

Research of Flight Safety and Technical Evaluation

Xiaonan Bo ^{1,*,#}, Chen Cui ^{2,#}, Jiancheng Lv ^{1,#}

¹ Department of Physics and Electronic Information, Henan Polytechnic University, Jiaozuo, Henan, China, 454003

² School of Software, Henan Polytechnic University, Jiaozuo, Henan, China, 454003

* Corresponding Author Email: 15138617651@163.com

[#]These authors contributed equally.

Abstract. Flight safety is the foundation of survival and development of civil aviation transport industry. Aircraft flight safety can be checked by aircraft flight parameters. We discuss a flight technology evaluation method based on flight parameter data. Firstly, data preprocessing is carried out, and the data is reduced by PCA dimension reduction. Then, BP neural network and decision tree model are established to train and test the data set. Neural network algorithm and machine learning algorithm are used for training. The results show that the BP neural network verification set is more accurate. The comparison table between the predicted results and the real results is shown in Table 3. Among the 15 samples, 14 are correct. The prediction accuracy reaches 93%..

Keywords: Flight safety, BP neural network, Decision tree, PCA dimension reduction.

1. Introduction

With the rapid development of civil aviation industry, it is more and more important to study flight safety. Rapid access recorder (QAR) [1] mainly records the flight parameters of an aircraft during flight, which is the most important data in aviation safety big data. The flight safety of aircraft is checked by the safety check of airline operation or by flight parameters. We are going to record a lot of flight parameter data in a route. A lot of data is not easy to analyze and evaluate, so it is very necessary for flight technical evaluation. It is conducive to monitoring and warning risks, reducing the probability of flight accidents, and preventing serious flight accidents caused by potential hazards.

Hao Shuai et al [2] carried out a research on flight safety interval method based on ADS-B in 2010. Liu Jiangping [3] carried out a BP neural network-based flight safety study of Hainan Airlines in 2014. Wen Chao et al [4] launched an aviation safety management evaluation system based on BP neural network in 2020. Gong Yanbin et al [5] proposed a research modeling method of aircraft fault diagnosis based on decision tree hierarchy in 2023. We are able to take some inspiration from their research.

We propose a flight technology evaluation method based on flight parameters. Firstly, we find some data on flight parameters. Due to excessive data, data preprocessing needs to be carried out. After the preliminary processing, there is still a lot of data, so PCA dimensionality reduction is used to conduct PCA training on the data, reducing the amount of data processing. Secondly, BP neural network and decision tree model are established to train the data set. Flight parameters are taken as independent variables, pilot qualifications as dependent variables. Finally, according to the training results of the model, flight technology evaluation and analysis are carried out.

2. Model assumptions and notation

2.1. Assumptions

- (1) It is assumed that the small amount of data is also representative [6].
- (2) It is assumed that the results of multiple tests are independent when the data is trained.
- (3) Assume that all data measurements are true and valid, without considering the effects of human error [7].

(4) Suppose that the technical evaluation of flight personnel does not consider the ranks of missing values greater than 15 percent.

2.2. Data preprocessing

Important notations used in this paper are listed in Table 1.

Table 1. Notations

Symbols	Meaning
b_j	Rate of contribution
e_i	Effective basis number
α_p	Accumulating contribution rate
$\widetilde{a}_{ij}$	Standardized data value
μ_j	Sample average
s_j	Sample standard deviation
λ_j	Eigenvalue
c_i	Different data types
$x_1, x_2, \dots, x_{100}$	100 flight parameters
W_{ij}	The value of a layer of neurons
R	Correlation coefficient matrix

3. Evaluation of flight technology based on flight parameters

3.1. Data preprocessing

3.1.1 Missing value processing

There are more than 2300 lines of data in this study, so these data need to be preprocessed. First, we need to delete the columns with more than 15% missing values. More than 200 lines of data in this article are found through the program search. The attachment still contains more than 2,100 lines of data after deleting these data. So the evaluation results will not be significantly affected. In fact, large amounts of data are inconvenient for evaluation and analysis. Then the data volume was reduced by PCA dimension reduction.

3.1.2 PCA dimensional reduction

Principal component analysis (principal the basic principle of component analysis) is to analyze the existing variable data and extract the larger information. It is an effective mathematical method for dimensionality reduction in multivariate statistics [8].

We use $(x_1, x_2, \dots, x_{100})$ respectively to represent 100 flight parameters and use $(i = 1, 2, 3, 4)$ respectively to represent the data situation of 4 different flight personnel. We construct the matrix $A = (a_{ij})_{4 \times 100}$. Next, we introduce the process and steps of PCA dimension reduction method.

(1) The first step is to standardize the original data. All the index values need to be converted to normalized data values $\widetilde{a}_{ij}$.

$$\widetilde{a}_{ij} = \frac{a_{ij} - \mu_j}{s_j}, i = 1, 2, \dots, 4; j = 1, 2, \dots, 100 \quad (1)$$

Meanwhile, the formulas for calculating μ_j and s_j are given as follows.

$$\begin{cases} \mu_j = \frac{1}{4} \sum_{i=1}^4 a_{ij} \\ s_j = \sqrt{\frac{1}{4-1} \sum_{i=1}^4 (a_{ij} - \mu_j)^2} \end{cases} \quad (2)$$

In fact, μ_j and s_j are sample mean and sample standard deviation. The formula for calculating standardized index variables is as follows accordingly.

$$\tilde{X}_j = \frac{X_j - \mu_j}{s_j}, j = 1, 2, \dots, 100 \quad (3)$$

(2) The coefficient matrix R is ought to be calculated.

$$Y_{ij} = \frac{\sum_{k=1}^{100} \tilde{a}_{ki} \cdot \tilde{a}_{kj}}{100-1}, i, j = 1, 2, \dots, 100 \quad (4)$$

(3) The eigenvalues and eigenvectors need to be evaluated.

$$\begin{cases} y_1 = u_{11}\tilde{x}_1 + u_{21}\tilde{x}_2 + \dots + u_{201}\tilde{x}_{20} \\ y_2 = u_{12}\tilde{x}_1 + u_{22}\tilde{x}_2 + \dots + u_{202}\tilde{x}_{20} \\ \vdots \\ y_{100} = u_{1100}\tilde{x}_1 + u_{2100}\tilde{x}_2 + \dots + u_{201000}\tilde{x}_{100} \end{cases} \quad (5)$$

(4) The most important thing in PCA is to calculate the eigenvectors [9]. We begin to calculate the aggregate valuation. We need to calculate the relevant coefficient matrix R eigenvalue λ_j and its corresponding eigenvector size. If the solution is a general type of eigenvector, the eigenvector will be obtained by combining the eigenvector together and constituted 100 new index variables.

$$\begin{cases} b_j = \frac{\lambda_j}{\sum_{k=1}^{100} \lambda_k}, j = 1, 2, \dots, 100 \\ \alpha_p = \frac{\sum_{k=1}^p \lambda_k}{\sum_{k=1}^{20} \lambda_k} \end{cases} \quad (6)$$

(5) Finally, the principal component was selected, the principal component load was calculated, and the comprehensive analysis was carried out [10].

$$z = \sum_{j=1}^p b_j y_j \quad (7)$$

3.1.3 PCA dimension reduction results

According to the above PCA dimension reduction principle, we draw the cumulative contribution rate table related to dimension reduction and calculate the cumulative contribution rate for each row of the data related to flight parameters. The results are shown in Table 2.

Table 2. Cumulative contribution rate calculation table

serial number	rate of contribution	cumulative contribution rate
1	23.80%	23.80%
2	17.50%	41.30%
3	15.40%	56.7%
4	14.50%	71.20%
5	11.50%	82.70%
6	6.20%	88.90%
7	3.10%	92.00%
8	3.30%	95.30%
9	3.20%	99.21%

According to the above analysis process, the cumulative contribution rate of the rubble graph was drawn. We see that the accuracy of the result reached 99% from the figure below when the 9th line of data was used. Therefore, these valid data can be used to calculate in the following calculation.

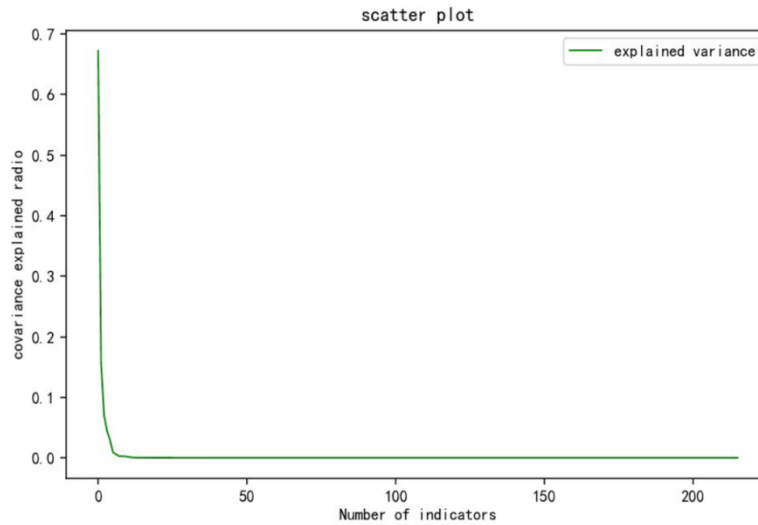


Figure 1. Aggregate contribution rate of PCA

3.2. Establishment of BP neural network model and decision tree model

We establish BP neural network model and decision tree model in order to ensure flight safety and evaluate pilots' flight techniques. The specific process is as follows.

3.2.1 Establishment of BP neural network model

BP neural network is a kind of multi-layer back propagation network structure, which can transmit data forward and error backward. After the error is constantly optimized, the output result is obtained after the minimum error [11]. Its learning algorithm is called inverse error transmission algorithm. It can be approximated by using a hidden layer of BP network for any continuous function in the closed interval, because a three-layer BP network can be used to complete the n dimension to m dimension mapping [12]. So the BP neural network adopted in this paper contains only one input layer, one hidden layer and one output layer [13]. Here are the steps for model building.

(1) The input and output of the first layer are as follows.

We choose the excitation function as piecewise linear function in order to determine the excitation function. The input and output we get at the first layer are samples of the actual input and output. Hence we get the formula for the linear function.

$$\varphi(v) = \begin{cases} 1 & v \geq 1 \\ \frac{1}{2}(1+v) & -1 < v < 1 \\ 0 & v \leq -1 \end{cases} \quad (8)$$

(2) The input and output of the second layer are as follows.

According to the principle of BP neural network, the sum of the value and threshold of the neurons of the latter layer whose input and output come from the former layer can be calculated by the following formula.

$$I_2 = W_{ij} \cdot X + B_{ij} \cdot ones \quad (9)$$

When the output of the second layer is made, it is assumed that the excitation function of the second layer is of type S, and the step excitation function is avoided here. Thus, the output of the second layer is as follows.

$$O_2 = \frac{1}{1+e^{-I_2}} \quad (10)$$

(3) The input and output of the third layer are as follows.

It can be seen from the above that the input of the latter layer comes from the sum of the value and threshold of the neurons of the previous layer. According to experience, the excitation function of the third layer is also a linear function, so the output matrix is $O_3 = I_3$.

3.2.2 BP neural network model test

(1) Determine the number of neurons and model training methods

First, the raw data was divided into 60% Training set, 20% Validation set and 20% Testing set [14]. Then, the number of hidden neurons and the training method of the model should be selected to complete the training of the corresponding model. The selection of the number of neurons determines the training effect to some extent. After a lot of attempts, the number of hidden neurons was finally selected as 4, and the model training method was Bayesian regularization.

(2) Examination of the results of neuron training models

It can be seen from figure 2 that the gradient and mean square error of the training data are gradually stable with the increase of algebra, and it can be seen that the results of the training model are relatively stable.

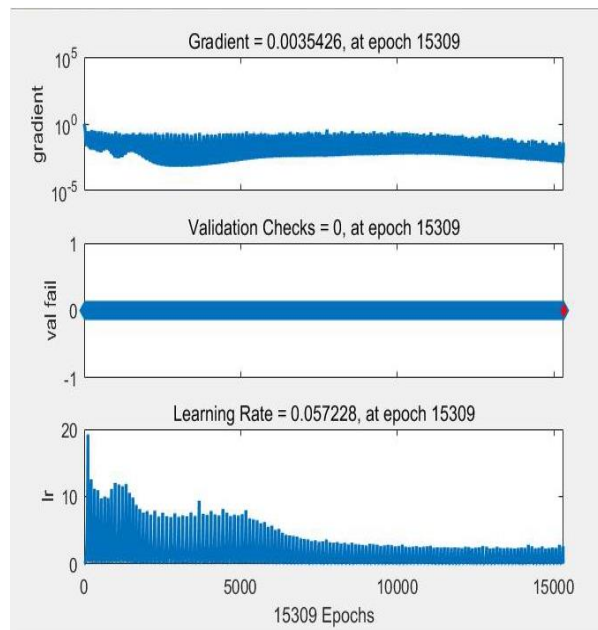


Figure 2. Gradient and mean square error of training data

3.2.3 Establishment of decision tree model

The main function of decision tree [15] model is to assist the command decision at the corresponding level and handle the relevant decision questions according to the rules set up in the model [16]. It is a machine learning method that uses probability and graph theory to establish a tree structure and compares different schemes in the tree. It is used to describe the tree structure for classifying instance objects, which is composed of nodes and edges [17].

Decision tree is classified by conditional judgment logic similar to if-else, one kind of supervised learning. Supervised learning is the method of learning with a batch of samples with a set of characteristics (attributes) and a classification result. It is learning from the results of classification and known samples [18].

ID3 algorithm selects the information gain as the measurement function of split nodes, traverses each attribute feature, uses the measurement to detect attributes, and completes the classification task through serial recursive construction [19]. Suppose N is a set of n samples, and suppose the class label has m different data values. If we define different types of data $c_i (i = 1, 2, \dots, m)$. We need to assume that n_i is the number of samples in c_i , so that given a sample information, classify it, and its expected information is represented by the following formula.

$$I(n_1, n_2, \dots, n_m) = -\sum_{i=1}^m p_i \log_2(p_i) \quad (11)$$

Note that $p_i = n_i/n$ in this formula is the probability of c_i in any sample value, and note that the logarithm function in this formula is base 2.

Suppose attribute A has v distinct values $a_1, a_2, \dots, a_v$, which divides attribute A into v subsets $\{s_1, s_2, \dots, s_v\}$, and then assume that the samples in s_j have the same value $a_j (1, 2, \dots, v)$. Then, assuming that s_{ij} is the number of its samples, the entropy or expectation of subsets divided by attribute A is given by the following formula.

$$E(A) = \sum_{j=1}^v ((s_{1j} + s_{2j} + \dots + s_{mj})/s) * I(s_{1j} + s_{2j} + \dots + s_{mj}) \quad (12)$$

According to the model principle, the smaller the entropy value, the higher the purity of subset division will be. For any given subset, their information expectations can be obtained by the following formula.

$$I(s_{1j} + s_{2j} + \dots + s_{mj}) = -\sum_{i=1}^m p_{ij} \log_2(p_{ij}) \quad (13)$$

$P_i = n_i/n$ in this formula is similar to the principle of formula (1). It is the probability of belonging to c_i in any sample. The information gain that can be obtained by branching in attribute A is as follows.

$$Gain(A) = I(s_1, s_2, \dots, s_m) - E(A) \quad (14)$$

Another algorithm used is the C4.5 algorithm, which selects attributes at nodes of all levels of the decision tree. It can use the gain ratio as the selection criterion for attributes, but of course this selection criterion is not unique.

$$\begin{cases} Split(A, s) = -\sum_{i=1}^c \frac{|s_i|}{|s|} \log_2 \frac{|s_i|}{|s|} \\ Gain(A, s) = \frac{Gain(s, A)}{Split(s, A)} \end{cases} \quad (15)$$

4. Results

4.1. BP neural network model solution results

Firstly, the data fitting of the training set and the test set is simulated. Figure 3 is the data fitting graph of the two values. It can be clearly seen from the data fitting diagram below that the two values fit well, indicating that when flight parameters are input as a data training set, the obtained test set is more accurate.

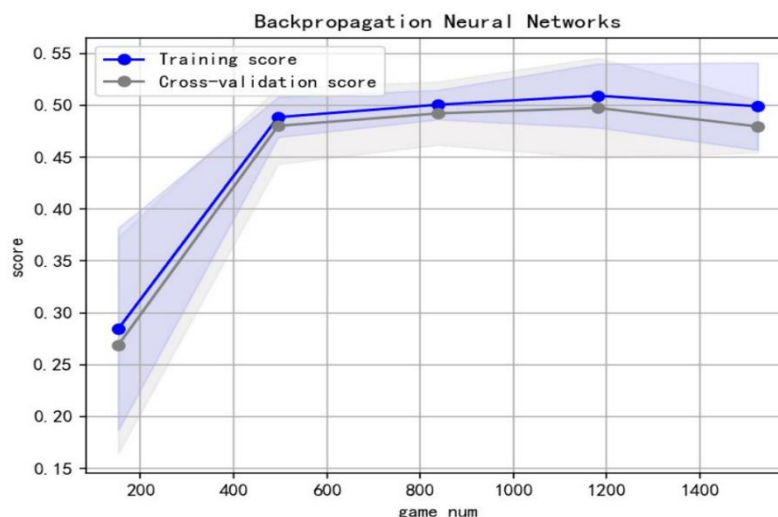


Figure 3. Data fitting of BP model

Next, the R -value of training set, test set and all data is simulated. In this model, R -value is one of the important data evaluation, and R -value represents the goodness index, which can be used to evaluate the fit of the model. Its value ranges from $[-1,1]$, and the closer the absolute value of R is to 1, the better. It can be seen from figure 4 that the R -values of the three groups of data are all at the level of close to 1, which indicates that the fitting probability of this model is very high.

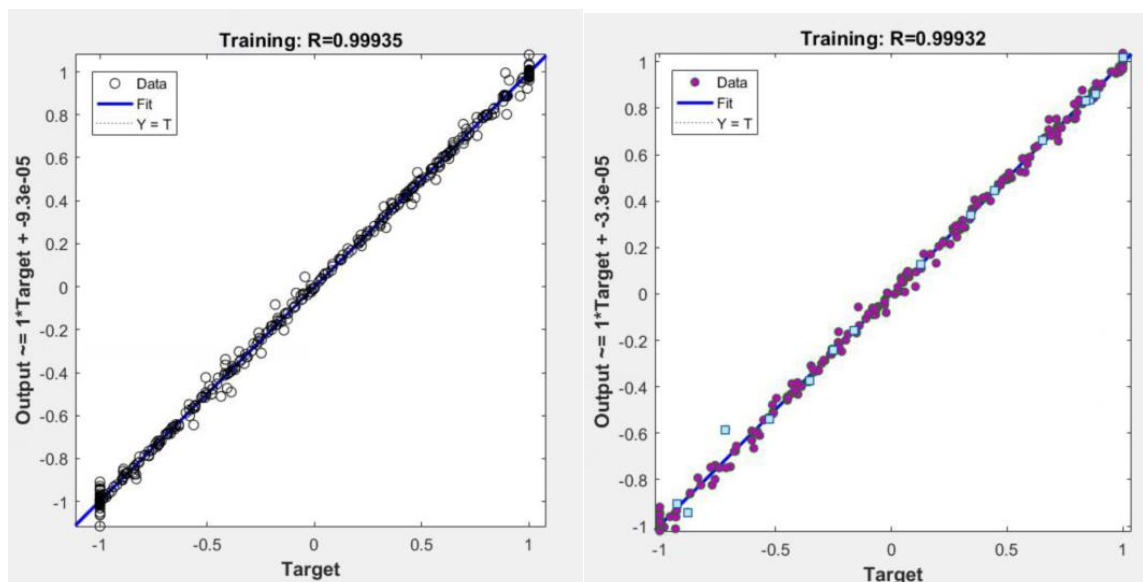


Figure 4. Historical residuals of training data

4.2. Decision tree model solution results

The solution of decision tree model is not much different from that of BP neural network. The simulation of training set and test set is also carried out by taking flight parameters as the input of training set. Relevant test results are obtained as shown in the following figure. This is why BP neural network model is chosen as the evaluation method.

There are two explanations for why the error of decision tree model is large. Firstly, the decision tree model is prone to overfitting, which is caused by the factors of the model itself. Secondly, the result of information gain in the decision tree is biased to those features with more numerical values for data with different samples [20], which is caused by data samples.

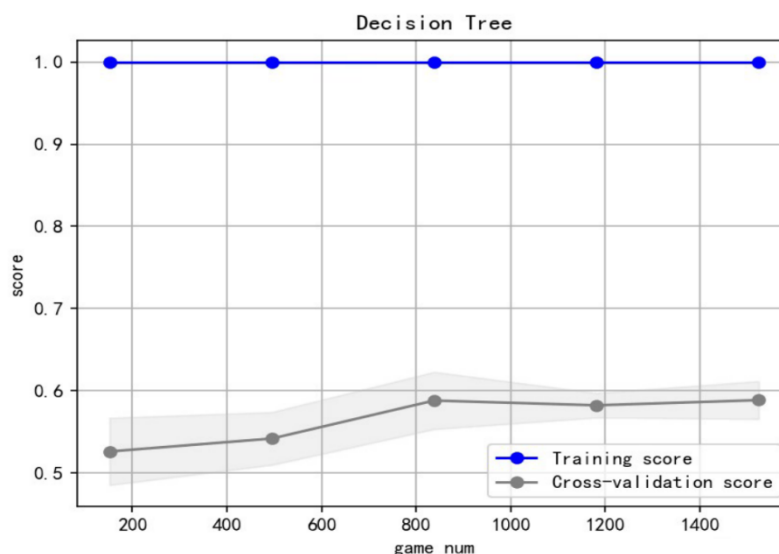


Figure 5. Data fitting of decision tree model

4.3. Comprehensive analysis and prediction evaluation

It can be seen from the previous analysis that the model evaluation method based on flight parameters adopts the BP neural network model, which can predict the qualification of pilots according to the parameters. BP neural network model has high accuracy, while machine learning algorithm cannot reach high accuracy, its data deviation is almost more than 60%, so this model is adopted. Table 3 shows a comparison of the predicted results with the actual results (in part). According to the flight parameters, the corresponding qualification level can be obtained. As can be seen from the table, 14 groups of 15 sample data are correct, so the prediction accuracy reaches 93%.

Table 3. Cumulative contribution rate calculation table

Predict outcome	True result	Parameter 1	Parameter 2	Parameter 3	Parameter 4	Parameter 5
A	A	-0.022	1.234	0.84	181.1	30116
M	M	-0.009	1.273	0.793	169	18716
A	A	-0.011	1.246	0.824	171.6	31120
F	F	-0.025	1.18	0.859	177.5	27612
F	F	-0.029	1.266	0.668	182.8	32752
F	F	-0.028	1.168	0.813	169.3	21716
M	M	0.027	1.207	0.699	181.3	29112
F	F	-0.114	1.301	0.785	174	31120
A	A	-0.029	1.293	0.891	167.8	32764
T	T	0.055	1.215	0.84	164.9	25652
F	F	-0.007	1.234	0.84	169.1	27616
A	J	-0.006	1.141	0.852	183.6	31120
F	F	0.018	1.266	0.66	177	32128
A	A	-0.028	1.18	0.879	175.9	31112
F	F	-0.008	1.118	0.84	174.5	30116

In this study, we mainly build BP neural network and decision tree model and find that BP neural network is closest to the actual situation. The model has strong nonlinear simulation ability, high prediction accuracy, strong self-learning and self-adaptation ability [21]. Finally, we conduct a comprehensive analysis of flight technology evaluation, which is convincing.

5. Conclusions

We preprocess the data through PCA dimension reduction and find that the accuracy of the result reaches 99% when the 9th line of data is used, so the 9th line of data will be used as the sample. Secondly, the BP neural network model and decision tree model are established respectively. Finally, the data deviation of the decision tree model is almost more than 60 percent. However, 14 groups of 15 sample data are correct when BP neural network is used. The prediction accuracy reaches 93 percent. So BP neural network test set is more accurate. The specific forecast results have been shown in Table 3. The model established in this paper is very simple and ingenious to solve the problem. And we make some simplification of the problem, which is easy to understand and answer the problem. But BP neural network has some shortcomings, such as slow convergence speed, the network is prone to local minimum, and the learning process often needs to be shaken. Therefore, it is necessary to improve the prediction accuracy in actual flight evaluation.

References

- [1] Chen Ke, Chen Jianqiang, Du Peng. Consideration on the development of Domestic BIM Modeling Software: Study on the Influencing factors based on PCA [C] Civil Engineering Graphics Branch of China

- Graphic Society. The 8th International Conference on BIM Technology -- Proceedings of Collaborative Application Innovation and Development in the Whole Life Cycle of Engineering Projects. Journal of Information Technology in Civil Engineering and Architecture Architecture), 2021:89-95.
- [2] Gong Yanbin, Man Jingbin, Gao Bo, Zhang Xinglei. A modeling method based on decision tree level of aircraft fault diagnosis research [J]. Mechanical and electrical information, 2023 (7) : 39-42.
- [3] Hao Shuai, Zhu Daiwu, Chen Ken. Research on flight safety Interval Method based on ADS-B [J]. Science Technology and Engineering, 2010, 10(33):8333-8336.
- [4] Liu Jiangping. Research on Flight Safety of Hainan Airlines Based on BP Neural Network [D]. Taiyuan University of Technology, 2014.
- [5] Nie J, Xu K, Chen Y. A glass classification model for glass artifacts based on PCA and SVM [C] Wuhan Zhicheng Times Cultural Development Co., Ltd.. Proceedings of 2023 International Conference on Mathematical Modeling, Algorithm and Computer Simulation (MMACS 2023). 2023:286-291.
- [6] Pan Zichao, Yang Wenjing, and Wang Xiaoshuang. " Design of Information communication Command Model Based on Decision Tree ". Proceedings of the 34th China Simulation Conference and the 21st Asian Simulation Conference. Ed.. , 2022, 927-931.
- [7] Tang Yichao, et al. " Analysis of China's Regional Digitalization Level Index and its Influencing Factors on GDP: Based on BP Neural Network Evaluation Model and BART Regression Model ". Proceedings of the 2021 (7th) National Statistical Modeling Contest for College Students (I). , 2021, 675-726.
- [8] Wang Jiawang. Based on AHP and BP network ship suppliers and multi-objective evaluation system research [C] China shipbuilding engineering society shipbuilding technology academic committee. 2021 conference on advanced manufacturing technology of the ship. [publisher unknown], 2021:164-174.
- [9] Wang Jiquan, Qiu Lichun, Wang Fulin, Wang Ying, Ren Yanjie, Sun Futian. Application of BP Neural Network in Predicting total power of agricultural machinery [J]. Mathematics in Practice and Understanding, 2013, 43(01):72-79.
- [10] Wan Zhonghong, Tao Chunfeng, Guo Rui, et al. Development and application of waveform component decomposition and reconstruction software based on Principal Component Analysis (PCA) [C] China Earth Sciences Union Annual Conference. 2021 China Earth Sciences Union Annual Conference Proceedings (9) -- Topic 25 Applied Geophysical Frontiers, Topic 26 Oil and Gas Field and Coal Field Geophysical Exploration, Topic 27 Water resources geophysical fine detection and multi-field data fusion. Beijing "electronic publishing house, 2021:175-178.
- [11] Wang Xinyu, Guo Zhenwei, Yu Dan, Liu Yimin, Cui Zhiguo. Hourly Load Prediction of Residential Heating Based on PCA-PSO-BP Neural Network [J]. Hvac, 2023 does (03) : 138-142 + 160.
- [12] Wen Chao, Zhang Li, Meng Bin, Zuo Quan. Evaluation of Airline Safety management System based on BP neural network [J]. Henan Science, 2020, 38(10):1660-1668.
- [13] Wang Hui, Zhang Wenjie, Liu Jie, Chen Linfeng, Li Zenan. Flight delay prediction model based on Classification regression Decision Tree algorithm [J]. Journal of Civil Aviation University of China, 2022, 40(03):35-40.
- [14] Xiao Dongsheng, Wang Bin, Lin Daoyuan, et al. Research on Evaluation of Network and Electric Combat Talents Based on BP Neural Network [C] Chinese Institute of Command and Control. The 10th China controller in the conference proceedings (I). The weapon industry press, 2022:230-234.
- [15] Xie Xin, Zhang Xianrong, Yang Jilin. Induced neighborhood equivalent relations improved ID3 decision tree algorithm [J]. Computer application research, 2022, 33 (01) 6:102-105 + 112.
- [16] Zhong Zhong, He Xiaofan, Ma Tao, et al. Dishwasher clean index prediction model based on BP neural network [C] China household electrical appliances association. Proceedings of China household appliances technology conference in 2022. "electrical appliances" magazine, 2023:1972-1976.
- [17] Zhu Enliang, Han Kang, Ma Xiaochen, Yang Qifan. Design of fast access recorder for civil Aircraft [J]. Aeronautical Computing Technology, 2021, 51(05):97-99+104.