5, when the tree height is set to 2, 4, the classification error rate of the training set and the test set is not much different and can be controlled within a relatively small range. The tree height When it is 4, the average classification error rate is smaller; when the height of the tree is set to 6, 8, the error rate is small, but the gap between the training set and the test set is too large, which is not suitable as the best parameter, and the tree height is selected as 4. The most reasonable.

The minimum leaf node weight and the value of "min_child_weight" are generally between 4-10. When the value of "min_child_weight" in this chapter is 4, 6, 8, 10, the classification error rate of the training set and test set of the XGBoost model is obtained as Shown in Fig. 6.

According to the comparison of the classification error rate of the XGBoost model when the min_child_weight is different in Fig.6, when the min_child_weight is set to 4, 6, the classification error rate of the training set and test set is not much different and can be controlled in a relatively small range, when the min_child_weight value is 6. The classification error rate is smaller; when the
The next parameters to be adjusted are the minimum loss function descent value gamma and the random sampling ratio subsample, and the random column number ratio colsample_bytree. Gamma represents the drop value of the corresponding loss function when each node is divided. If the algorithm is more conservative, the value of gamma is larger. In this chapter, the size of gamma is adjusted between 0-0.5 based on experience, with an interval of 0.1 each time. Through experiments, it is found that the best gamma value is 0.1, and the accuracy rate is 0.985. The random column number ratio colsample_bytree represents the generation time of the decision tree, and the random sampling ratio subsample represents the ratio of the sampled samples to the entire sample. By continuously adjusting the parameters, the optimal combination of colsample_bytree and subsample is obtained when the value of the former is 1, and the value of the latter is 0.8, when the model has the best effect.

4.2 Improved XGBoost model result analysis

On the basis of setting the optimal values of various parameters, import the shearer state data set for training and testing. The model training steps are shown in 2.3. After the model is trained, 200 sets of test data are imported into the XGBoost state evaluation model, and the evaluation accuracy rate, the recall rate of each health state, and the value of the comprehensive evaluation parameter F1 are calculated to judge the model. The evaluation accuracy is an indicator of the overall quality of the evaluation model. The closer the data is to 1, the better the overall evaluation effect of the model. However, there is an imbalance in the health evaluation samples of the shearer. The problem is that the number of healthy samples is far greater than that of unhealthy samples. A single accuracy index cannot qualitatively evaluate the effect of the model. Therefore, the number of correctly classified samples in each type of sample accounts for the proportion of the type of healthy samples. The applicability of the evaluation model to the problem of sample imbalance. At the same time, in order to avoid the shortcomings of the single accuracy rate and recall rate evaluation index, the comprehensive evaluation index F1 value of the two is used to comprehensively reflect the effect of the evaluation model. The closer the F1 value is to 1, the better the classification effect of the evaluation model. By running the program, the multi-class error rate mirror of the test set during the evaluation process is shown in Fig.4.10, and the specific evaluation results of the model obtained are represented by a confusion matrix as shown in Fig.7.
In the XGBoost evaluation model, the ordinate represents the actual health level, and the abscissa represents the health level predicted by XGBoost. It can be seen from Fig.8 that among the 200 sets of data, there are 197 sets of status data that are correctly classified, that is, the corresponding health status can be accurately obtained for the 197 sets of status data, and 1 set of data belongs to the "healthy" state and is classified as "good" state, there are two sets of data belonging to the "degraded" state and classified as a fault state, but the state prediction result is only one level different from the actual state, and the effect on the result is not particularly large. The overall evaluation effect of the model is good, the overall accuracy rate is as high as 98.50%, the accuracy rate of the "healthy" state level is 100%, the accuracy rate of the "good" state level is 98.57%, and the accuracy rate of the "degraded" state level is 100%, "fault" The state level assessment accuracy rate is 89.47%, the average recall rate of the four health status levels is 98.38%, the average F1 level is 97.61%, and the average recall rate and F1 value are relatively high, indicating that the model is effective for each state of the shearer simulation data set And the overall evaluation effect is good, which can effectively solve the problem of unbalanced coal shearer state data set.

| number | C11  | C12  | C13  | C14  | ... | C42  | C43  | C44  | status |
|--------|------|------|------|------|-----|------|------|------|--------|
| 1      | 0.6177 | 0.4652 | 0.5068 | 0.5711 | 0.5844 | 0.5785 | 0.5402 | 3     |
| 2      | 0.8412 | 0.7312 | 0.9184 | 0.8269 | 0.754 | 0.8585 | 0.8412 | 1     |
| 3      | 0.9795 | 0.7044 | 0.8633 | 0.9066 | 0.8097 | 0.7656 | 0.9803 | 1     |
| 4      | 0.5286 | 0.4622 | 0.3936 | 0.4482 | 0.496 | 0.4132 | 0.4459 | 4     |
| 5      | 0.7335 | 0.7287 | 0.7225 | 0.7514 | 0.6686 | 0.6975 | 0.7702 | 2     |
| 7      | 0.6028 | 0.6252 | 0.6092 | 0.5076 | 0.4869 | 0.4721 | 0.5919 | 3     |
| 8      | 0.9037 | 0.8262 | 0.9255 | 0.8539 | 0.9258 | 0.7975 | 0.8123 | 1     |
| 996    | 0.8635 | 0.8743 | 0.9148 | 0.8426 | 0.9032 | 0.9397 | 0.9009 | 1     |
| 997    | 0.7402 | 0.6307 | 0.6338 | 0.6992 | 0.7513 | 0.664 | 0.7887 | 2     |
| 998    | 0.5023 | 0.5018 | 0.5019 | 0.5513 | 0.5709 | 0.3946 | 0.5595 | 4     |
| 999    | 0.7898 | 0.7192 | 0.7107 | 0.7687 | 0.6072 | 0.7517 | 0.7227 | 2     |
| 1000   | 0.8948 | 0.8285 | 0.9766 | 0.7743 | 0.8072 | 0.9436 | 0.7174 | 1     |
5 conclusion

In view of the strong data correlation between the multi-dimensional monitoring parameters of the shearer and the existence of data redundancy, the SP correlation coefficient combining the Spearman correlation coefficient and the Person correlation coefficient is proposed to analyze the correlation of the shearer parameters. Based on the construction of the shearer health evaluation system; in view of the unsatisfactory classification effect of a single classifier and the imbalance of category data in the data set of the shearer, the integrated learning method is introduced into the health evaluation work of the shearer and studied. Based on XGBoost-based shearer health evaluation method, and optimize its key parameters. Experimental results show that the average accuracy rate of the XGBoost-based health assessment method can reach 98.50%, the average recall rate is 98.38%, and the average F1 value can reach 97.61%, which can effectively solve the imbalance problem of the data set.

Acknowledgments This research was funded by the National Natural Science Foundation of China (grant number 51875451 and 51834006)

references

[1] Zhai Wenrui, Li Xiangong, Wang Jiaqi, et al. Research on performance degradation assessment method and application of shearer[J]. Industry and Mine Automation, 2020, 46(12):57-63+100.

[2] Cao Xiangang, Lei Yinan, Gong Yurong, et al. Research on the evaluation method of shearer health based on combined weighting method[J]. Coal Science and Technology, 2020, 48(06):135-141.

[3] Si Lei. Research on the key technology of shearer intelligent control[D]. Xuzhou: China University of Mining and Technology, 2015.

[4] Tobon-Mejia D A, Medjaher K, Zerhouni N, et al. A Data-Driven Failure Prognostics Method Based on Mixture of Gaussians Hidden Markov Models[J]. IEEE Transactions on Reliability, 2012, 61(2):491-503.

[5] Chen Faguang, Zhou Buxiang, Zeng Lanyu. Transformer operating state assessment model based on multi-information fusion[J]. Journal of Electric Power System and Automation, 2013, 25(04):140-144.

[6] Ali J B, Chebel-Morello B, Saidi L, et al. Accurate bearing remaining useful life prediction based on Weibull distribution and artificial neural network[J]. Mechanical Systems & Signal Processing, 2015, 56-57: 150-172.

[7] Li Peng, Li Dan, Li Xiwang, et al. Health Assessment of Power Dispatching Automation Equipment Based on Fuzzy Neural Network[J]. Computer System Applications, 2019, 28(02): 207-212.

[8] Xu Yan, Chen Xin. Transformer state assessment method based on cooperative game and cloud model[J]. Electric Power Automation Equipment, 2015, 35(3): 88-93.

[9] Mandal S, Maiti J. Risk analysis using FMEA: Fuzzy similarity value and possibility theory based approach [J]. Expert Systems with Applications. 2014, 41, 3527-3537.

[10] Merle G, Roussel M, Lesage J, et al. Quantitative Analysis of Dynamic Fault Trees Based on the Coupling of Structure Functions and Monte Carlo Simulation [J]. Quality & Reliability Engineering, 2014, 32(1): 7-18.

[11] J. Schmidhuber, Deep learning in neural networks: An overview. Neural Netw, vol. 61, pp. 85–117, Jan. 2015.

[12] Ge D, Yang Y. Reliability analysis of non-repairable systems modeled by dynamic fault trees with priority AND gates [J]. Applied Stochastic Models in Business & Industry, 2015, 31(6): 809-822.

[13] Chen Yanglong, Ma Yanheng, Yin Yuanwei. Equipment State Assessment Based on Convolutional Neural Network, Journal of Physics: Conference Series. 2019;40:48-52.

[14] LAl Yue-hua, DONG Hai-ping, CHENG Tao, et al. Reliability Estimation of Aviation Pyrotechnics System Based On Mixed Beta Distribution [J]. Journal of Donghua University (English Edition), 2014, 31(06): 785-788.

[15] Dong Xinghui, Zhang Xinniao, Zheng Kai, et al. Evaluation of wind turbine health status based on combined weighting and cloud model[J]. Acta Energia Sinica, 2018, 39(08): 2139-2146.
[16] Zhao Dongming, Cheng Yanming, Cao Ming. Health Assessment of Diesel Generators Based on Convolutional Neural Networks[J]. Computer Science, 2018, 45(S2): 152-154.

[17] CHEN T, GUESTRIN C. XGBoost: A scalable tree boosting system [C]. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, United States, 2016: 785-794.

[18] Li Jinbin, Xia Tian, Huang Shuo, et al. State evaluation of integrated disconnecting circuit breakers based on XGBoost[J]. High Voltage Technology, 2020, 46(05): 1800-1806.