

Estimation of Shipment Size in Seaborne Iron Ore Trade

X. Zhou & Q. Hu

Shanghai Maritime University, Shanghai, China

ABSTRACT: Shipment size is unavailable and important in AIS-based trade volume estimates. A method of shipment size estimates based on AIS (Automatic Identification System) data and BP neural network is proposed. The ship's length, width, designed draught, current draught and deadweight ton are input parameters, the actual shipment size of the ship is output value, and the BP neural network is trained to estimate the actual shipment size of the iron ore carriers. Then, the AIS data is used to calculate the iron ore trade volume in 2018. Compared with customs data, the annual error of import volume of China is less than 0.5%. The result shows that the proposed method is accurate and practical.

1 INTRODUCTION

The analysis of global trade and market is usually based on official customs data, which is relatively macroscopic and lagging behind. It is suitable for macro analysis (see Fu et al 2006; Liu 2014; Chen 2015). The other method is the aggregation of inter-regional seaborne trade flows by tracking vessel movements over time (see Nossun 1996; Haji S 2013; Adland R et al 2017), which is microscopic and flexible. According to Nossun, Fearnleys, a shipping broking firm, summarized the shipping information provided by the ship brokers to get the reviews of the world bulk trades. With the emergence of AIS technology, the real-time broadcasting of AIS messages makes the shipping data more timely, convenient and objective. The AIS data were used to summarize and analyze the global container trade circulation in Haji S (2013) and the global crude oil trade circulation in Adland R (2017). Both of the statistics were compared with the official data, and the impact of the difficulty in determining the actual shipment size of ships was mentioned.

The AIS data does not include the true shipment size of the ship, the information related to the shipment size is the deadweight ton. Deadweight ton is the shipment size when the ship is fully loaded. But in the actual transportation, the ships are always no load or half load, and it is difficult to get the centimeter deadweight ton which reflects the actual shipment size of the ship. Ren et al (2017) established the state recognition system of ship load based on neural network. However, the object of the study is inland ships, and only the ships' load state can be recognized, it's not possible to make an accurate estimate of the shipment size. In this paper, BP neural network is used to estimate the true shipment size of ships, so as to improve the accuracy of AIS-based trade volume estimates and analysis of marine market. This method is aimed at iron ore carriers engaged in import and export trade.

2 DATA SOURCE AND PARAMETER ANALYSIS

2.1 Data source

The AIS data of the ships used in this paper is all provided by ShangHai Maili Marine Technology Co.,Ltd (<http://www.hifleet.com>) and stored after receiving and decoding.

The ships involved in this paper are all ocean bulk carriers used for international iron ore trade, and the information of such ships are included in Lloyd's maritime archives. Therefore, in order to avoid the situation that ship's AIS data is mis-input by sailors, this paper combines Lloyd's maritime archives to improve the static information of AIS data.

The actual shipment size of the ships used in this paper is derived from the Line-up of port agency LBH which includes ship name, departure port, departure time, shipment size, destination, etc. In order to exclude the impact of other factors, the selected ships are all bulk carriers delivered from Australia directly to China, and multi-port unloading ships have been excluded.

Match the Lloyd's maritime files with ship AIS data by Maritime Mobile Service Identify (MMSI) and ship name to get a more complete and accurate ship static information (name, call sign, MMSI, IMO, vessel type, designed draft, length, width, etc.) and dynamic information (longitude, latitude, speed, draft, etc.). Then, match the true shipment size for each ship with Line-up through the name, departure port and departure time.

2.2 Characteristic extraction

AIS data include static information, dynamic information and voyage-related information, there is no obvious linear relationship between all kinds of information, so we need to select the appropriate information for nonlinear modeling, in order to find the nonlinear mapping relationship between the ship's AIS messages and actual shipment size. BP neural network has obvious advantages in the modeling of nonlinear relations and is widely used in information fusion (see Hu X et al 2011), track prediction (see Gan S et al 2016; Rong Zhen 2017), information recognition (see Zhang et al 2009; Zhu et al 2012) and other studies.

After analysis and screening, the following features are selected:

- 1 Length and width of vessel. It is a major factor affecting the shipment size because it's closely related to the ship's cargo capacity.
- 2 Deadweight ton. Deadweight ton = Displacement of the fully loaded ship - the weight of empty ship. It can reflect the shipment size when the ship is fully loaded.
- 3 Designed draft. It is the draft when the ship is fully loaded.
- 4 Current draft. It changes when the shipment size changes.

The ratio of current draft to designed draft and the ratio of shipment size to deadweight ton are theoretically positively correlated.

Take the above five features as inputs and the true shipment size as output of BP neural network, and obtain the mapping relationship between the features and the true shipment size through supervised learning.

2.3 Data processing and storage

Match the AIS data with the Line-up data through the ship name, departure port and departure time to get the actual shipment size. Store the AIS data with actual shipment size in My SQL database, and set MMSI as the primary key for convenience of query and change. Storage structure is shown in figure 1.

Data preprocessing:

- 1 Dynamic data. The draft of the ship is entered manually by the crew, so the information may be entered incorrectly or changed untimely, or the draft for last voyage may be retained. For the data with unreasonable draft, search the ships' historical track on <http://www.hifleet.com> through MMSI. Observe its draft changes after departure. If there is any changes before arriving next port, choose the latter.
- 2 Static data. The AIS data were matched with Lloyd's Marine archives through MMSI and ship name, so as to improve the accuracy of AIS static data. Through preprocessing, the accuracy and integrity of training data are guaranteed.

MMSI	departure time	dwt	length	width	desdraught	curdraught	shipmentsize
477351200	2018-01-08 12:38:36	207980	300	50	18	18	170000
563008300	2018-01-09 14:04:46	210000	300	50	18	18	190000
477351200	2018-02-16 09:31:57	207980	300	50	18	18	170000
357427000	2018-02-16 23:56:49	181461	292	45	18	16	180000
255805873	2018-02-18 12:41:43	208000	299	50	18	17	190000
311749000	2018-03-04 20:13:49	203266	300	50	17	17	190000
636013937	2018-01-28 09:00:59	260995	330	59	19	20	250000
.
.
.
477510300	2018-01-09 23:39:56	400606	360	65	23	23	390000
477510300	2018-01-10 07:26:56	400606	360	65	23	23	390000
566869000	2018-01-22 22:04:40	262779	328	57	18	18	250000
477617800	2018-01-24 08:20:07	297719	327	55	21	21	290000

Figure 1. Data storage sample

3 ESTIMATION MODEL OF SHIPMENT SIZE BASED ON BP NEURAL NETWORK

3.1 Network structure

BP neural network is a kind of back propagation learning algorithm (see Yu 2011) proposed by the team of scientists led by Rumelhart and McMelland in 1986. It can store a large number of input-output mode mapping relations. Its learning rule is the steepest descent method. The weights and thresholds of the network are continuously adjusted by back propagation to minimize the network errors. BP neural network consists of a series of simple units which are closely related to each other. Its topological structure includes input layer, hidden layer and output layer. Here we build a BP neural network with

one hidden layer, five input layer nodes and one output layer node. The structure is shown as follows: $x_1 \cdots x_5$ are the inputs which are length, width, designed draft, current draft and DWT (deadweight ton); y is the output, the true shipment size; ω_{ij} is node i -to-node j weight, and ω_{jk} is node j -to-node k weight. Each neuron node has a certain number of inputs and unique output.

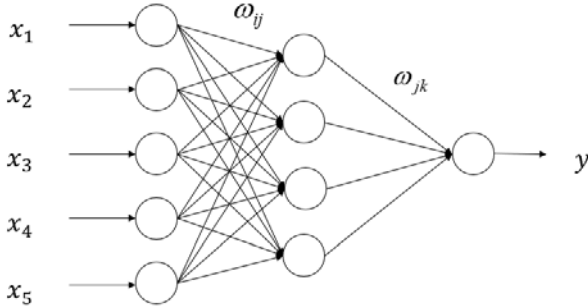


Figure 2. BP neural network structure

3.2 Data standardization

BP neural network has a strong dependence on normalization. In order to avoid the influences caused by magnitude difference of input data, the data normalization is necessary. Common methods for data normalization include min-max standardization and z-score method. The former is adopted here. The conversion function is as follows:

$$x^* = \frac{x - \min}{\max - \min} \quad (1)$$

x is sample data, \min is the minimum value of the sample data, \max_x is the maximum value of the sample data, and x^* is the converted data. The original data can be mapped to values between 0 and 1 by data standardization. After the prediction, the output results should be anti-normalization to get practical significance.

3.3 Hidden layer node

Too few nodes will affect the accuracy of network, while too many nodes may lead to overfitting. The principle to determine the number of nodes in the hidden layer is to take as compact a structure as

possible under the requirement of accuracy. Here we choose the number of nodes with an empirical expression (see Zhang et al 2008):

$$M < \sqrt{m+n} + a \quad (2)$$

M is the number of hidden layer nodes, n is the number of input layer nodes, m is the number of output layer nodes, a is an integer among 1~10. Here, n is 5, m is 1, M ' optimum range is 3~12.

3.4 Network's training

Steps of this BP algorithm is described as below (see Wang 2013):

Step 1 Initialize the parameters such as net structure, layer numbers, nodes number of each layer, the weights and thresholds, and set the appropriate network learning rate and transfer function.

Step 2 Calculate the hidden layer output H . n is the number of input layer nodes, l is the number of hidden layer nodes, θ_j is the threshold of hidden layer neuron, $f(x)$ is the transfer function of hidden layer -- sigmoid function.

$$H_j = f\left(\sum_{i=1}^n \omega_{ij} x_i - \theta_j\right) \quad j = 1, 2, \dots, l \quad (3)$$

$$f(x) = \frac{1}{1 + e^{-x}} \quad (4)$$

Step 3 Calculate the output Y . m is the number of output layer nodes, θ_k is the threshold of output layer neuron, $\phi(x)$ is the transfer function of output layer.

$$Y_k = \phi\left(\sum_{j=1}^l H_j \omega_{jk} - \theta_k\right) \quad k = 1, 2, \dots, m \quad (5)$$

$$\phi(x) = x \quad (6)$$

Step 4 Error calculation. y_k is the desired output and Y_k is the actual output.

$$e_k = y_k - Y_k \quad k = 1, 2, \dots, m \quad (7)$$

Step 5 Update the weights.

$$\omega_{ij} = \omega_{ij} + \eta H_j (1 - H_j) x_i \sum_{k=1}^m \omega_{jk} e_k \quad i = 1, 2, \dots, n; j = 1, 2, \dots, l \quad (8)$$

$$\omega_{jk} = \omega_{jk} + \eta H_j e_k \quad j = 1, 2, \dots, l; k = 1, 2, \dots, m \quad (9)$$

$$\theta_j = \theta_j + \eta H_j (1 - H_j) \sum_{k=1}^m \omega_{jk} e_k \quad j = 1, 2, \dots, l \quad (10)$$

$$\theta_k = \theta_k + e_k \quad k = 1, 2, \dots, m \quad (11)$$

Step 6 The updated weights and thresholds are used to recalculate the error. If not, return to step 2 until the error is less than the set error.

3.5 Simulation and results

1650 data of Australia-China and Brazil-China iron ore ships are put into BP neural network. The number of hidden layer nodes is 4. 1600 of them are randomly selected as training data and 50 as testing data. The training and testing are carried out according to the above process. The test results are shown as follows:

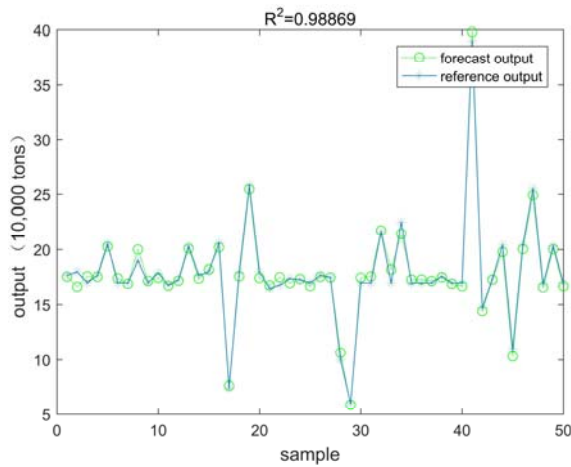


Figure 3. Comparative analysis of prediction results

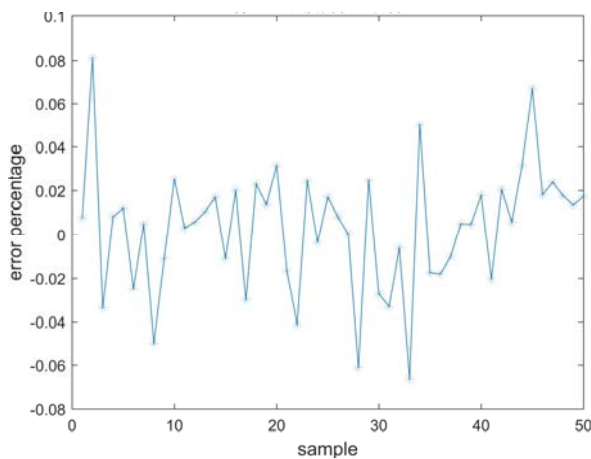


Figure 4. Error percentage

Figure 3 is a comparison between the predicted results and the expected results of the neural network. It can be seen from the graph that the predicted results are good. In order to quantify the comparison results, a determination coefficient R^2 is set to express the goodness of fit of the model. The closer the value is to 1, the better the fit of the model is. Figure 4 is the error percentage of the predicted results. In this test, R^2 is above 0.99, and the error is concentrated within 0.04%. The result is satisfactory.

4 EXAMPLE VERIFICATION

4.1 Verification process

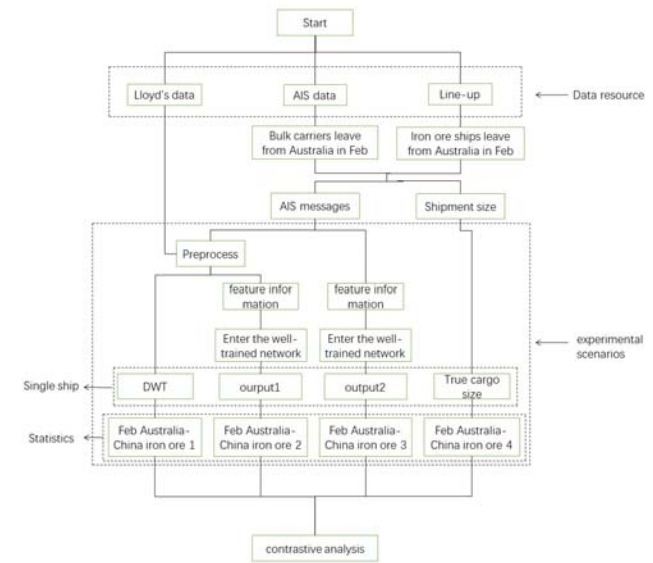


Figure 5. Model Accuracy Testing Process

In order to verify the feasibility of the proposed method, an experiment was carried out (Figure 5). The AIS data of bulk carriers leaving Australian ports in February 2018 and going to China are matched with the Line-up of iron ore ships in Australian ports in February 2018. Then, a data set of iron ore ships leaving from Australia in February is obtained. The data set contains identity information of ships (MMSI, name), model input parameters (length, width, designed draft, current draft, deadweight ton) and actual shipment size. Statistics and analysis are performed in the following four experimental scenarios:

Scenario 1 Preprocess the AIS data of iron ore ships to improve incomplete and inaccurate data. Replace the actual shipment size with deadweight ton, and the weekly statistics of all the deadweight tons of the ships is carried to obtain the 1st trade volume of Australia-China in February.

Scenario 2 Preprocess the AIS data of iron ore ships. Put the input parameters (length, width, designed draft, current draft, deadweight ton) into the well-trained network. Replace the actual shipment size with the network's output, and the weekly statistics is carried to obtain the 2nd trade volume of Australia-China in February.

Scenario 3 Select the AIS data of iron ore ships without preprocessing. Put the input parameters (length, width, designed draft, current draft, deadweight ton) into the well-trained network. Replace the actual shipment size with the network's output, and the weekly statistics is carried to obtain the 3rd trade volume of Australia-China in February.

In the case of a large amount of AIS data, data preprocessing takes so much time. So this scenario is set to test the robustness of the model and observe whether it can still have good predicted results in the case of data missing.

Scenario 4 Do the weekly statistics with true shipment sizes and obtain the 4th trade volume of Australia-China in February.

4.2 Verification result

50 data are randomly selected, and the shipment size of every ship under different experimental scenarios are as follows:

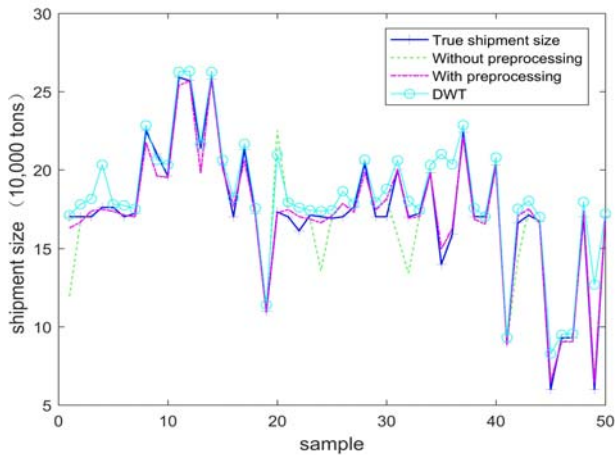


Figure 6. Shipment size under four experimental scenarios

As shown in Figure 6, under the condition of data preprocessing, the data is relatively complete, and the network prediction results are in good agreement with the actual shipment size; In the case of incomplete data, there are a few results not satisfactory, but most of the network prediction results are good, indicating that the model has a certain degree of robustness. The deadweight ton is generally larger than the true shipment size.

The statistics of Australia-China iron ore seaborne trade volume in February 2018 are as follows:

Table 1. Statistics under different experimental scenarios (million tons)

	True shipment size	With data preprocessing	Without data preprocessing	DWT
2018/2/1-2018/2/7	13.10	13.06	13.02	13.58
2018/2/8-2018/2/14	14.69	14.58	14.12	15.03
2018/2/15-2018/2/21	12.06	11.92	11.91	12.46
2018/2/22-2018/2/28	16.39	16.13	16.08	17.05
February	56.24	55.69	55.14	58.12

Calculate the prediction error e according to the following formula (x_p - predicted value, x_t - true value):

$$e = (x_p - x_t) / x_t \quad (12)$$

Table 2. Statistics error comparison under different experimental scenarios

	With data preprocessing	Without data preprocessing	DWT
2018/2/1-2018/2/7	-0.27%	-0.60%	3.69%
2018/2/8-2018/2/14	-0.75%	-3.88%	2.27%
2018/2/15-2018/2/21	-1.10%	-1.17%	3.38%
2018/2/22-2018/2/28	-1.61%	-1.87%	4.03%
Average error	-0.93%	-1.88%	3.34%

The statistical results show that, in the case of complete data, the statistical error is the smallest, -

0.93%; In the case of incomplete data, the statistical error is -1.88%. Both results are better than the statistical error of deadweight tons, which is 3.34%.

5 PRACTICAL APPLICATION

After the actual shipment size is predicted by well-trained network, the AIS data is summarized to get the seaborne trade volume of iron ore in 2018:

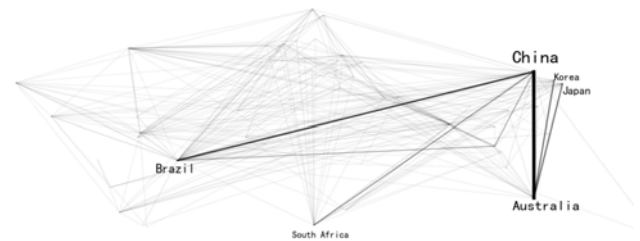


Figure 7. Iron ore circulation in 2018

(Line's transparency indicates the volume of the trade. The major trading countries are identified, and the font size of countries name are related to the volume of trade. Lines are links between countries where trade takes place, not vessels' tracks)

Table 3. The top ten trade routes

leavecountry		trade volume (million tons)		arrivecountry
Australia	→	715	→	China
Brazil	→	222	→	China
Australia	→	71	→	Japan
Australia	→	56	→	Korea
South Africa	→	32	→	China
Malaysia	→	29	→	China
Brazil	→	23	→	Malaysia
Brazil	→	18	→	Netherlands
Australia	→	17	→	Taiwan,China
Brazil	→	17	→	Japan

Table 4. The top 10 countries in terms of net exports and net imports

Range	exporter	net exports(million tons)	importer	net imports(million tons)
No.1	Australia	876	China	1070
No.2	Brazil	368	Japan	103
No.3	South Africa	51	Korea	75
No.4	Canada	37	Netherlands	34
No.5	Chile	16	Taiwan, China	24
No.6	Norway	16	Oman	15
No.7	Peru	15	France	13
No.8	Malaysia	11	Germany	11
No.9	Iran	10	United Kingdom	9
No.10	Mozambique	4	Bahrain	8

To verify the accuracy of the data, the import volume of China (mainland) is compared with the customs statistics. According to customs statistics, the annual import volume is 1064.78 million tons, and the statistical error with AIS-based statistics is 0.5%. Figure8 shows the monthly statistics comparison.

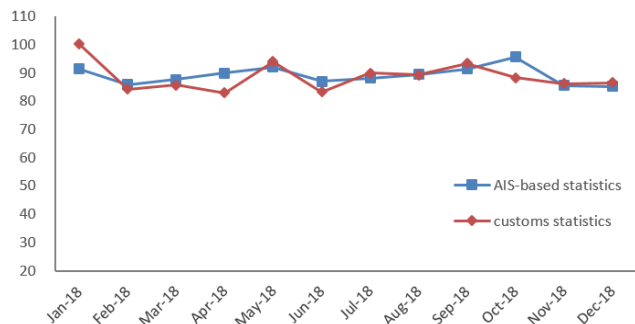


Figure 8. China's monthly iron ore imports in 2018

The primary reasons for the difference of two statistical results:

- 1 In the statistics based on AIS data, once a ship is berthed, it is included in the import statistics. The time is slightly different from the customs statistics time.
- 2 The dropout of AIS signal.

6 CONCLUSION

This paper proposes a shipment size of iron ore estimation method based on AIS data and BP neural network. 1650 pieces of data were used in the neural network for training and testing. Weekly AIS-based statistics of Australia-China trade volume in February in four scenarios is done to show the superiority of the proposed method. Then, iron ore carriers' AIS data is put into the well-trained network to obtain the shipment size of them. And the global iron ore circulation in 2018 is published by the AIS-based trade volume statistics method. The import volume of China is selected for comparison with the customs statistics data, and the annual statistical error is less than 0.5%.

This study can improve the accuracy of seaborne trade volume statistics. It has practical significance for relevant departments, companies and individuals to predict the market and make decisions. This paper has realized the prediction of the shipment size of ships, and the results are satisfactory. The next step will be aimed at the prediction of the cargo category and the corresponding volume, in order to play a more detailed and effective reference role for the actual decision-making.

REFERENCES:

- [1] Fu xiaoqi, Xie wen, Zheng gui-huan, et al. Forecast and analysis of China's import and export in 2006[J]. management review, 2006, 18(1):24-27+65.
- [2] Chen wei. Import and export trade prediction based on linear ARIMA and nonlinear BP neural network combination model [J]. Statistics and decision-making, 2015(2):47-49.
- [3] Liu xianfeng. Prediction of oil trade volume from 2014 to 2017 based on wavelet analysis and ARIMA combination model [J]. Theory and practice of finance and economics, 2014(4):117-121.
- [4] Nossum, B. "The evolution of dry bulk shipping, 1945-1990". Self-published. Oslo.1996.
- [5] Haji S , O'Keeffe E , Smith T . Estimating the global container shipping network using data and models[J]. Estimating the Global Container Shipping Network Using Data & Models.
- [6] Adland R, Jia H, Strandenes S P. Are AIS-based trade volume estimates reliable? The case of crude oil exports[J]. Maritime Policy & Management, 2017, 44(1):1-9.
- [7] Ren jie, Zhang ao, Yan liping, et al. Research on the status identification of inland river ships [J]. Instrumentation technology, 2017(6):38-40.
- [8] ShangHai Maili Marine Technology Co.,Ltd.<http://www.hifleet.com>
- [9] Hu X, Lin C. A Preliminary Study on Targets Association Algorithm of Radar and AIS Using BP Neural Network ☆[J]. Procedia Engineering, 2011, 15:1441-1445.
- [10] Gan S, Liang S, Li K, et al. Ship trajectory prediction for intelligent traffic management using clustering and ANN[C]// Ukacc, International Conference on Control.
- [11] Zhen rong, Jin yongxing, Hu qinyou, et al. Vessel Behavior Prediction Based on AIS Data and BP Neural Network[J]. China maritime industry, 2017, 40(2):6-10.
- [12] Zhu jin-shan, Sun li-cheng, Yin jian-chuan, et al. Model and simulation of ship signal recognition based on BP neural network [J]. Journal of applied science and engineering, 2012, 20(3):455-463.
- [13] Zhang y w, cui w b, wu g t, et al. Improvement analysis of BP neural network for ship remote monitoring system [J]. China maritime, 2009, 32(2):14-19. (in Chinese)
- [14] Yu tao. BP network adaptive learning rate algorithm analysis [D]. Dalian university of technology, 2011.
- [15] Zhang qingqing, he xingshi. Improved method for node selection of BP neural network and its application [J]. Journal of xi 'an university of technology, 2008, 22(4):502-505.
- [16] Wang xiaochuan. Analysis of 43 cases of MATLAB neural network [M]. Beijing university of aeronautics and astronautics press, 2013.