ISSN: 2710-0170

Container throughput forecasting of Nansha Port based on ARIMA-RBF

Wenxian Wang, Jian Zhang *
School of Transportation, Wuyi University, Jiangmen 529020, China.
* Corresponding Author

Abstract

Predicting container throughput is fundamental for port management and the scheduling of handling equipment. Based on an analysis of the mechanisms of the ARIMA model and the RBF model, this paper investigates the daily container throughput patterns and data stationarity characteristics of Nansha Port in Guangzhou, using survey results. By integrating the time series forecasting capability of the ARIMA model with the nonlinear processing ability of the RBF neural network, an ARIMA-RBF combined forecasting model is established to predict the container throughput of Nansha Port. This model accounts for both the linear and nonlinear characteristics of port container throughput and demonstrates superior predictive performance compared to the traditional ARIMA forecasting model.

Keywords

Port container throughput, ARIMA-RBF combined forecasting model, Nansha Port.

1. Introduction

Port throughput reflects the status of international trade, and as hubs of global commerce, ports are playing increasingly important roles. Local governments have been consistently increasing investments in port infrastructure, especially in container terminals. Accurate container port throughput forecasts hold significant reference value for port operators and are crucial for their decision-making processes.

As the most important metric for ranking ports, container throughput prediction has been a focal point of research among scholars in port management. Time series methods are commonly employed, with the Autoregressive Integrated Moving Average (ARIMA) model and its variants being widely used across various fields. Pan Changhong [1] used the coefficient generation method to predict the demand for empty containers in a region and modified the ARMA model. Liu Yu [2] improved prediction accuracy by using ARIMA and support vector regression for different parameter vectors based on their fluctuations, then combining these forecasts. Rao Changrui [3] decomposed several intrinsic mode functions (IMFs) using EMD technology, followed by forecasting with ARIMA and SVR and integrating the results. Zhang Peng [4] employed wavelet decomposition on raw data and utilized deep learning models to adjust parameters and model structures to obtain the forecasting model. Kong Linlin et al. [5] proposed an ARIMA model for port container throughput forecasting. Rashed et al. [6] constructed an ARIMA model to forecast container throughput at the Port of Antwerp. Min K C et al. [7] used a Seasonal Autoregressive Integrated Moving Average (SARIMA) model for container throughput forecasting. Juan Huang [8] conducted a comparative analysis of various univariate forecasting methods, including seasonal dummy variable regression, hybrid grey forecasting model, and SARIMA, finding SARIMA most suitable for seasonal container throughput forecasting.

ISSN: 2710-0170

Container throughput at ports exhibits regular variations throughout the year, displaying certain time series characteristics with high similarity in traffic trends within each cycle. Additionally, container throughput is influenced by factors such as climate and unexpected events, introducing a degree of non-stationarity. Addressing these characteristics, this study adopts a combined forecasting model that integrates the ARIMA model with the Radial Basis Function (RBF) neural network. By analyzing the time series characteristics and stationarity, the ARIMA model is constructed to calculate the forecast errors between predicted and actual port container throughput. The RBF neural network model is then used to fit and predict these errors, resulting in the final forecast of port container throughput.

2. ARIMA-RBF forecasting model

2.1. ARIMA-RBF forecasting model

Time series models utilize internal information to extrapolate trends for forecasting purposes. Classic time series models include the following types:

1. Autoregressive Model AR(p)

The mechanism of the Autoregressive Model AR(p) involves predicting the current or future values based on historical data over p periods. The key lies in determining the value of the order p. The calculation formula for this model is as follows:

$$y_t = \mu + \sum_{i=1}^{p} \gamma_i y_{t-i} + \varepsilon_t \tag{1}$$

Where y_t represents the original time series, μ represents a constant, p represents the order of the autoregressive model, γ_i represents the coefficients, and ε_t represents the error between the actual and predicted values of the time series.

2. Moving Average Model MA(q)

The mechanism of the Moving Average Model MA(q) involves using accumulated error terms to represent error fluctuations, and on this basis, predicting current or future values. The calculation formula for this model is as follows:

$$y_t = \mu + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i} + \varepsilon_t \tag{2}$$

Where p represents the order of the moving average model, γ_i represents the coefficients.

3. Autoregressive Moving Average Model ARMA(p,q)

The Autoregressive Moving Average Model ARMA(p,q) can be seen as a combination of the Autoregressive Model AR(p) and the Moving Average Model MA(q). The calculation formula for this model is as follows:

$$y_t = \mu + \sum_{i=1}^{p} \gamma_i y_{t-i} + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i} + \varepsilon_t \tag{3}$$

It can be observed that as an extension of the ARMA(p,q) model, the ARIMA(p,d,q) model first transforms the original time series into a stationary time series through d-order differencing, and then uses the ARMA(p,q) model to model the differenced series.

2.2. RBF neural network

The structure of the Radial Basis Function (RBF) neural network comprises three layers: the input layer, the hidden layer, and the output layer. The topology of this structure is illustrated in Figure 1. In this network, the input layer transmits the input data to a high-dimensional space. The hidden layer space is constituted by the basis of the high-dimensional space, calculating the distance between the hidden layer and the input layer, which is then used as the input to the radial basis function. The output results are linearly optimized and weighted to form the output layer, thus avoiding the problem of local minima. The radial basis function is generally a Gaussian function, with the calculation formula as follows:

$$g(X_t - d_i) = \exp(-\frac{\|X_t - d_i\|^2}{2\sigma^2}) \tag{4}$$

The functional relationship from the input layer to the output layer is calculated as follows:

$$G_k(X_t) = \sum_{i=1}^{n} w_{ik} \exp(-\frac{\|X_t - d_i\|^2}{2\sigma^2}) \tag{5}$$

In the above formulas, $\|\cdot\|$ denotes the Euclidean norm, n represents the number of nodes in the hidden layer, X_t represents the training set of the t-th input sample, $X_t = (x_{t1}, x_{t2}, \ldots, x_{tm})^T$, $X_t \in R^m$. m represents the sample size. d_i is the center of the hidden layer nodes, $d_i \in R^m$. w_{ik} represents the connection weights from the hidden layer to the output layer, with n being the number of nodes in the output layer.

The output value of sample X_t is defined as $G(X_t) = [G_1(X_t), G_2(X_t), \ldots, G_n(X_t)]^T$. Let the expected output value of X_t be Z_t. Thus, the mean square performance index σ is calculated as follows:

$$\sigma = \frac{1}{M}\sum_{t=1}^{n} \| Y_t - G(X_t) d_i \|^2 \tag{6}$$

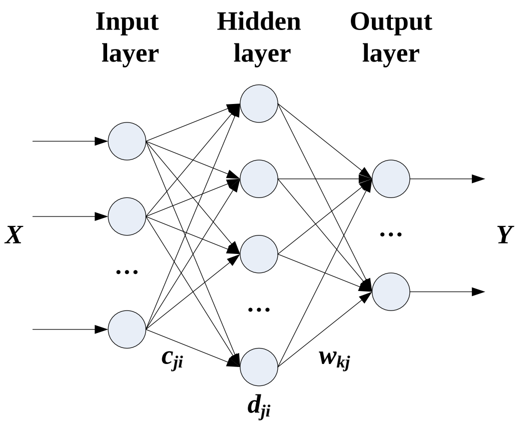

Figure 1: Schematic diagram of RBF neural network structure

2.3. ARIMA-RBF combined forecasting model

Combining the linear and nonlinear characteristics of container throughput variation at Nansha Port, the ARIMA model is used to capture the autocorrelation and partial autocorrelation within the series for traffic flow prediction. Additionally, the RBF neural network is employed to characterize the nonlinear features of the container throughput time series at Nansha Port. The specific steps of the model are as follows:

Step 1. Use the ARIMA model to perform linear predictions on the container throughput time series at Nansha Port.

Step 2. Use the RBF neural network model to perform nonlinear predictions on the sequence of errors between the ARIMA model's predictions and the actual container throughput.

Step 3. Sum the prediction results of the two models to obtain the final container throughput forecast.

Based on the above analysis, the ARIMA-RBF combined forecasting model for container throughput at Nansha Port is expressed as follows:

ISSN: 2710-0170

$$\begin{cases} y_t = L_t + N_t \\ e_t = y_t - \hat{L}_t \\ \hat{N}_t = f(e_t) \\ \hat{y}_t = \hat{L}_t + \hat{N}_t \end{cases} \tag{7}$$

Where L_t represents the linear component of the container throughput time series, and N_t represents the nonlinear component of the container throughput time series. $\hat{L}_t$ is the prediction value of the ARIMA model at time *t*. e_t is the prediction error. *f* is the nonlinear function fitted by the neural network, $\hat{N}$ is the prediction result of the RBF neural network. $\hat{y}_t$ is the container throughput forecast value of the combined model at time *t*.

3. Case study

3.1. Throughput data analysis

Based on the container throughput statistics at Guangzhou Nansha Port from January 1 to December 31, 2022, the corresponding time series is depicted in Figure 2. The figure shows that the container throughput at Nansha Port exhibits noticeable periodic fluctuations. Additionally, it can be observed that the throughput generally fluctuates around values between 350,000 and 450,000, with no significant seasonal variation.

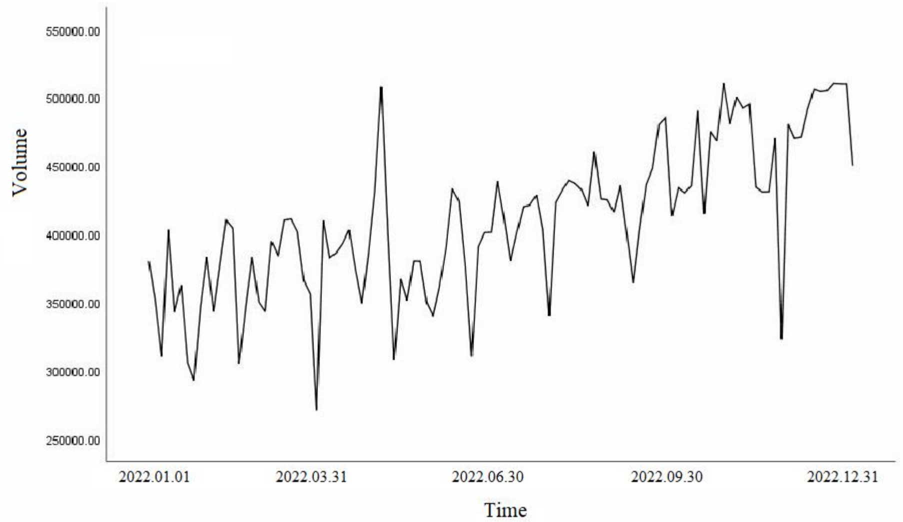

Figure 2: Time series of container throughput at Nansha Port in 2022

3.2. Data stationarity test

Using Eviews software, the t-statistic and the probabilities of various critical values for the container throughput time series at Nansha Port were calculated, with the specific results shown in the table below.

Table 1: Unit root test

	t-statistic	p-probability
ADF	-5.316	0.02
Significance levels 1% level	-3.872	
Significance levels 5% level	-3.129	
Significance levels 10% level	-3.042	

ISSN: 2710-0170

The unit root test statistic ADF = -5.316 is less than the ADF critical values at the 1%, 5%, and 10% significance levels set by Eviews. Additionally, the probability is 0.02, indicating that the null hypothesis of a unit root can be rejected. Therefore, it can be concluded that the container throughput time series is stationary.

3.3. ARIMA model Identification

The identification of the ARIMA model essentially involves model order determination, specifically the selection of the values for the parameters (p,d,q). In this study, values of p and q ranging from 0 to 6 were examined using the AIC (Akaike Information Criterion) test to determine the appropriate values. According to the calculation results from Eviews software, when p=4 and q=3, the AIC value is minimized at 14.92. Given that the number of differences d=0, the basic model for the container throughput time series at Nansha Port can be preliminarily determined as ARIMA(4,0,3). Further validation is required using the BIC (Bayesian Information Criterion) and the adjusted R-squared value.

The fitting degree, adjusted R-squared, and BIC statistical values of the model are shown in Table 2. When the parameters p=4 and q=3, both the adjusted R-squared and BIC values are optimal. Therefore, the ARIMA(4,0,3) model is finally determined to be the basic model.

Table 2: Statistical results of ARIMA(4,0,3) model

Model suitability statistics		Ljung-Box Q (18)			
Stable R-square	BIC	Statistics	DF	Significance	Group values
0.361	12.096	13.674	11	0.246	0

3.4. ARIMA-RBF model forecasting

In this study, the container throughput data of Nansha Port for the first 280 days of 2022 was selected as the training sample, and the residual data for 84 days was used as the test sample to predict the residuals for the next 10 days. The Mean Absolute Percentage Error (MAPE) index was used to measure the accuracy of the prediction results, calculated as follows:

$$MAPE = \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times 100 \tag{8}$$

The prediction results and MAPE values for the container throughput from December 24 to December 30, obtained using the ARIMA basic model and the ARIMA-RBF combined model, are shown in Table 3.

Table 3: Comparison of ARIMA and ARIMA-RBF forecast results

Date	Actual throughput	ARIMA Forecast results	MAPE	ARIMA-RBF Forecast results	MAPE
12.24	452221	445463	1.49%	448004	0.93%
12.25	453587	439473	3.11%	441136	2.75%
12.26	453701	438559	3.34%	440392	2.93%
12.27	449485	424358	5.59%	436983	2.78%
12.28	444177	419128	5.63%	428203	3.60%
12.29	445901	428326	3.94%	430364	3.48%
12.30	443146	429281	3.12%	431883	2.54%

ISSN: 2710-0170

As can be seen, the average absolute percentage error of the container throughput predictions for Nansha Port obtained using the ARIMA model is 3.75%, which is higher than the error of 2.72% obtained using the ARIMA-RBF combined model. This indicates that the ARIMA-RBF combined model provides better predictive performance.

4. Conclusion

This paper constructs an ARIMA-RBF combined model to predict the container throughput at Guangzhou Nansha Port. This method addresses the linear patterns in container throughput variation by first fitting an ARIMA model. Then, to capture the nonlinear characteristics of the ARIMA prediction errors, an RBF neural network model is used. This approach simultaneously accounts for both the linear and nonlinear features of container throughput variations at Nansha Port. According to the actual prediction results of the ARIMA-RBF combined model for Nansha Port's container throughput, this combined forecasting model significantly reduces prediction errors and demonstrates good applicability in the field of routine container throughput forecasting at ports.

Acknowledgements

General Projects of the Water Transport Economy Research Center 2023 (SYJJ2023B02) and Jiangmen science and technology planning project (No.2021030101730004367 & No. 2022030100950003858).

References

[1] C.H. PAN, H.M FAN. Regional empty container demand forecast model based on the production coefficient method (Journal of Dalian Maritime University), Vol. 38 (2012) No.1, p.45-46.

[2] Y. LIU:. *Container throughput forecasting based on hybrid model of VMD-ARIMA-HGWO- SVR* (Ph. MS., Lanzhou University, China 2018), P.38.

[3] C.R. RAO: *Container Throughput Prediction Paradigm with Major Events* (Ph. MS., Beijing Jiaotong University, China 2021), P.35.

[4] P. ZHANG: *Tianjin Port Container Throughput Forecasting Based on Deep Learning* (Ph. MS., Tianjin University of Technology, China 2021), P.27.

[5] L.L. KONG, L. LIU, W.X. XU et al: Prediction of the container throughput in a port based on time sequence analysis (Forest Engineering), Vol. 32 (2016) No.5, p.106-110.

[6] Y. Rashed: Modeling of Container Through put in Northern Adriatic Ports Over the Period 1990-2013(Journal of Transport Geography), (2016) No.52, p.131-142.

[7] K.C. Min, H.K. Ha. Forecasting the Koreas Port Container Volumes With SARIMA Model (Physica), Vol. 32 (2014), No.6, p.600-614.

[8] J. Huang. Container throughput forecasting for inter-national ports in Taiwan (Journal of Marine Science and Technology), Vol.28, (2020) No.5, p.456-469