ISSN: 2688-8653

Research on Flight Technology and Risk Assessment Based on Machine Learning

Yuqing Li[1], Hao Chen[2]

[1]School of Statistics and Applied Mathematics, Anhui University of Finance and Economics, Bengbu, China;

[2]School of Management Science and Engineering, Anhui University of Finance and Economics, Bengbu, China.

Abstract

With the development of the times and the continuous advancement of science and technology, the safety of aircraft has become more and more important in the civil aviation industry. Ensuring the safety of aircraft has become a top priority in the development of the current civil aviation industry. Therefore, the comprehensive use of existing data for scientific aircraft flight process management and monitoring of aircraft flight process is an effective means to reduce the occurrence of flight accidents. This article aims to use the machine learning model as the main body to establish aircraft flight technology and risk assessment models, so as to prevent possible flight accidents. First, use the 3 laws and IQR test data to eliminate outliers, and then use the method of mean value filling and interpolation filling to fill in the vacant values of the data, and then "landing G value, attitude (pitch angle), stick volume, disk volume " and other data columns are combined to increase the reliability and persuasiveness of the data. Finally, the key indicators related to flight safety are extracted through principal component analysis and their importance is explained.Secondly, the random forest algorithm and GA-BP neural network model are used at the same time, the flight parameters of the aircraft are used as the training set, and the pilot's flying skills are used as the test set. Through prediction, it is found that the pilot's flying skills can be evaluated by evaluating the flight parameters.Finally, establish an automatic early warning mechanism including the altitude early warning mechanism, the descent rate early warning mechanism, the No. 68 airport early warning mechanism, and the overrun early warning mechanism to prevent possible safety accidents and reduce the incidence of flight accidents.This paper provides a comprehensive and systematic flight safety solution by modeling and analyzing flight data. Using a variety of analysis methods, it provides targeted solutions and provides multiple early warning mechanisms to improve the technical level of flight personnel and reduce the incidence of flight accidents..

Keywords

3 σ laws ;IQR test; principal component analysis; random forest; GA-BP neural network.

1. Background

China's civil aviation passenger traffic is growing at a rate of more than 20% per year. According to data, in 2021, China's civil aviation passenger traffic will exceed 100 million for the first time, a historic breakthrough. Relevant departments have repeatedly issued policies to support development. The rapid development of China's civil aviation industry has brought about economic growth and safety problems. For example, the "3.21" air crash that occurred in 2022. Safety problems will occur during the flight, and once a safety accident occurs, the losses caused are irreparable. Therefore, the issue of flight safety is the primary issue we face. The cause of

ISSN: 2688-8653

the accident may be caused by various factors, including aircraft failure, airport and crew, etc. With the development of civil aviation and the advancement of science and technology, technical means can be used to monitor and control the safety of flight flights. In order to deal with irreparable losses, the data of the whole flight segment can be mined for analysis to form the flight quality record of specific personnel. By processing and modeling records of different flight crews, flight routes, airports and specific flight conditions, factors that affect flight safety can be better derived, and reasonable and effective measures can be taken to reduce the probability of accidents.

2. Key data extraction based on principal component analysis

2.1. Outlier processing

2.1.1. Law of 3 σ

The 3 σ law is also known as the 68-95-99.7 principle. In statistics, the 68-95-99.7 principle is in a normal distribution, and the distance from the mean is less than one standard deviation, two standard deviations, and three standard deviations. The more precise figures are 68.27%, 95.45% and 99.73%. The calculation formula is as follows:

$$\Pr(\mu\text{-}1\sigma \le X \le \mu\text{+}1\sigma) \approx 0.682689492137086 \tag{1}$$

$$\Pr(\mu-2\sigma \le X \le \mu+2\sigma) \approx 0.954499736103642 \tag{2}$$

$$\Pr(\mu-3\sigma \le X \le \mu+3\sigma) \approx 0.99730020393674 \tag{3}$$

where x is the observed value of a normally distributed random variable, μ is the mean of the distribution, and σ is the standard deviation.

The data collected for the agreed indicators are respectively $X_1, X_2, X_3 \ldots\ldots X_n$, then the arithmetic mean is:

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{4}$$

The standard deviation is:

$$\sigma = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2} \tag{5}$$

The given data is regarded as a random variable, and it is assumed that the given data obey a normal distribution $N(\mu, \sigma^2)$. According to the 3 σ principle, the probability of these random variables falling in the $(\mu-3\sigma, \mu+3\sigma)$ interval is 0.9973, and the probability of exceeding the $(\mu-3\sigma, \mu+3\sigma)$ interval is only $1-0.9973=0.0027<0.003$, that is, the average number of 1000 trials exceeding the $(\mu-3\sigma, \mu+3\sigma)$ interval is less than 3 times. This is a small probability event, which generally does not happen in one experiment. . Therefore, it can be considered that in subsequent tests, if the data exceeds the $(\mu-3\sigma, \mu+3\sigma)$ interval, it is abnormal data.

ISSN: 2688-8653

2.1.2. IQR TEST

IQR, also known as the interquartile range, is a robust statistical technique used to represent a quantity of data dispersion and is often used to check for abnormalities in data. Quartiles divide the data into 4 parts, namely the upper quartile, median, and lower quartile. Its formula is as follows:

$$Q_1 = \frac{N+1}{4} \tag{6}$$

$$Q_2 = \frac{2(N+1)}{4} \tag{7}$$

$$Q_3 = \frac{3(N+1)}{4} \tag{8}$$

Among them, Q_1 is the lower quartile, indicating that the data below this value accounts for 25% of the total; Q_2 is the median, indicating that the data below this value accounts for 5% of the total; Q_3 is the upper quartile, indicating that the data below this value accounts for 5% of the total; Data below the numerical value account for 75% of the total; is the number of results.

The interquartile range IQR is expressed as: $IQR = Q_3 - Q_1$

The outlier detection interval under the IQR criterion is: $[Q_1 - 1.5IQR, Q_2 + 1.5IQR]$, and the data beyond this range will be eliminated.

2.1.3. Blank value processing

After eliminating the abnormal data through the 3σ law and the IQR criterion, fill the vacant data with interpolation method.

Figure 1 is a heat map of the correlation between each data column after removing outliers.

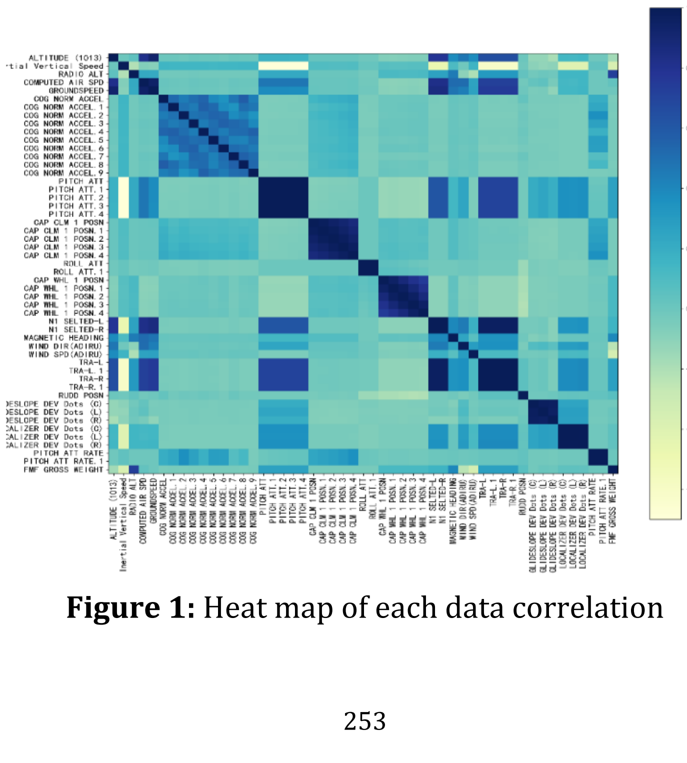

Figure 1: Heat map of each data correlation

2.2. Principal component analysis method to extract key data

In flight safety issues, there are many influencing factors related to it. If multiple factors are considered in the analysis of the problem, it will cause a large amount of overlap of information and make the problem more complicated. Therefore, this paper uses principal component analysis to explore the factors that are highly correlated with flight safety issues.

Principal component analysis is a statistical method that converts multiple factors into a few important influencing factors. Its main function is to achieve data dimensionality reduction and retain the efficient contribution rate of original data. The specific steps of principal component analysis are as follows:

(1) Raw data standardization

In a sample set, there are j random variables, and each random variable contains i data. Normalize X_{ij} to remove dimension effects. Make calculation more convenient. The normalization formula is as follows:

$$x^*_{ij} = \frac{x_{ij} - \bar{x}_j}{\sqrt{D(X_j)}} (i = 1, 2, \ldots, n : j = 1, 2, \ldots, p) \tag{9}$$

$$\bar{X}^*_j = \frac{1}{n}\sum_{i=1}^{n} x^*_{ij} = 0(\text{j=1, 2, . . . , p})$$
$$D(X^*_j) = \frac{1}{n}\sum_{i=1}^{n}(x^*_{ij} - \bar{X}^*_j)^2 = 1(j = 1, 2, \ldots, p) \tag{10}$$

Among them, $\bar{X}_j$、$\sqrt{\mathrm{D}(X_j)}$ are the mean value and standard deviation of the jth variable, respectively.

(2) Calculate the eigenvalues and corresponding eigenvectors of the correlation coefficient matrix.

The eigenvalue vector corresponding to the correlation coefficient matrix is a linear conversion coefficient, and the eigenvalue is the variance of the main component, and its variance contribution rate decreases with the order of the eigenvalues.

Construct the correlation coefficient matrix R of X as:

$$R = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1p} \\ r_{21} & r_{22} & \cdots & r_{2p} \\ \cdots & \cdots & \cdots & \cdots \\ r_{pl} & r_{p2} & \cdots & r_{pp} \end{bmatrix} \tag{11}$$

$$r_{ij} = \frac{Cov(X_i, X_j)}{\sqrt{DX_i}\sqrt{DX_j}} (i, j = 1, 2, \ldots, p) \tag{12}$$

Get the characteristic root arrangement of R:

$$\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p \geq 0 \tag{13}$$

ISSN: 2688-8653

Its unit eigenvector is:

$$a_1=\begin{bmatrix}a_{11}\\a_{21}\\\cdots\\a_{p1}\end{bmatrix},a_2=\begin{bmatrix}a_{12}\\a_{22}\\\cdots\\a_{p2}\end{bmatrix},\ldots,a_p=\begin{bmatrix}a_{1p}\\a_{2p}\\\cdots\\a_{pp}\end{bmatrix} \tag{14}$$

The expression of the i-th principal component F_i is:

$$F_i=a_{1i}X_1+\ldots+a_{pi}X_p(i=1,2,\ldots,p) \tag{15}$$

(3) Number of principal components

The principal component is selected according to the cumulative variance ratio of the principal component, and the contribution rate is:

$$\eta_i=\frac{\lambda_i}{\sum_{k=1}^{p}\lambda_k}(i=1,2,\ldots,p) \tag{16}$$

The sum of all principal component contributions:

$$\sum_{i=1}^{p}\eta_i=1 \tag{17}$$

Record the sum of the contribution rates of the first e principal components:

$$M_e=\frac{\sum_{i=1}^{e}\lambda_i}{\sum_{k=1}^{p}\lambda_k}(e<p) \tag{18}$$

When M_e exceeds 85%, it is considered that the principal component can represent the information X of the data, satisfy the data will be, and reduce the loss of data at the same time.

(4) Result analysis

According to the steps of principal component analysis, the contribution rate of each indicator is obtained, among which "ALTITUDE (1013)" has the highest weight, which is 30.97%; "MAGNETIC HEADING" is 3.03%; the total contribution rate of the first nine indicators More than 85%. According to the definition of principal component analysis, these nine indicators are strongly related to flight safety and are some key data items related to flight safety. Figure 2 is the weight map of principal component analysis.

ISSN: 2688-8653

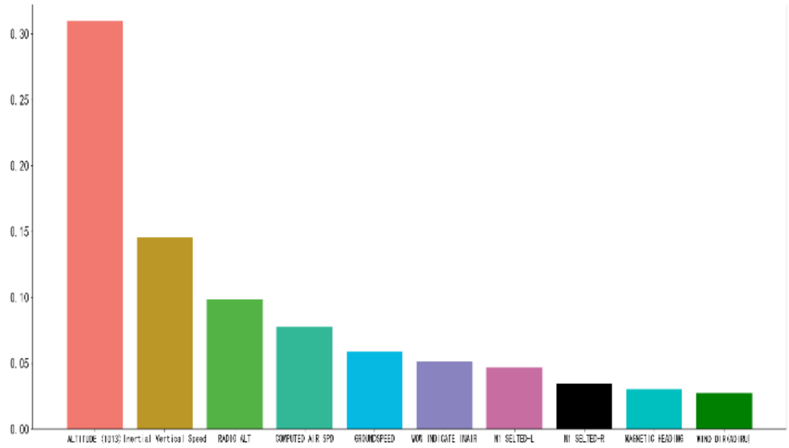

Figure 2 : Weights of principal component analysis

3. Evaluation of flight techniques based on random forest and GA-BP

3.1. Classification algorithm based on random forest

Random Forest is a classification and regression tree technique. Randomly and iteratively sample data and variables to produce a large set of classification and regression trees or forests. A random forest consists of multiple decision trees, each of which is also different. When building a decision tree, since it is not known which part of the data has abnormal samples, it is also impossible to determine which features can determine the classification result, so there must be a choice to put it back.

The random forest prediction steps are as follows:

(1) Randomly select a part of samples from the training data with replacement to form a new training set containing n samples for training, and use n samples to train a decision tree as the sample at the root node of the decision tree. Repeat t times to get t training set.

(2) When the M attributes of each sample need to be split at each node of the decision tree, m attributes are randomly selected from the M attributes, and the condition $m << M$ needs to be met. Then a certain strategy is adopted from the m attributes to select 1 attribute as the splitting attribute of the node.

(3) (3) Repeat step 2, the formation of each node in the decision tree formation process is split according to step 2. Then the parameters corresponding to the unselected features are eliminated, and then a decision tree is constructed for the new training set without pruning. Each decision tree will also have different training results due to different samples and features. Its classifier is used to define the expected number of categories k and the number m of predictor variables for each node to grow the tree to generate a predictive model.

The final classification result of the system adopts the simple majority voting method, and the final classification decision is:

$$H(x) = \arg\max \sum_{i=1}^{k} I(h_i(x) = Y) \tag{19}$$

Among them, $H(x)$ represents the combined classification model, h_i is a single decision tree classification model, Y represents the output variable, and the function $I(Y)$ is an indicative function, that is, the final classification is determined by the way of majority voting.

The decision tree error curve is shown in Figure 3 below:

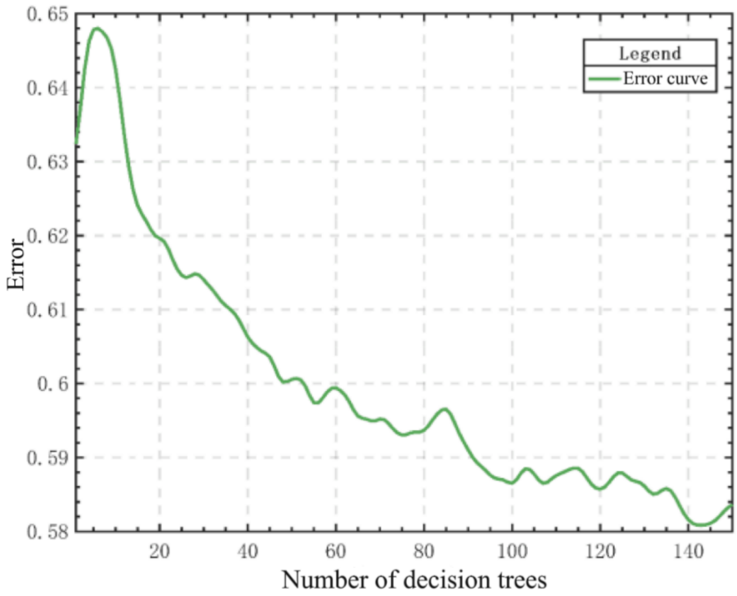

Figure 3: Decision tree error curve

It can be seen from the above figure that the more the number of decision trees, the less error. 20% of the data is used as the training set, and the confusion matrix of the training set is shown in Figure 4. It can be seen from the figure that the model has good adaptability and high accuracy.

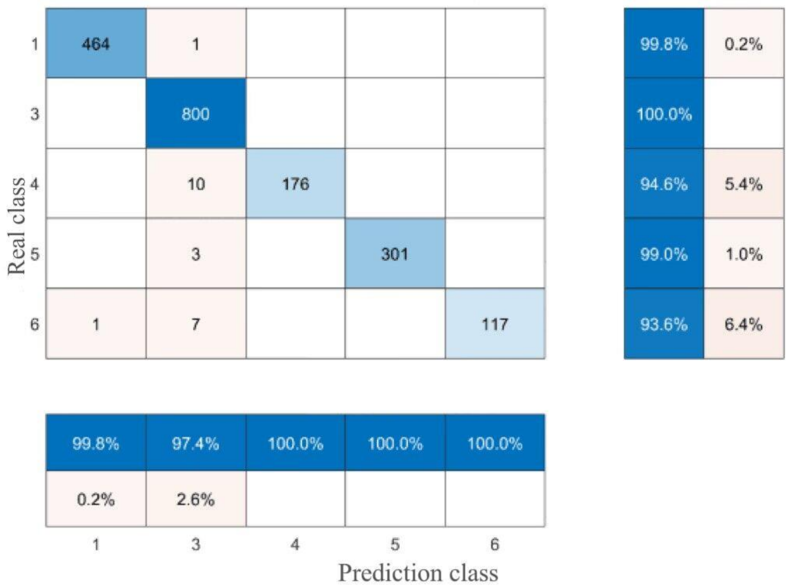

Figure 4: Training set of GA-BP model

The prediction results of the random forest are shown in Figure 5 below:

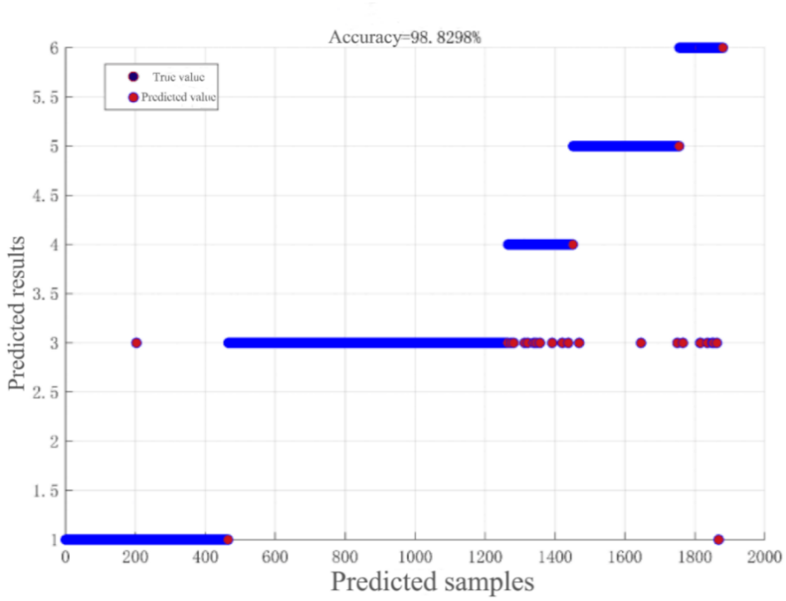

Figure 5 : Test set of GA-BP model

As can be seen from the above figure, the comparison accuracy rate of the training set prediction results is 98.8298%

3.2. Neural Network Prediction Model Based on GA-BP

The neural network algorithm is essentially an error echelon descent algorithm, but there are still some defects in some complex nonlinear optimization problems. The neural network is only optimizing locally, training the samples one by one, but not performing global optimization on all the data before training, so the convergence speed is slow. At the same time, due to the shortcomings of BP neural network, such as low learning efficiency and prone to local minima.

Genetic Algorithm, referred to as GA, always searches the entire space, and it is easy to obtain the global optimal solution. The GA genetic algorithm just makes up for the shortcomings of the BP neural network that cannot be optimized globally. At the same time, on the one hand, since the initial weight of the BP neural network is randomly given, the number of times of each training and the final weight are different, resulting in the non-uniqueness of optimization; on the other hand, the random given The initial weights lead to more training times and slower convergence. The GA genetic algorithm can roughly search out a certain range of weights as the initial weights of the BP neural network, and then solve the problems that the BP neural network is prone to local minimum and slow convergence speed.

In this paper, the genetic algorithm is used to optimize the BP neural network to form a hybrid algorithm GA-BP algorithm to improve the accuracy of model prediction. A three-layer BP neural network is used for prediction, that is, the input layer, the hidden layer and the output layer. There is no connection between the neurons in the layer, and the inter-layer neurons are connected to each other. Therefore, this paper combines GA genetic algorithm and BP neural network, uses genetic algorithm to train and optimize the initial value, uses genetic algorithm to perform global optimization, narrows the range of initial value, and then uses BP neural network to train and obtain accurate The initial weight value, and finally use the generalization ability of the BP neural network to predict the input value.

3.2.1. Traditional BP neural network model

The traditional BP neural network prediction principle is shown in Figure 6:

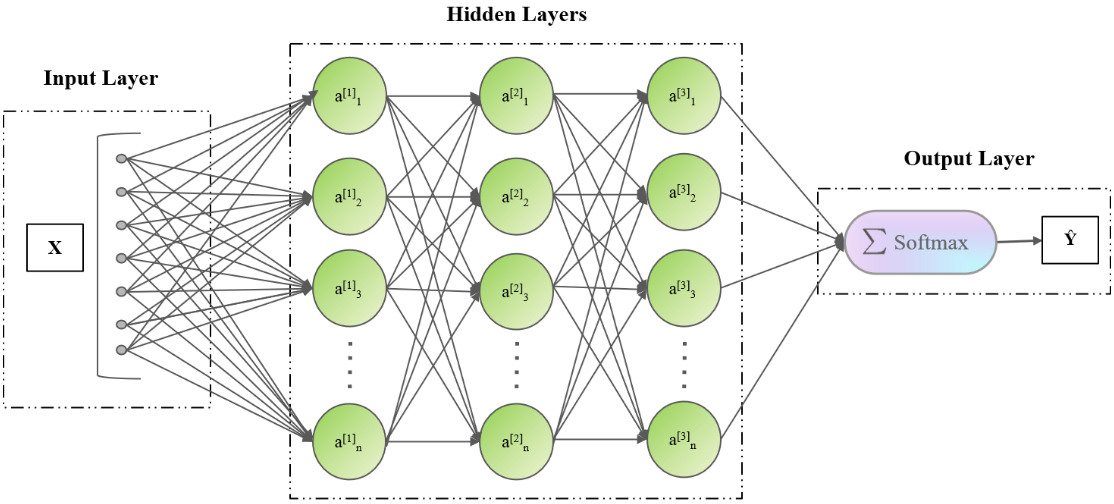

Figure 6: BP neural network prediction principle

According to the above schematic diagram, a BP neural network prediction model for evaluating aircraft flight technology is constructed: the flight parameters are used as the input layer $X(x_{1k}, x_{2k} \cdots x_{nk})$; the pilot's technical level is used as the output Y. Use a large number of sample data to train the network, so that different input volumes can get different output volumes. When the error between the output value and the teacher signal is less than a certain allowable value, the training ends.

3.2.2. Establishment of GA-BP neural network prediction model

Combining the genetic algorithm with the BP neural network, first use the genetic algorithm to optimize the training, narrow the search range of the initial value, and then combine the genetic algorithm and the gradient descent-based introspective propagation training method of the neural network to form a hybrid training method.

The flowchart of GA-BP neural network prediction is shown in Figure 7:

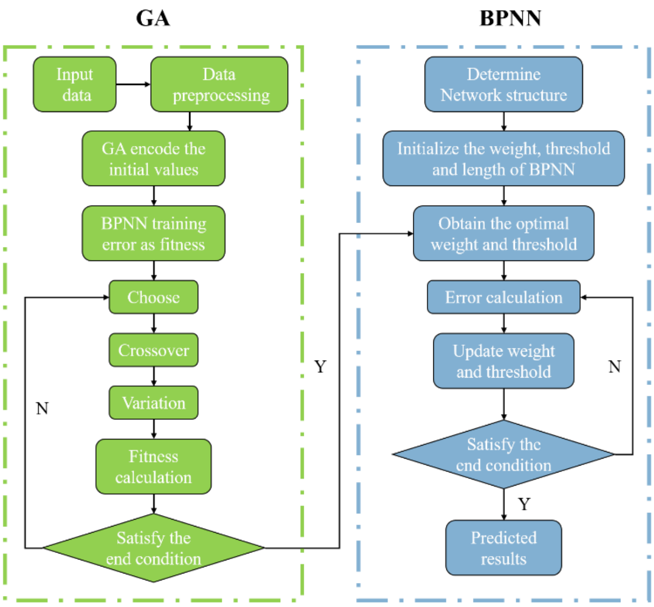

Figure 7: Flow chart of GA-BP model

The specific steps of GA optimizing the BP neural network prediction model are as follows:

(1) Coding scheme.

In order to obtain high-precision weights and thresholds, this paper uses a binary encoding method. When encoding, first assume that all weights and thresholds are within a certain range, $W_{ij} \in (W_{\min}, W_{\max})$, the weights and thresholds are represented by Q bit binary numbers, and the difference between the actual threshold (or weight) and the value represented by a binary string The relationship is as follows:

$$W_{ij} = (W_{ij})_{\min} + \frac{binreplace(Q)}{2^Q}[(W_{\max}) - (W_{\min}) + 1] \tag{20}$$

Among them, W_{ij} is the actual weight, $binreplace(Q)$ is a binary number represented by Q bit string, and $(W_{\min}, W_{\max})$ is the range of continuous weights. Then all the weights and the 0/1 codes corresponding to the thresholds are connected in series to form a weight combination of the network.

(2) Generate initial population

Let the population size be P. Randomly generate the initial population $X = (X_1, X_2, \cdots, X_P)^T$ of P individuals, given a selected range of data, use a linear interpolation function to generate a real vector $x_1, x_2, \cdots, x_S$ of individuals X_i in the population as a chromosome of GA, representing the initial weight and threshold distribution of a neural network, then the individual The length of is the sum of the number of weights and the number of thresholds of the neural network, namely:

$$n = r \times s_1 + s_1 \times s_2 + s_1 + s_2 \tag{21}$$

Among them, r is the number of input layers, s_1 is the number of hidden layers, and s_2 is the number of output layers.

(3) Determine the fitness function of the individual

Evaluate the individual according to the fitness, decode the chromosome obtained in (1), assign the initial weight and threshold of the BP neural network, input the training samples for neural network training, calculate the error value of the network, and reach the set value Accuracy obtains the network training output value, and then constructs the fitness function. For the minimum value, the formula can be used:

$$f(x) = \frac{1}{\frac{1}{2}\sum_{K=1}^{P}(y_k - \hat{y}_k)^2 + 1} \tag{22}$$

Among them, $\frac{1}{2}\sum_{K=1}^{P}(y_k - \hat{y}_k)^2$ is the output error of the neural network, and $y_k, \hat{y}_k (k = 1, 2, \cdots, P)$ are respectively the expected output and the actual output of the network.

According to the fitness function, the fitness value of each individual is calculated, and then the individuals are sorted according to the size of the fitness value. The larger the fitness value, the stronger the adaptability of the individual.

The figure below shows the fitness curve:

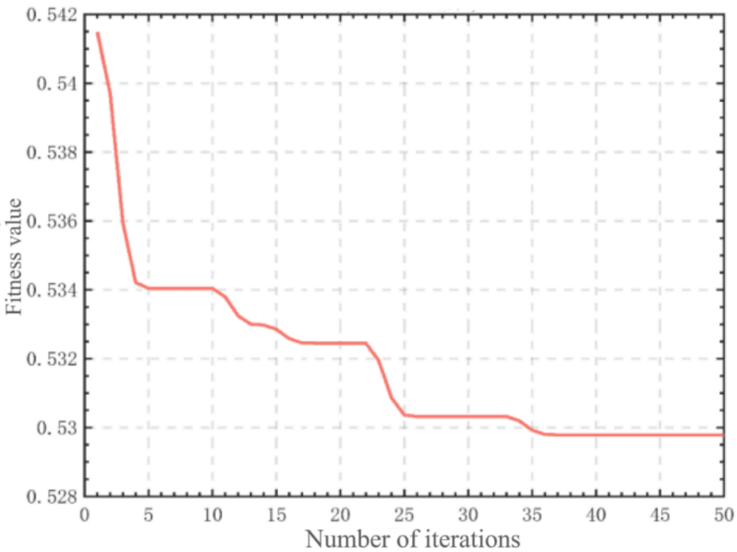

Figure 8: Fitness change curve

(4) choose

The fitness ratio method is used to select the operator, that is, each given selection probability is proportional to its corresponding fitness, and the selection probability is:

$$p_i = \frac{f_i}{\sum_{i=1}^{P} f_i}, i = 1, 2, \cdots, P. \tag{23}$$

Among them, f_i is the fitness value of the i-th individual, and P is the population size. In actual learning, the individual with the largest fitness function value is directly assigned to the next generation.

(5) cross

Randomly select two weight individuals from the population according to the crossover probability, and then randomly set an intersection point in the individual character string, and the crossover operation of the kth gene x_k and the l-th gene x_l of the intersection point at the j position is respectively for:

$$\begin{aligned} x_{kj} &= x_{kj}(1-b) + x_{lj}b, \\ x_{lj} &= x_{lj}(1-b) + x_{kj}b \end{aligned} \tag{24}$$

Among them, b is a random number between [0,1].

(6) Mutation operation

Randomly select the jth gene of the i-th individual from the population with a certain probability for mutation operation, namely:

$$w = \begin{cases} w_{ij} + (w_{ij} - w_{\max}) f(g), r \geq 0.5 \\ w_{ij} + (w_{\min} - w_{ij}) f(g), r < 0.5 \end{cases} \tag{25}$$

$$f(g) = r_2 (1 - \frac{g}{G_{\max}}) \tag{26}$$

Among them, $w_{\max}$和$w_{\min}$ are the upper and lower bounds of the gene value, r is a random number between [0,1], r_2 is a random number, g is the current iteration number, and $G_{\max}$ is the maximum evolution algebra. Variation reflects the gene mutation phenomenon of the occasional human in biological inheritance, so its probability is generally very small.

(7) Repeat steps (3) to (6)

The initial weight and threshold distribution are continuously modified and evolved, and then the optimal individual obtained by GA is decomposed into the connection weight and threshold of the BP neural network, which are used as the initial value and threshold of the BP neural network prediction model, and then the BP neural network The network prediction model is trained to output the predicted global optimal solution.

3.2.3. Solution of GA-BP neural network prediction model

The aircraft parameter measurement data in Annex III is used to predict samples, wherein 80% of the original data are selected as training samples, and the remaining 20% are used as test samples, that is, to predict the flight technology of the aircraft.

As shown in Figure 9, it is the confusion matrix of the training set.

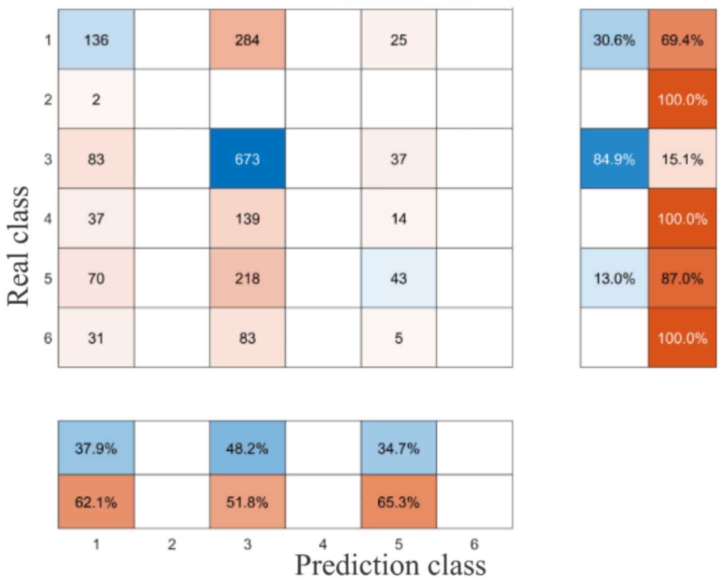

Figure 9: Confusion matrix of GA-BP model training set

After training, the test sample is predicted by the network that has learned about the "pilot qualification". As shown in Figure 10, it is the confusion matrix of the test set.

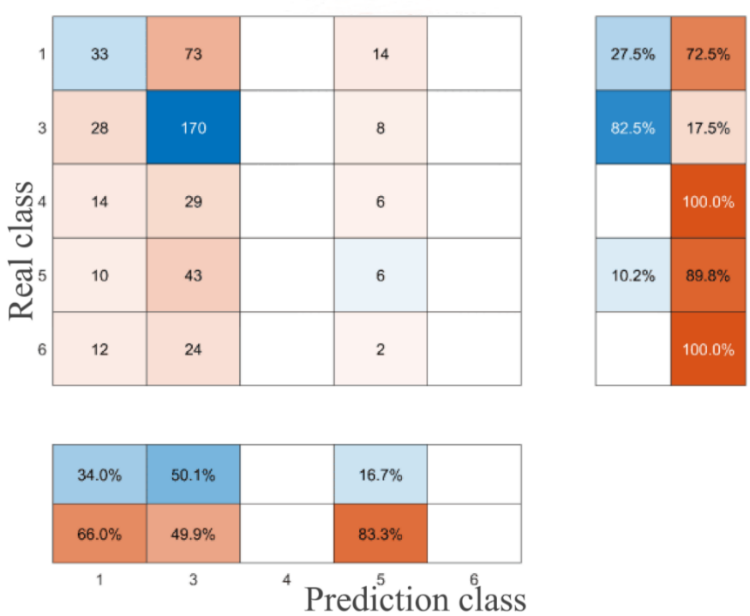

Figure 10: Confusion matrix of GA-BP model test set

Through two prediction algorithms, the given 116 columns of flight parameters are reduced to 8 dimensions and then the model is constructed for analysis. From the result graph, we can see that the prediction accuracy rate is as high as 98.83%. Pilots with "different qualifications" are highly consistent with their aircraft flight status during flight operations, indicating that this evaluation of pilots' flying skills based on flight parameters is reliable.

4. Establish an automated early warning mechanism

As a safety manager of an airline, through solving the above four problems, an automatic early warning mechanism is established to prevent possible safety accidents and provide guarantee for the safe operation of the aircraft. The following is the content of the automatic early warning mechanism.

4.1. Altitude warning mechanism

Through the extraction of nine indicators that are strongly related to aircraft safety in a pair of questions, it is found that ALTITUDE (1013), that is, the altitude has the greatest impact on aircraft safety. At the same time, combined with the data in Appendix 1, when the altitude of the aircraft exceeds 20,000 feet, the aircraft may have problems such as a significant reduction in the available power of the engine and a significant deterioration in maneuverability. Therefore, an altitude warning mechanism needs to be established. When the altitude of the flying aircraft exceeds 20,000 feet, the alarm system on the aircraft will remind the pilot that the altitude is higher.

4.2. Early warning mechanism for decline rate

It can be seen from the principal component analysis method that the aircraft descent rate has a relatively large impact on aircraft safety. When the descent rate of the aircraft exceeds 1800 feet, accidents such as aircraft crashes will occur. Therefore, an early warning mechanism for aircraft descent rate should be established. When the aircraft's descent rate exceeds 1,800 feet, the warning system will remind the pilot to slow down the descent speed.

ISSN: 2688-8653

4.3. Overrun early warning mechanism

Through the analysis of the data in Attachment 2, it is found that Aircraft No. 26 flying on the "Airport 68-Airport 92" route is most likely to exceed the limit of the distance from 50 feet to touchdown with a warning level of 2 in the state of landing. Therefore, it is necessary to establish an overrun early warning mechanism. When aircraft No. 26 takes off from Airport No. 68 and lands at Airport No. 92, especially when the aircraft is in the landing state, the pilot must pay attention to whether the aircraft will exceed the limit. , to ensure that the aircraft reaches its destination safely.

5. Model Evaluation

5.1. Advantages of the model

(1) Random forest has high accuracy, can reduce over-fitting problems, and can effectively reduce the influence of noise and outliers on the model. By reducing the dimensionality of high-dimensional data, randomly selecting samples with replacement to build a decision tree can avoid the impact of strong correlation between features. Random forest is more robust to noise, missing values, and wrong values in the data, has strong robustness, and is not easy to overfit.

(2) When using the traditional neural network model, we need to set the hyperparameters of the network structure, number of layers, number of nodes, etc., and for different data samples, repeated experiments are required to bring the optimal Model. However, the genetic algorithm-based neural network can find the most suitable network parameters and hyperparameters through automatic search and optimization, and can also make the network more suitable for various data sets and improve the applicability of the model. In the traditional neural network, it will fall into the deadlock of the local optimal solution, but when combined with the genetic algorithm, it can jump out of the local optimal solution, and then continue to search for the global optimal solution.

5.2. Disadvantages of the model

(1) In the principal component analysis, it is necessary to ensure that the amount of information after dimensionality reduction is kept at a high level, and the extracted principal components must have an explanation that can give the actual background and meaning. Moreover, the interpretation meaning of the principal components after dimensionality reduction has a certain degree of ambiguity, which cannot be as clear as the meaning of the original variables.

(2) In the neural network algorithm based on the genetic algorithm, because the genetic algorithm will search for different global optimal solutions under different initial conditions, the result is uncertain, and a large number of experiments are required, which will also cause high calculation costs.

6. Model improvement and promotion

6.1. Model Improvements

During machine learning training, the computational resource consumption of models is an important issue, especially when the models become more complex and large. In order to solve this problem, optimization algorithms such as gradient descent and regularization algorithms should be adopted in future research to improve the training speed and memory utilization, thereby reducing the resource consumption of the model.

6.2. Generalization of the model

Dimension reduction processing of huge data to study its relevance and other issues can also be applied to the research of other vehicles, such as high-speed rail and other road vehicles, and

ISSN: 2688-8653

can also be used to evaluate the technical problems of its drivers. In the previous topic, the flight parameters are highly consistent with the pilot's grade evaluation. We can further explore and understand how to build a complete evaluation system to evaluate the pilot's grade through flight parameters. The actual research in this paper is to evaluate and predict flight parameters accordingly. When solving similar problems, a model that meets the given conditions of the problem can be reconstructed through relationship analysis and data processing, and a reasonable evaluation and optimization can be performed through algorithms.

References

[1] Zhao Hesong, Wang Yuanyuan, Sun Aimin. A water intake anomaly detection method based on ARIMA model and 3σ criterion [J]. Water Conservancy Informationization, 2022.

[2] Yang Keke, Liu Lilong, Chen Jun. Ionospheric TEC anomalies during earthquakes based on sliding interquartile range method [J]. Journal of Guilin University of Technology, 2019.

[3] Wang Yuan, He Ying, Wu Zhenyu. Back analysis method of unsaturated seepage parameters based on principal component analysis [J]. Water Conservancy Planning and Design, 2023.

[4] Yu Yufeng, Zhu Yuelong, Wan Dingsheng, Guan Xingzhong. Anomaly Detection of Hydrological Time Series Based on Sliding Window Forecasting [J]. Computer Applications, 2014.

[5] Yang Fan, Feng Xiang, Ruan Ling, etc. Research on the correlation between water branches and ultra-low frequency dielectric loss based on the Pearson correlation coefficient method, 2014.

[6] Nie Yunli, Gao Guozhong. Shale gas "sweet spot" classification method based on random forest [J/OL]. Oil and Gas Reservoir Evaluation and Development: 1-15 [2023-04-15] 2023.

[7] Yi Huangzhi, Gao Fei. Subway deformation prediction model based on GA-BP neural network [J]. Journal of Hefei University of Technology (Natural Science Edition), 2021.

[8] Sun Keda. Research on GA-based BP Neural Network Road Traffic Accident Prediction Model [D]. Zhejiang University of Technology, 2008.

[9] Li Song, Liu Lijun, Xie Yongle. Genetic Algorithm Optimizing BP Neural Network for Short-term Traffic Flow Chaotic Prediction [J]. Control and Decision Making, 2011.