Advances in Economics, Business and Management Research, volume 68

International Symposium on Social Science and Management Innovation (SSMI 2018)

Forecasting Port Throughput Model of Tianjin Port based on Wavelet Analysis and Machine Learning

Bingchun Liu [a], Shaofeng Feng [b, *]

School of Management, Tianjin University of Technology, Tianjin, 300384, P.R. China.

[a]375887180@qq.com, [b]sffeng@hotmail.com

Abstract. With the development of national economy, the importance of port growing port is the national foreign trade portal, is the driving force for the development of the city, with the improvement of freight volume, how many directly affects the port cargo throughput of port layout planning, therefore, to effectively predict the Tianjin port cargo throughput is very important to the correct port development policy. This article embarks from the analysis of Tianjin port cargo throughput changes year by year, from 2001 to 2017, Tianjin port cargo throughput are summarized factors index, prediction model was established based on seven kinds of machine learning algorithms and wavelet analysis, in Tianjin, for example, through the training data model, and then the model prediction data were compared with the actual data analysis, weighing error, select the optimal model, it is concluded that the prediction data and the actual data, the basic of Tianjin port cargo throughput forecast has positive significance.

Keywords: Tianjin port, cargo throughput, machine learning, forecasting, wavelet analysis.

1. Introduction

Tianjin port is an important support for the coordinated development of Beijing-Tianjin-Hebei region and the construction of the north international shipping core zone. It is one of the most convenient ports in Xiongan new area, an important strategic fulcrum of "One Belt and One Road", an important link connecting northeast Asia with central and western Asia, and an important part of Tianjin pilot free trade zone. In 2017, the cargo throughput of Tianjin port was 500 million tons, and the container throughput exceeded 15 million TEUs. It is of great significance to study the cargo throughput of Tianjin port. One is to serve the national strategy. The implementation of the Beijing-Tianjin-Hebei coordinated development strategy urgently requires Tianjin port to constantly improve the collection and transportation system and promote the efficient and coordinated development of the port transportation system. We will establish an efficient and modern logistics system centered on ports and further strengthen the role of port transportation as a hub in global supply chains and trade chain networks. Second, we will improve the business environment and build a convenient, efficient, green and intelligent logistics system. Third, promote the development of port logistics industry. Port logistics development has experienced cost concept to profit idea and comprehensive logistics service concept, in the pursuit of high efficiency and low cost at the same time, strengthen the customer service, improving customer experience, provide to the customer as the center of the complete logistics service, so as to speed up the depth of the port logistics and information technology integration, promote the development of container logistics, e-commerce, realize resources sharing and information connectivity. Only through scientific and reasonable prediction of the cargo throughput of Tianjin port can resources be allocated more effectively, and the operation and management of Tianjin port can be intelligentized, trade and logistics facilitation and innovation and sharing ecology can be realized.

2. The Wavelet Analysis

As an analytical tool which can adjust the size of analysis window automatically with the change of frequency, wavelet transform is applied in many fields such as signal processing, computer vision, image processing, speech analysis and synthesis. It mainly includes continuous wavelet transform, discretization of wavelet change, multi-resolution analysis and Mallat algorithm. Its properties include linearity, translation invariance, expansion covariance and redundancy.

The wavelet transform itself has the characteristics of non-stationarity, non-linearity and high signal-to-noise ratio. The wavelet theory is developed according to the demand of time-frequency localization and has adaptive and mathematical microscopic properties, especially suitable for non-stationary and non-linear signal processing.

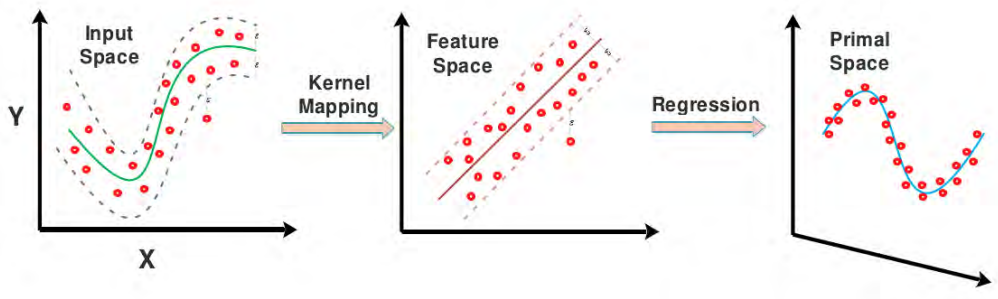
The principle of using wavelet for data prediction is as follows: firstly, Mallat algorithm is used to decompose the data, and the data after decomposition is smoothed. After that, the reconstructed data becomes a stable time series of approximate significance, and then the approximate signal of the original data is obtained. Finally, the prediction model is used for time series prediction, such as AR, MA, ARMA, etc.

3. Forecasting Port Throughput Model

According to the statistics of cargo throughput of Tianjin port, this paper estimates the cargo throughput of Tianjin port by building different prediction models. In this paper, seven machine learning evaluation methods are used to compare the ability of the prediction method, so as to select the prediction value with more practical reference significance.

From 2001 to 2017 of Tianjin port cargo throughput is increasing year by year and remain stable on the whole, consider to choose the SVR, RDG, reg, las, dt, GBR and KRDG prediction model of Tianjin port cargo throughput forecast, and through the pros and cons of each evaluation model to predict the error, more observation meaningful to compare the model data and lay the foundation.

SVR: Support vector regression (SVR) is a nonlinear regression model based on statistical learning theory, which can well capture the nonlinear modes hidden in the original data, so SVR model can be used to predict the high-frequency and nonlinear components in the container throughput series. Vapnik established support vector machines (SVM) based on structural risk minimization (SRM) principle in the 1990s, and applied the idea of SVM to solve classification problems to regression problems, and then the SVR model was obtained. x SVR is the basic ideas of the original data by nonlinear function ϕ F mapped to higher dimension feature space, and then the linear regression in the feature space.



By using the insensitive loss function and introducing relaxation variables and tantalum, the SRM problem in SVR can be described as follows:

$$\begin{cases} \min\limits_{\omega,b,\xi_i,\hat{\xi}_i} \left\{ \frac{1}{2}\|\omega\|^2 + C\sum\limits_{i=1}^{m}(\xi_i + \hat{\xi}_i) \right\} \\ s.t.\ f(x_i) - y_i \le \varepsilon + \xi_i \\ y_i - f(x_i) \le \varepsilon + \hat{\xi}_i \\ \xi_i \ge 0, \hat{\xi}_i \ge 0, i = 1,2,\cdots,m. \end{cases}$$

ε is a pre-specified parameter, C is a regularization parameter, indicates the tradeoff between training error and model complexity, $f(xi) = \omega^{T}\phi(xi) + b$. After introducing the Lagrange multiplier αi, $\hat{\alpha} i$ and the kernel function K, the equation can be written as:

$$f(x_i) = \sum_{j=1}^{m} (\hat{a}_j - \alpha_j) K(x_i, x_j) + b$$

In this paper, we choose the usual Gauss kernel function:

$$K(x_i, x_j) = exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right)$$

4. Raw Data and Processing

In order to verify the error between the predicted data and the actual data, the experimental data was divided into two parts: training data and test data. Among them, the first 80% of the data was used as training data for the training model, and the last 20% of the data was used as test data. Then, the port throughput of the 20 percent of the data is predicted by using the trained model, and the actual port throughput of the 20 percent of the data is compared with that of the actual port throughput, and the calculation error is used to analyze the advantages and disadvantages of the model.

Table 1. Throughput experimental data of Tianjin port from 2001 to 2017

year	Throughput of last year (million tons)	Tianjin GDP (million yuan)	Tianjin tertiary industry GDP(million yuan)	Tianjin total import and export of goods (million yuan)	Annual throughput (million tons)
2001	114.84	1919.09	881.3	1241.94	130.23
2002	130.23	2150.76	997.47	1558.82	163.45
2003	163.45	2578.03	1150.81	2005.77	208.72
2004	208.72	3110.97	1319.76	2869.51	244.06
2005	244.06	3905.64	1658.19	3645.79	257.23
2006	257.23	4462.74	1902.31	4409.66	317.5
2007	317.5	5252.76	2250.04	4886.13	359
2008	359	6719.01	2886.65	5500.00	380
2009	380	7521.85	3405.16	4366.74	421.04
2010	421.04	9224.46	4238.65	5613.49	453.35
2011	453.35	11307.28	5219.24	7060.57	476
2012	476	12893.88	6058.46	7895.88	568.3
2013	568.3	14442.01	6979.6	8777.18	539.64
2014	539.64	15726.93	7795.18	10984.21	540
2015	540	16538.19	8625.15	7804.37	551
2016	551	17885.39	10093.82	7010.37	501
2017	501	18595.38	10786.74	7646	

5. Model Performance Analysis

Through seven kinds of machine learning algorithms, respectively for 2001 to 2012 data for training, it is concluded that the training model, can find seven kind of model fitting sex is stronger, the minimum error may reach 0.2%, common error is about 4% or so, and then from training model predicts throughput from 2013 to 2016, and compared with the actual data, can be found that the minimum error is about 4.6%, SVR model has good prediction ability, and Tianjin port throughput.

Continue to analyze the above, however, we found that part of the machine learning algorithm for training the model prediction data error is bigger, even up to about 33%, we can clearly see that from 2001 to 2012, Tianjin port cargo throughput a steady upward trend, similar to the linear growth, but a higher growth in 2013, from 2014 to 2016, steady, and about 10%, a sharp fall in 2017 as a result

of the irregular data change of above parts of machine learning algorithm is the main reason for the error prediction results.

Table 2. Experimental training error and test error

year	train	test
svr	0.045635704	0.04590491
rdg	0.040199707	0.23497575
reg	0.040199663	0.33539357
las	0.047140742	0.25206155
dt		0.06781142
gbr	0.002419326	0.04606250
krdg	0.040954129	0.18317423

In 2017, according to the ministry of environmental protection, the port of Tianjin stopped receiving motor transport coal into the port. Bulk coal has been transferred to railway port in Tianjin port. Based on the continuous expansion of railway transport capacity, Tianjin port has strengthened cooperation with various ports in Hebei province to upgrade its business structure. During the period of transition, bulk cargo throughput fell sharply, at the same time, due to the original car in the course of carriage, round-trip transport bulk and ore respectively, the comprehensive transportation cost is low, after the bulk freight reduce, into a single transport ore, circular, increased the ore transport costs, also led to a decline in Tianjin port ore throughput.

6. Summary

For Tianjin port cargo throughput forecast, which is based on the previous year, the cargo throughput of the tertiary industry GDP, GDP in tianjin, tianjin and tianjin import and export of goods, total does not include consideration of major scientific and technological research and various policies and major accidents, so part of the machine learning model produce the bigger error, if use the same methods from 2001 to 2012 with data segmentation and forecast, the model can achieve higher precision of prediction. However, we also found that the model of Wavelet-SVR algorithm training, even though there are many influencing factors, still achieved a high prediction accuracy, which has a great reference significance for the prediction of cargo throughput of tianjin port.

7. Funding

This work was supported by the National Natural Science Foundation of China [grant number 71503180]; Major Project of Tianjin Education Committee and Social Science [grant number 2017JWZD16]; Tianjin science and technology project [18ZLZDZF00040].

References

[1]. Chen T T, Chen Y Y. Port Cargo Throughput Forecast Based on BP Neural Network[J]. Computer & Modernization, 2009.

[2]. Marks J R, Placone J, Hristova K, et al. Spontaneous Membrane- Translocating Peptides by Orthogonal High- Throughput Screening[J]. Journal of the American Chemical Society, 2011, 133(23):8995-9004.

[3]. Wang Bingdan. Prediction of container throughput based on SVM [D]. Beijing Jiaotong University, 2011.