

Research on Classification and Prediction of Underwater Navigation Adaptation Area Based on BP Neural Network

Haotong Wang^{*,#}, Xijun Gao[#], Zhiguo Bai[#]

School of Track Intelligent Engineering, Dalian Jiaotong University, Dalian, China

* Corresponding author: wang15842752121@163.com

[#]These authors are co-first authors.

Abstract: Aiming at the problem of classification and prediction of adaptive area for underwater navigation and positioning, this paper studies the processing of gravity anomaly data and the method of adaptive calibration. Aiming at the problem of calibration and recognition of the current adaptation area, this paper proposes a method based on hierarchical clustering analysis and BP neural network. Firstly, a hierarchical clustering analysis model is proposed to divide the suitability of each underwater navigation and positioning area. Through data preprocessing, interpolation encryption and significance analysis, the suitability calibration of the area is completed. Secondly, a BP neural network prediction model is established to predict the adaptability of the region, and the performance of the model is improved by normalization processing and network parameter optimization. Finally, through experimental verification, the classification results of the adaptation area and the prediction results of the BP neural network are obtained. The prediction accuracy is 82.7929% and the correlation coefficient is 0.94641, which verifies the effectiveness of the proposed method. This method provides theoretical support and decision-making basis for the navigation accuracy improvement of underwater vehicles and marine economic development planning.

Keywords: Significance Analysis, Hierarchical Cluster Analysis, Normalization Processing, BP Neural Network Prediction.

1. Introduction

Under the background of the construction goal of "strong marine province," the innovation in the field of marine high-tech is of great significance, especially the classification and prediction technology of underwater navigation and positioning[1]. This technology is one of the key core technologies for underwater vehicles to perform tasks, and gravity-assisted navigation is one of the main methods to achieve high-precision navigation. The calibration and recognition technology of the adaptation area is one of the most challenging problems. Its core is to interpolate and encrypt the gravity reference map of the studied sea area. Gravity anomaly refers to the deviation between the actual observed gravity value and the theoretical normal gravity value, which is mainly caused by the uneven distribution of the earth's material density. In the area where the gravity anomaly changes significantly, the navigation system can obtain higher positioning accuracy. In the area where the gravity anomaly changes flat, the positioning accuracy of the navigation system will be significantly reduced. Therefore, it is very important to establish an effective adaptation area classification prediction model to ensure the navigation accuracy of underwater vehicles.

In the existing research, Qinfeng Xu et al.[2] introduced the basic algorithm principle of BP neural network, and took the financial crisis early warning model of China's real estate industry as an example to elaborate the specific application method of BP neural network in the field of prediction. Based on the background of big data and Internet finance, Zhang D et al.[3] proposed a stock price prediction model based on BP neural network. However, these studies have some limitations. First, they mainly focus on a single field (such as finance or real estate), and lack targeted research on complex nonlinear

problems such as marine gravity anomaly data. Secondly, the model optimization method is more traditional, which does not fully solve the problems of slow convergence speed and easy to fall into local minimum value of BP neural network. In addition, the existing research has insufficient adaptability in data preprocessing and feature extraction, and it is difficult to directly migrate to the classification prediction task of marine gravity adaptation area.

Based on the above analysis, this paper has carried out the following work in view of the shortcomings of the current research. Firstly, the gravity anomaly data is preprocessed and interpolated to construct a refined reference map. Secondly, through the significance analysis and hierarchical clustering method, the division and calibration of the adaptation area are completed. Finally, the BP neural network prediction model is constructed, optimized and evaluated. The experimental results show that the model has high accuracy and robustness in the classification and prediction of the adaptation area, which provides effective support for the accuracy guarantee of the underwater navigation system.

2. Principle of correlation method

2.1. Adaptability grade classification based on hierarchical clustering analysis

Hierarchical clustering analysis[4] is an unsupervised learning method that achieves classification by gradually merging or splitting data samples. The core of this method is to construct a tree pedigree diagram to visually display the hierarchical relationship between data samples, and finally determine the classification results by setting the threshold or the number of categories. In the problem of suitability classification, the first step is to preprocess the gravity anomaly data, including interpolation encryption and

significance test, to quantify the statistical characteristics of each region. Then, the Euclidean distance[5] is used to calculate the similarity between regions. The formula is:

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (1)$$

where x and y represent the test statistics vectors of the two regions, respectively.

In the clustering process, the single connection method, the full connection method or the average connection method can be selected as the measure of the distance between classes, which are suitable for data with different distribution characteristics. The single connection method takes the minimum distance as the distance between classes, which is suitable for finding clusters with long strip distribution. The full connection method takes the maximum distance as the inter-class distance, which is suitable for finding compactly distributed clusters. The average connection rule balances the

advantages and disadvantages of the first two methods by calculating the average distance between samples.

By gradually merging similar regions and generating a tree pedigree diagram, the tree diagram can be cut according to the actual needs to obtain the final classification results. For example, in this paper, the region is divided into six types of suitability levels, including 'extremely poor suitability', 'poor suitability', 'poor suitability', 'strong suitability', 'strong suitability' and 'strongest suitability'. The advantage of hierarchical clustering analysis is that it does not need to preset the number of categories, and the visualization results are intuitive and suitable for exploratory data analysis.

2.2. Prediction of underwater navigation suitability level based on BP neural network

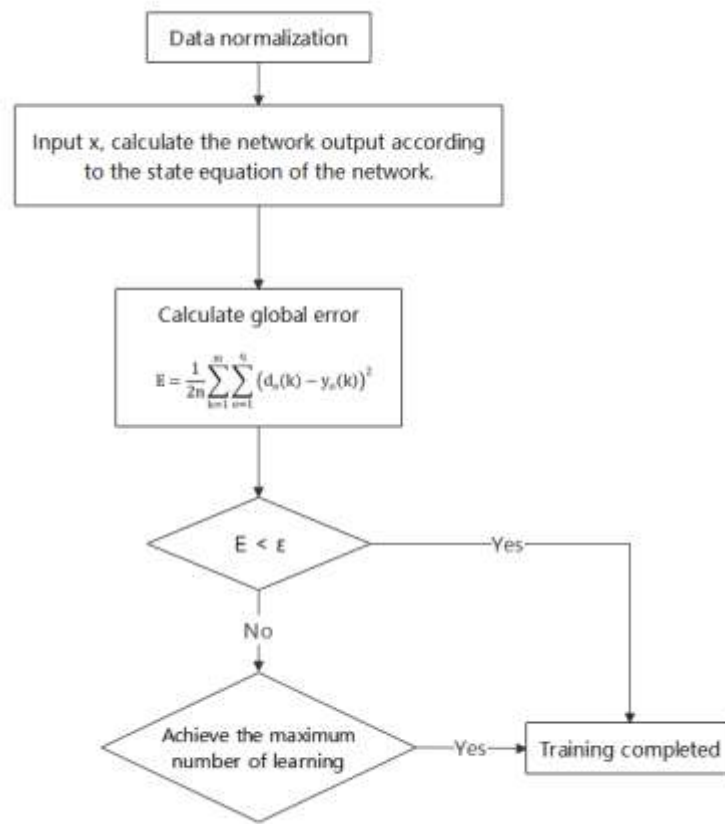


Figure 1. BP neural network algorithm flow chart

BP (Back Propagation) neural network[6] is a multi-layer feedforward neural network, which adjusts the weights and thresholds through the error back propagation algorithm to realize the modeling of nonlinear relationships. Its core principle includes two stages: forward propagation and error back propagation. Here are the model building steps:

1. Input layer and feature selection: The test statistics of the divided area, the mean and variance of the gravity anomaly value are used as the input feature X , and the dimensional influence is eliminated after normalization.

2. Hidden layer design: Double hidden layer structure is adopted. The activation function is hyperbolic tangent function (TanSig), and its expression is:

$$\text{TanSig}(s) = \frac{e^s - e^{-s}}{e^s + e^{-s}} \quad (2)$$

It has smooth gradient characteristics and is suitable for dealing with nonlinear problems.

3. Output layer and loss function: The output layer uses a linear function (Purelin) to map the predicted value of the fitness level, the loss function uses mean square error (MSE), and the error is minimized by the gradient descent method.

4. Back Propagation Optimization: The weight adjustment formula is:

$$\Delta w(n) = -\eta \nabla E(n) + \alpha \Delta w(n-1) \quad (3)$$

where η is the learning rate and α is the momentum factor, which is used to accelerate convergence and avoid local minima.

3. Simulation analysis

3.1. Data pre-processing

Data from <http://topex.ucsd.edu/>

3.1.1. Treatment of outliers and missing values

In dealing with outliers, this paper uses a method based on quartiles and interquartile range to process data. This method first calculates the Q1 and Q3 of the data, and then calculates the IQR. Next, we use lb and ub to determine which data are

outliers. Then, the outliers are removed and processed as missing values. Finally, 459 outliers were determined. The distribution map of the size of the gravity anomaly value is as follows:

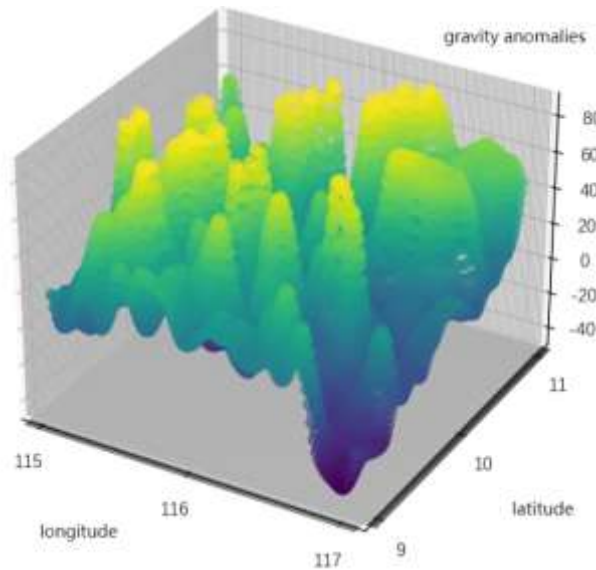


Figure 2. $1^\circ \times 1^\circ$ gravity anomaly value A reference map

3.1.2. Interpolation encryption processing

In order to make the accuracy of the benchmark map drawn higher and more accurate, this paper chooses to interpolate and encrypt it, and uses the moving average method in the process. After removing the outliers above, there will be some missing values in the data. In order to accurately describe the trend and pattern of the data, this paper needs to fill these missing values. In order to fill these missing values, this paper uses the fillmissing function in MATLAB, which is based on the moving average method to estimate and fill the missing values. The moving average method is a technique for smoothing time series data. It uses the weighted average of

the past values of the data point sequence to estimate the current value. Finally, the encrypted data is inserted.

3.1.3. Data normalization and network parameter configuration

In order to facilitate the establishment of BP neural network prediction model, it is necessary to normalize the data[7]. Firstly, this paper assigns a random number in the interval (-1,1) to each connection weight (the specific weight is given in the appendix). The learning rate R is set to 0.01, and the error function e is $y = \text{tansig}(s)$ and $y = \text{purelin}(s)$. The specific image of the function is as follows:

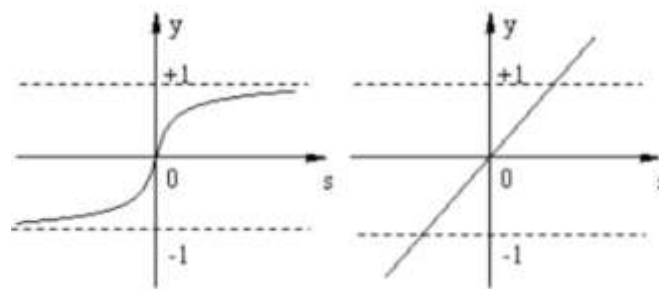


Figure 3. $y = \text{tansig}(s)$ Image (left) $y = \text{purelin}(s)$ (right)

Then, the calculation accuracy value ϵ is given to be 0.000001,

3.2. Results of regionalized classification of data

In order to divide the region reasonably and complete the suitability calibration of each region, this paper will carry out significance analysis[8]. Firstly, it is divided into $0.1^\circ \times 0.1^\circ$ regions, and the gravity anomaly value is obtained after division.

Then, the gravity anomaly values of four points in a region are analyzed significantly. In order to facilitate the subsequent cluster analysis, this paper quantifies the gravity anomaly values of four points in a region, that is, the significance test

and the maximum number of learning times M is 982. of single sample T is carried out. After the significance analysis, the p-value divided by region is as follows:

Table 1. p-values by region

1.72	1.61	1.39	3.42	2.06	1.19
-0.55	0.22	1.88	1.9	6.79	3.12
0.25	2.08	5.9	2.01	1.48	12.55
1.06	1.47	2.13	1.78	3.15	2.78
3.38	2.23	1.27	1.97	1.33	3.04
11.06	3.09	1.29	-1.94	-0.15	-7.27

The test statistics obtained by the obtained regional

division are shown in the following table:

Table 2. Test statistics by region

0.1841	0.2057	0.2591	0.0418	0.1309	0.3199
0.6212	0.8424	0.1572	0.1535	0.0065	0.0523
0.8211	0.1286	0.0097	0.1374	0.2343	0.0011
0.3666	0.2367	0.1234	0.1737	0.0512	0.069
0.0431	0.1121	0.2952	0.1438	0.2755	0.0557
0.0016	0.0538	0.2881	0.1477	0.8929	0.0054

If the p value is less than the selected significance level, it can be considered that there is a significant relationship between the variables. If the p value is greater than the significance level, it can be considered that there is no significant relationship between the variables. The three-dimensional graph of the specific test statistic drawn by is as follows:

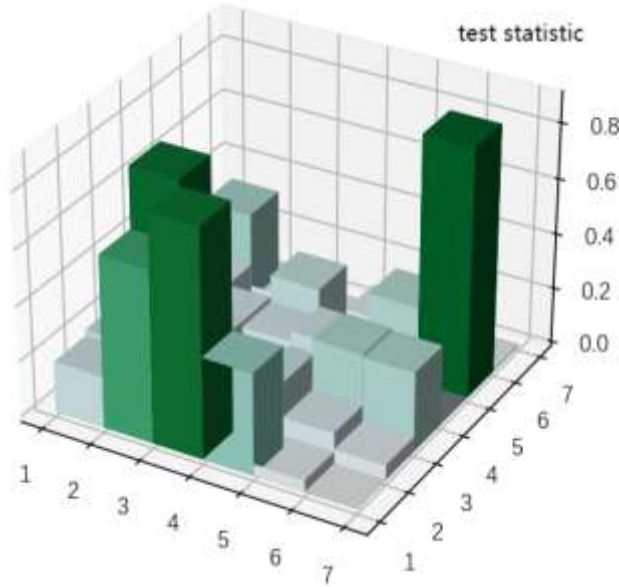


Figure 4. Three-dimensional test statistics histogram in the region

In order to observe the statistical test statistics more clearly, the dimension is reduced, and the two-dimensional diagram

is drawn as follows:



Figure 5. Two-dimensional statistical test distribution

In order to complete the regional classification of the data, it is necessary to perform hierarchical clustering analysis on the test statistics to solve the categories of each divided area. In this paper, through the hierarchical clustering analysis method, some samples with greater similarity between each other are specifically found to be aggregated into one category, and other samples with greater similarity between each other are aggregated into another category. Closely related samples

are aggregated into a small classification unit, and distantly related samples are aggregated into a large classification unit, until all samples are aggregated, and different types are divided one by one to form a classification system from small to large. Finally, the whole classification system is drawn into a pedigree chart, which is used to express the relationship between all samples. The pedigree chart is as follows:

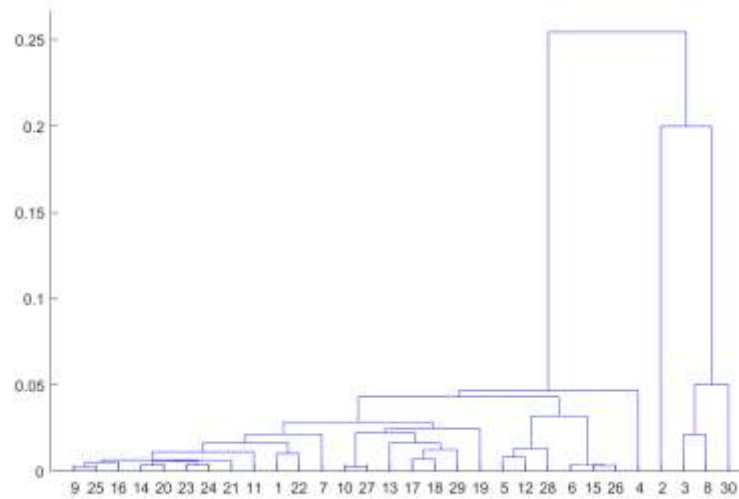


Figure 6. Cluster pedigree diagram

3.3. Classification results of adaptation region

region, each region is named first, as follows:

In order to clearly show the suitability calibration of each

Table 3. Naming of each region

Longitude & latitude	11.0068°- 10.8431°	10.6794°- 10.5155°	10.3516°- 10.1877°	10.0236°- 9.8594°	9.6951°- 9.5308°	9.3664°- 9.2019°
115.0083°-115.175°	1	7	13	19	25	31
115.3417°-115.5083°	2	8	14	20	26	32
115.675°-115.8417°	3	9	15	21	27	33
116.0083°-116.175°	4	10	16	22	28	34
116.3417°-116.5083°	5	11	17	23	29	35
116.675°-116.8417°	6	12	18	24	30	36

The 6 groups obtained by hierarchical cluster analysis are as follows:

Table 4. The suitability calibration of each region

Adaptability is very poor	5、6、12、15、19、26、28、32、33、34、35、36
Poor suitability	4
Adaptability is relatively poor	1、7、9、10、11、13、14、16、17、18、20、21、22、23、24、25、27、29、31
Adaptability is relatively strong	2
Strong adaptability	3、8
Adaptability is very strong	30

3.4. BP neural network model training and prediction results

3.4.1. BP neural network model training

Through the normalization of the data and the

configuration of the network parameters, as well as the construction of the BP neural network prediction model, the BP neural network prediction model is trained, and the analysis images of the predicted value and the real value are as follows:

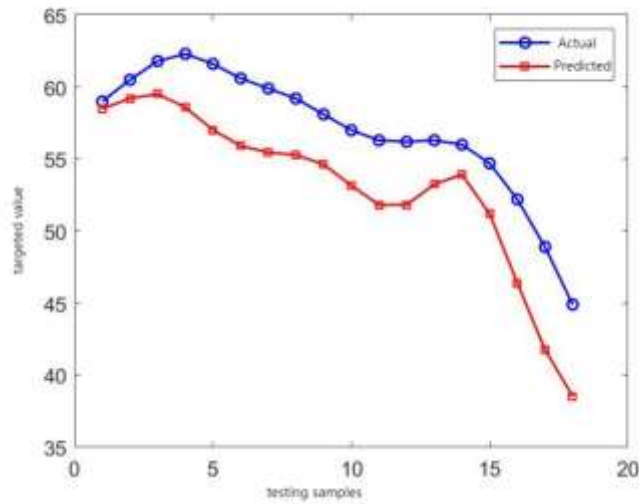


Figure 7. Comparison of predicted and actual values Line chart

The image of the mean square error of the training set, the validation set, the test set and the overall change with the number of training times is as follows:

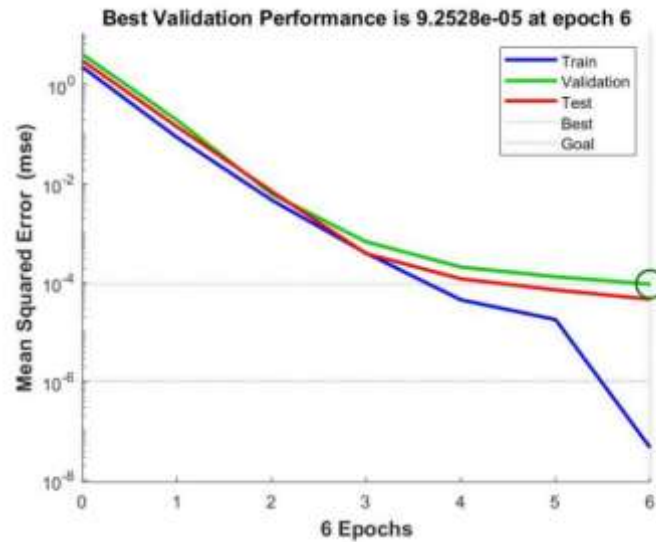


Figure 8. Mean square error of training set, validation set, test set and population

The training images of each stage of BP neural network are as follows:

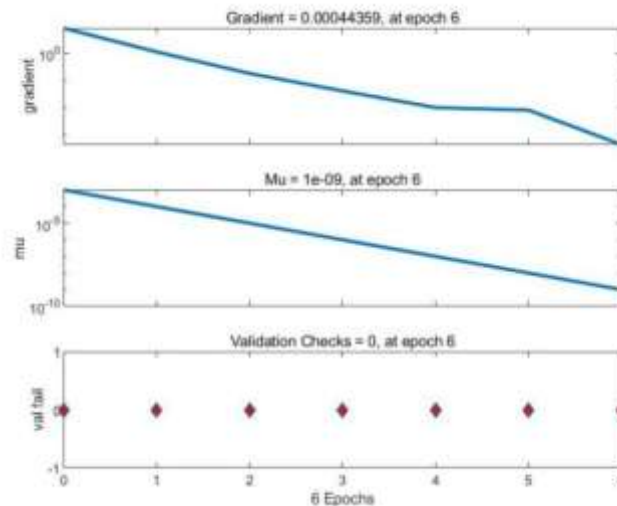


Figure 9. Training images of each stage

The correlation analysis images of each sample set and the overall are as follows:

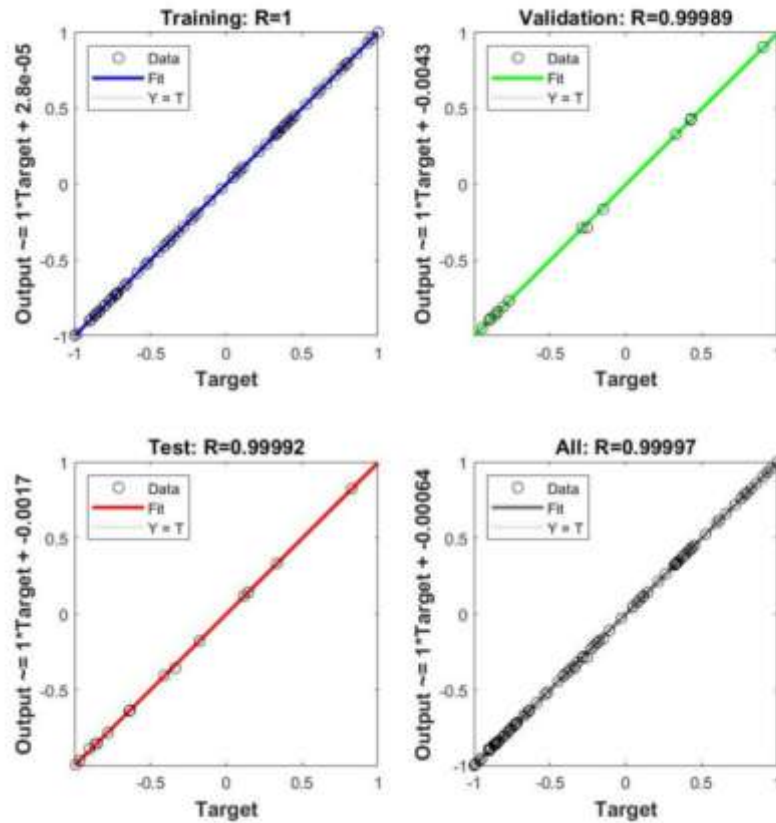


Figure 10. Correlation analysis image

The prediction error of BP neural network test set is as follows:

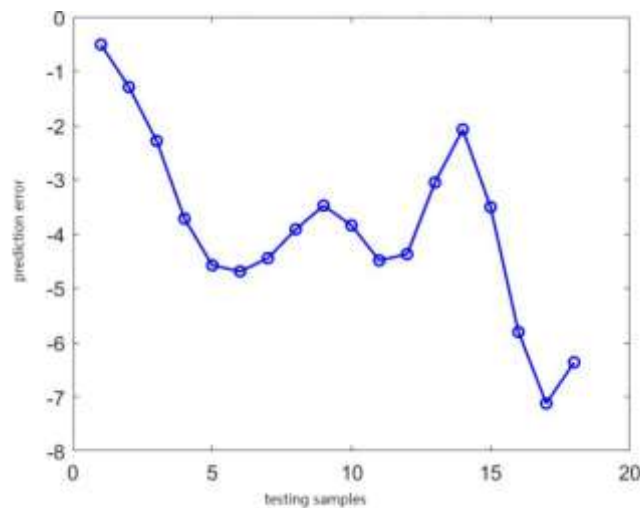


Figure 11. Test error line chart

3.4.2. BP neural network model evaluation and prediction results

After the model training is completed, the model needs to be further evaluated to determine its functional performance. The trained model is applied to the test set to evaluate the performance of its prediction results.

Finally, the error square sum SSE is 2672.606. The mean absolute error MAE[9] is 9.1474. The mean square error MSE[10] is 148.4781. The root mean square error RMSE is

12.1852. The mean percentage error MAPE[11] is 17.2071 %. The prediction accuracy is 82.7929 %. The correlation coefficient R is 0.94641.

The result set of its tests is as follows:

Table 5. BP neural network prediction model according to the test results of the data in Annex I

numbering	actual value	BP predicted value	error
1	59.0000	59.2584	0.2584
2	60.5000	61.5403	1.0403
3	61.80006	63.8767	2.0767
4	62.30006	64.9090	2.6090
5	61.6000	63.8905	2.2905
6	60.60006	61.7871	1.1871

4. Conclusions

This paper focuses on the classification and prediction technology of the adaptation area of underwater navigation and positioning. It focuses on solving the problems of preprocessing, region division and adaptability calibration of gravity anomaly data, and constructs an efficient BP neural network prediction model. Through data interpolation encryption, significance analysis and hierarchical clustering analysis, the classification and calibration of the adaptation area were successfully completed. Through the training and optimization of BP neural network, high-precision regional suitability prediction is realized. The simulation results show that the model has excellent performance in error control and prediction accuracy, which provides reliable technical support for the accuracy guarantee of underwater navigation system.

References

- [1] Ma H, Mu X, He B. Adaptive navigation algorithm with deep learning for autonomous underwater vehicle[J]. Sensors, 2021, 21(19): 6406.
- [2] Xu Q, Kong W. Research on forecast model based on BP neural network algorithm[C]//Journal of Physics: Conference Series. IOP Publishing, 2021, 1982(1): 012065.
- [3] Zhang D, Lou S. The application research of neural network and BP algorithm in stock price pattern classification and prediction[J]. Future Generation Computer Systems, 2021, 115: 872-879.
- [4] Ran X, ** Y, Lu Y, et al. Comprehensive survey on hierarchical clustering algorithms and the recent developments[J]. Artificial Intelligence Review, 2023, 56(8): 8219-8264.
- [5] Hohmann M, Devriendt K, Coscia M. Quantifying ideological polarization on a network using generalized Euclidean distance[J]. Science Advances, 2023, 9(9): eabq2044.
- [6] Song S, **ong X, Wu X, et al. Modeling the SOFC by BP neural network algorithm[J]. International Journal of Hydrogen Energy, 2021, 46(38): 20065-20077.
- [7] Huang L, Qin J, Zhou Y, et al. Normalization techniques in training dnns: Methodology, analysis and application[J]. IEEE transactions on pattern analysis and machine intelligence, 2023, 45(8): 10173-10196.
- [8] Grabski I N, Street K, Irizarry R A. Significance analysis for clustering with single-cell RNA-sequencing data[J]. Nature Methods, 2023, 20(8): 1196-1202.
- [9] Hodson T O. Root mean square error (RMSE) or mean absolute error (MAE): When to use them or not[J]. Geoscientific Model Development Discussions, 2022, 2022: 1-10.
- [10] Zhou J, Li X, Ding T, et al. On the optimization landscape of neural collapse under mse loss: Global optimality with unconstrained features[C]//International Conference on Machine Learning. PMLR, 2022: 27179-27202.
- [11] Malburg L, Hoffmann M, Bergmann R. Applying MAPE-K control loops for adaptive workflow management in smart factories[J]. Journal of Intelligent Information Systems, 2023, 61(1): 83-111.