Prediction of Chinese Port Cargo Volume Based on BP Algorithm

Chenwei Zhong *
College of Transport and Communications, Shanghai Maritime University, Shanghai, 201306, China.

*Corresponding author e-mail: 202030610155@stu.shmtu.edu.cn.

Abstract. With the implementation of maritime power strategy, China attaches more and more importance to port economy. As a powerful indicator to measure Marine economy, port cargo volume can reflect whether the effectiveness of port construction in China is remarkable at the present stage. In this study, the author takes the cargo volume data of China's ports from 2010 to 2017 as the training data, and predicts the cargo volume of a January in 2018 and 2019 based on the time series method combined with BP neural network to determine the feasibility of the prediction model.

Keywords: Predict; Time series method; BP neural network solving speed; Prediction model.

1. Research background and research methods

1.1. The research background
Since China became the world's largest trading country in 2013, more than 90% of China's foreign trade goods are completed by water transportation, and the port collection and distribution system is the core system of cargo transportation. Therefore, port construction has always been one of the main tasks of China's maritime power construction, and a port's cargo carrying capacity is too high for port throughput expectation, which will lead to idle infrastructure in the port and waste of investment funds. If the port throughput forecast value is too small, it will cause the backlog of goods at the port, affect the delivery of goods by suppliers, and then affect the customer satisfaction. Only reasonable prediction can make the comprehensive and efficient development of ports and reasonable and sustainable construction. How to carry out reasonable port construction and layout and determine the port development strategy are the tasks to be undertaken by the port throughput prediction, so the port throughput prediction is of great significance.

Tang Fei (2015) built a BP neural network model and made an example prediction of Luzhou Port by taking years as a unit and using time series. Xu Jinhe (2009) predicted the throughput of Taicang Port before 2015 by using the cubic exponential smoothing forecasting method, and put forward suggestions for the further development of Taicang Port on this basis. Liu Changjian and Zhang Qingnian (2007) used BP neural network model to forecast containers and found that this processing method could forecast throughput in a relatively short period. Xie Xiaoshan (2007) combined genetic
algorithm and BP neural network to forecast the train passenger volume and compared it with various methods.

1.2. The research methods

Therefore, this study draws on the experience of predecessors and adopts BP neural network model to forecast the cargo volume of China's ports. The training and test data are shown in Table 1 below.

**Table 1. Statistics of China's port cargo volume from 2010 to 2019**

| Unit: 10,000 tons | 2010 | 2011 | 2012 | 2013 | 2014 | 2015 | 2016 | 2017 | 2018 | 2019 |
|-------------------|------|------|------|------|------|------|------|------|------|------|
| Jan.              | 28779| 31169| 32473| 37910| 44888| 49119| 48668| 52244| 52323| 57264|
| Feb.              | 23640| 27654| 32245| 33146| 39662| 42051| 42519| 46482| 46587| 49404|
| Mar.              | 29421| 34637| 36647| 38068| 45883| 46316| 47518| 50688| 51215| 57183|
| Apr.              | 30514| 35602| 39192| 41492| 50170| 48898| 50300| 54728| 55909| 59910|
| May.              | 31303| 36469| 40476| 42282| 50109| 52116| 52829| 58813| 58444| 61782|
| Jun.              | 31174| 35553| 38762| 42266| 51856| 53649| 54900| 59054| 61995| 64948|
| Jul.              | 30659| 36761| 36843| 43778| 51082| 53590| 53428| 57353| 59289| 64115|
| Aug.              | 31900| 36841| 37988| 41933| 52088| 54164| 54230| 57583| 62075| 64106|
| Sept.             | 32281| 37528| 39796| 42863| 52438| 55069| 55010| 57891| 63587| 67433|
| Oct.              | 33810| 37734| 40337| 43037| 52693| 56359| 57593| 59985| 64022| 67093|
| Nov.              | 32831| 37013| 41219| 44729| 53933| 57083| 60856| 60606| 65102| 67778|
| Dec.              | 33810| 37004| 40537| 43337| 52600| 56005| 57003| 53292| 64022| 66210|

In this study, the data of 10 years were divided into 10 years, 10 to 17 years as training data, 18 and 19 years as testing data, January to November data as input values of neural network, and December data as output values. The network is predicted, and the function library of MATLAB platform is used to call BP neural network model to train and predict the data, and the number of hidden layers is adjusted according to the final error.

2. Introduction to BP neural network

**Figure 1. Topological structure diagram of three-layer BP neural network**

The process of BP neural network is mainly divided into two stages. The first stage is the forward propagation of signal, from the input layer through the hidden layer, and finally to the output layer. The second stage is the back propagation of error. From the output layer to the hidden layer and finally to the input layer, the weight and bias of the hidden layer to the output layer are successively adjusted. The weight and bias of the input layer to the hidden layer are shown in the figure, which is a BP neural network with three-layer structure.

There are a few important formulas:

1) Initialization of the network
Assume that the number of nodes in the input layer is $n$, the number of nodes in the hidden layer is $l$, and the number of nodes in the output layer is $m$. The weight of input layer to hidden layer $w_{ij}$, the weight of the hidden layer to the output layer is $w_{jk}$, the bias of the input layer to the hidden layer is $a_j$, and the bias of the hidden layer to the output layer is $b_k$. The learning rate is $\eta$ and the excitation function is $g(x)$. Wherein, the excitation function is Sigmoid function.

In the form of:

$$g(x) = \frac{1}{1 + e^{-x}}$$  \hspace{1cm} (1)

2) Output of hidden layer

As shown in the three-layer BP network above, the output of the hidden layer $H_j$ is:

$$H_j = g\left(\sum_{i=1}^{n} w_{ij} x_i + a_j\right)$$  \hspace{1cm} (2)

3) Output of the output layer

$$O_k = \sum_{j=1}^{l} H_j w_{jk} + b_k$$  \hspace{1cm} (3)

4) Error calculation

$$E = \frac{1}{2} \sum_{k=1}^{m} (T_k - O_k)^2$$  \hspace{1cm} (4)

Where $T_k$ is the expected output, as shown in the above formula $i = 1 \cdots n$, $j = 1 \cdots l$, $k = 1 \cdots m$. Remember $e_k = T_k - O_k$.

5) Update the weight

The updated formula of weight is as follows:

$$w_{ij} = w_{ij} + \eta H_j (1 - H_j) x_i \sum_{k=1}^{m} (w_{jk} e_k)$$  \hspace{1cm} (5)

$$w_{jk} = w_{jk} + \eta H_j e_k$$  \hspace{1cm} (6)

6) Update the offset

The updated formula of bias is as follows:

$$a_j = a_j + \eta H_j (1 - H_j) \sum_{k=1}^{m} (w_{jk} e_k)$$  \hspace{1cm} (7)

$$b_k = b_k + \eta e_k$$  \hspace{1cm} (8)

3. Model construction and solution

3.1. Model building

1) Determination of the number of network layers

In a typical BP neural network model, the input and output are all one layer. In particular, the number of layers of the hidden layer should be selected according to the actual situation. Therefore, the number of layers of the network and the number of hidden layers are closely related. Although the number of hidden layers is proportional to the nonlinear mapping ability of BP neural network, excessive number of hidden layers will complicate the network structure and affect the training speed and effect of the network. The study found that a single hidden layer of multiple neuron nodes can realize the mapping from multidimensional input to multidimensional output. Therefore, a single hidden layer can fully meet a variety of research applications, so a three-layer BP neural network was selected for the study.

2) Number of input layer nodes

According to my training requirements, the input layer node is the cargo volume from January to November every year, that is, the input layer node is 11.

3) Number of nodes in the output layer
This prediction is to predict the freight volume of a single month (that is, December), so the node of the output layer is 1.

4) Number of hidden layer nodes

The number of neurons in hidden layer nodes selection more difficult than the above two layers, often need to according to the experience and experimented to draw, so there will not be an exact function expression to calculate, but there is a certain number of hidden layer nodes and input and output layer, hidden layer node number does not represent a network effect is good, too much excessive number of nodes will make the network training process is long. Result is not necessarily the most ideal expectations and tend to use the experience and trial and error method, through the empirical formula \( M = \sqrt{m + n} \) (M, n number of input layer and output layer nodes, respectively) calculate a rough range to determine the number of nodes, then based on the number of nodes in the network was trained to find the minimum error is compared with the number of nodes in the network structure. After repeated training, it is found that the error of 10 hidden layers is small, so the hidden layer node is selected as 10.

### 3.2. Implementation of MATLAB algorithm

The matlab code is as follows:

```matlab
clc
clear
data=xlsread('C:\Users\zcw\Desktop\matlab\data2.xls');
input=data(1:11,:);
output=data(12:end,:);
k=rand(1,10);
[m,n]=sort(k);
% set training value and test value
input_train=input(:,n(1:8));
output_train=output(:,n(1:8));
input_test=input(:,n(9:end));
output_test=output(:,n(9:end));
% data normalization
[inputn,inputs]=mapminmax(input_train);
[outputn,outputs]=mapminmax(output_train);
inputn_test=mapminmax('apply',input_test,inputs);
% establishment of neural network
net=newff(inputn,outputn,10,{'tansig','purelin'},'trainlm');% toolbox functions directly referenced
net.trainParam.epochs=5000; % maximum training times
net.trainParam.goal=0.05;% target minimum error
net.trainParam.lr=0.05;% Learning rate
net=train(net,inputn,outputn);% start training
BPoutput=net(inputn_test);% data in
prediction=mapminmax('reverse',BPoutput,outputs);% inverse normalization of output data
error = abs(prediction - output_test)/output_test;% calculating relative error
% Drawing
figure(1)
plot(output_test,'-*b') ;hold on
plot(prediction,'-o','color','y');
legend(' Act. ',' Est. ');
title('BP prediction chart');
```

After Matlab running, as shown in the figure below, one iteration converges.
Then through the test of the test set, we can get the prediction chart as shown in the figure below. We can observe that the error is basically controlled at about 5%, which basically meets our expected 5% error requirement.

Figure 2. Matlab operation diagram

After several times of operation, it is found that it can be controlled at about 5%, but there are still large errors, as shown in the figure below.

Figure 3. Test error analysis

After several times of operation, it is found that it can be controlled at about 5%, but there are still large errors, as shown in the figure below.

Figure 4. Prediction chart of 6 runs
4. Conclusion
In this paper, MATLAB software is used to establish BP artificial neural network to predict the cargo throughput of China's ports. It is found that the prediction error can be basically controlled within 5%, but there are still large errors. Through the study of literature, it is found that the initial weight and threshold setting of BP neural network has great randomness, which will affect the accuracy of training. Therefore, further research on how to modify the initial weights and thresholds of BP neural network may be carried out in the future.

There are many factors affecting the port freight volume. This study only forecasts the data of previous years, that is, the time series method, without considering other factors, so the study has some limitations.

References
[1] Yang Ke. Application of BP neural network optimized by genetic algorithm in Lianyungang Port Throughput Prediction [D]. Shenzhen University, 2017.
[2] Tang Fei. Prediction of Luzhou port cargo throughput based on BP neural network [J]. Logistics engineering and management, 2015,09:58-59 + 33.
[3] Xu Jinhe. Application of cubic exponential smoothing method in throughput prediction of Taicang port [J]. Logistics engineering and management, 2009,06:53-54 + 62.
[4] Liu Changjian, Zhang Qingnian. Dynamic prediction of container throughput based on time series BP neural network [J]. Water transport engineering, 2007,01:4-7 + 11.
[5] Cargo and passenger throughput of ports above Designated Size in 2019. Ministry of transport. [EB / OL]. (May 13, 2019).
[6] http://xxgk.mot.gov.cn/jigou/zhghs/201905/t20190513_3198921.html.
[7] Chen Ming. Matlab neural network principle and example refinement [M]. Beijing: Tsinghua University Press, 2013.