Study on the Sub-period Forecast of the Chang Jiang River Trunk Freight Volume Based on Support Vector Machines

Jun Li
School of Engineering, Sun Yat-sen University
Guangzhou 510275, China
stslijun@mail.sysu.edu.cn

Licheng Wan
School of Engineering, Sun Yat-sen University
Guangzhou 510275, China
wanlch3@mail2.sysu.edu.cn

ABSTRACT
The accurate forecast of the freight volume on the Chang Jiang River trunk line is of great significance for the freight policies of the Chang Jiang River. However, the existing researches on the freight volume forecast are difficult to obtain the satisfactory results with only consideration of its non-natural influence factors. This study aims to develop a sub-period forecast model of the freight volume of the Chang Jiang River trunk line based on the support vector machines (SVM) in consideration of the both non-natural factors and natural factors affecting the freight volume. The low-water period and high-water period are calculated separately according to the actual periodic water condition in the Chang Jiang River. The results show that the accuracy of the sub-period forecast model is nearly doubled compared with the traditional forecast model, and the proposed method is practical for more accurate forecast of the Chang Jiang River trunk freight volume and can also be applied to other similar forecast scenarios.

KEYWORDS: Chang Jiang River; Trunk line; Freight volume; Sub-period forecast; Support vector machine

1 INTRODUCTION
The Chang Jiang River, as the largest river in China, plays an important role in the economic development in our country, especially in the riverside area. The optimal freight efficiency in the Chang Jiang River can not be separated from the reasonable freight transportation policy. The freight volume (FV) provides data reference for the formulation of shipping policy of the Chang Jiang River. So how to improve the forecast accuracy of the freight volume along the Chang Jiang River has become an urgent problem to be solved by the shipping department.

There are so many researches on the forecast of waterway freight. Most of the existing waterway freight volume forecast model are based on the non-natural influence factors and time series. Moscoso-López et al. (2016) deemed that advanced models for freight volume forecast are essential to improve the port level-service and competitiveness, and presented two forecast models to predict the freight volume. The models were developed and tested based on artificial neural networks and support vector machines. Jie et al. (2015) presented genetic-BP neural network model to forecast the freight volume of Chang Jiang River Ro-Ro transport. In addition, time-series models were applied to predict the waterway freight volume (for reviews, see e.g. Wan and Zhu, 2006; Song et al., 2017). Besides, Patil and Sahu (2016) developed the forecast model with additive regression and time series techniques. The gross domestic product and crude oil production were selected as economic indicators.
to explain the multivariate regression model and the result indicated that the regression model produced more optimistic forecasts than the time series model.

The existing models contained only the non-natural influence factors of freight volume. However, the freight volume is affected by various factors other than the non-natural ones (Zheng and Kim, 2017). The regression models with non-natural factors only are difficult to accurately reflect the trend and regularity of freight volume. In view of the above problems, the sub-period forecast model with consideration of water condition in Chang Jiang River, proposed in this paper, will be discussed in the following sections.

The impact of influence factors on freight volume is complex and it is difficult to use linear expression (Yang and Xu, 2016). In recent years, many methods based on machine learning have been applied to the freight volume forecast (Yin et al., 2016; Mrowczynska et al., 2017). Artificial neural network is a typical application (Maiti et al., 2012). Artificial neural network can deal with the nonlinear functions, and the prediction results are generally better than the traditional methods (Chen et al., 2003). In the 1990s, support vector machines became an important method to solve nonlinear problems (Vapnik, 1999). Support vector machines are mainly used for classification and regression, and can better deal with small samples, nonlinear and high dimensional data sets (Gupta et al., 2015). Compared with artificial neural network, support vector machines have better generalization ability and are more suitable for nonlinear regression prediction problem (Yoon et al., 2011).

In this paper, we propose a sub-period forecast model of the Chang Jiang River trunk freight volume based on support vector machines. First we analyze the influence factors of the freight volume and explain the relevant data used in the model. Second, we develop the sub-period forecast model with consideration of the actual water condition of the Chang Jiang River and analyze the forecast results of the sub-period model and full-period model. And finally, we draw the conclusions.

The paper is structured as follows. Section 2 provides the analysis of the influence factors. Section 3 presents the sub-period forecast model and finally section 4 concludes.

2 ANALYSIS OF INFLUENCE FACTORS

2.1 Non-natural Influence Factor

The trunk line of the Chang Jiang River described in this paper refers to the main channel of the Chang Jiang River flowing through nine cities and two cities including Shanghai, Jiangsu, Anhui, Jiangxi, Hunan, Hubei, Chongqing, Sichuan, Yunnan, Qinghai and Tibet. The freight volume of the Chang Jiang River trunk line is the direct data representation of the freight demand of that (Wang et al., 2017). The essence of analyzing the factors affecting the freight demand of the Chang Jiang River trunk line is to analyze the factors affecting the freight demand of that (Wang et al., 2017). The main non-natural factors affecting the transport of goods include economic factors, political factors, technical factors, transport network factors and transport capacity factors (Nuzzolo and Comi, 2014). Among these factors, economic factors and transport capacity factors have the greatest impact on freight demand (Patil and Sahu, 2016). This study mainly analyzes the impact of economic factors and transport capacity factors on freight transport, and makes use of economic factors and transport capacity factors to make a reasonable forecast of the freight volume on the Chang Jiang River trunk line.

In the Chang Jiang River trunk freight transport, the main types of freight include four major categories, containers, dry bulk, groceries, and liquid groceries. The Statistical Bulletin of the Development of the Waterway Transport Industry in 2012 analyzed the proportions of the above four major categories of goods in the total freight volume of the Chang Jiang River trunk line, as shown in Figure 1.
Containers mainly include refrigerated containers and general containers. Reefer containers are generally loaded with foodstuffs such as fruits, vegetables and aquatic products. Ordinary containers are generally loaded with a large amount of household items such as flour, liquid fuel and bumf and so on, which are closely related to the total retail sales of social consumer goods in China. Dry bulk, groceries and liquid bulk mainly include coal, steel, metal ores, mining and construction materials and oil. With the structural upgrading of cities along the Chang Jiang River, transforming the pattern of economic growth is more and more important. These are the important material bases for the industrialization of cities along the Chang Jiang River and even China, and which are strongly related to secondary industry, gross domestic product (GDP), etc. Besides, Chang Jiang River trunk freight, as an important member of waterway transport in China, is closely linked with the overall volume of water transport and the turnover volume of water transport in our country. Based on the above analysis, this study selects the national GDP, the added value of the secondary industry (SIV), the total retail sales of social consumer goods (TRS), the total volume of waterway transport (WTV) and the turnover volume of waterway transport (WRT) as the non-natural influence factors to develop the freight volume (FV) forecast model. The variables are given in Table 1. Economic factor and transport capacity factor are latent variables, and the observed variables are GDP, SIV, TRS, WTV and WRT, respectively.

Table 1 Input Variables Considered in the Forecast Model

<table>
<thead>
<tr>
<th>Latent Variables</th>
<th>Observed Variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>Economic factors</td>
<td>GDP, SIV, TRS</td>
</tr>
<tr>
<td>Transport capacity factors</td>
<td>WTV, WRT</td>
</tr>
</tbody>
</table>

2.2 Natural Influence Factor

Water level is an important indicator of sailing conditions and the level can directly affect the navigation (Millerd and Dufournaud, 1996; Liu et al., 2014). Similarly, the water level of the Chang Jiang River also has a great influence on the sailing of the Chang Jiang River.

Generally, the Chang Jiang River is in a low-water period from December to March of the following year, while other months are in the high-water period (Bing et al., 2011). The analysis of data from different periods showed that the freight volume in low-water period and high-water period presented different regularities. Therefore, this study selects the water level as the natural influence factor of the freight volume and proposes a sub-period forecast model with seasons as the time unit and the low-water period and the high-water period as the separation points.
2.3 Data and Correlation Analysis

In view of the data availability on the data source website, this study selects the data of freight volume and influence factors from 2011 to 2016 as the research data to make short-term forecast analysis on the freight volume of the Chang Jiang River trunk line. All data is collected from the National Data websites and the People's Republic of China Ministry of Transportation website.

According to the analysis in Section 2.1, this study uses GDP, SIV, TRS, WTV, WRT and FV values to develop forecast models. Selection of the input variables is one of an important problem when developing forecast models (Ghorbani et al., 2017). In order to verify the quantitative correlation between influence factors and the freight volume of the Chang Jiang River, the correlation analysis is carried out (Ghorbani et al., 2013). The specific correlation coefficient and the significance level are given in Table 2.

<table>
<thead>
<tr>
<th></th>
<th>FV</th>
<th>GDP</th>
<th>SIV</th>
<th>TRS</th>
<th>WTV</th>
<th>WRT</th>
</tr>
</thead>
<tbody>
<tr>
<td>FV</td>
<td>1.000</td>
<td>0.885**</td>
<td>0.784*</td>
<td>0.899**</td>
<td>0.981**</td>
<td>0.841**</td>
</tr>
<tr>
<td>GDP</td>
<td>0.885**</td>
<td>1.000</td>
<td>0.954**</td>
<td>0.940**</td>
<td>0.944**</td>
<td>0.906**</td>
</tr>
<tr>
<td>SIV</td>
<td>0.784*</td>
<td>0.954**</td>
<td>1.000</td>
<td>0.806**</td>
<td>0.867**</td>
<td>0.891**</td>
</tr>
<tr>
<td>TRS</td>
<td>0.899**</td>
<td>0.940**</td>
<td>0.806**</td>
<td>1.000</td>
<td>0.914**</td>
<td>0.815**</td>
</tr>
<tr>
<td>WTV</td>
<td>0.981**</td>
<td>0.944**</td>
<td>0.867**</td>
<td>0.914**</td>
<td>1.000</td>
<td>0.905**</td>
</tr>
<tr>
<td>WRT</td>
<td>0.841**</td>
<td>0.906**</td>
<td>0.891**</td>
<td>0.815**</td>
<td>0.905**</td>
<td>1.000</td>
</tr>
</tbody>
</table>

* The correlation is significant at a confidence level (double test) of 0.05.
** The correlation is significant at a confidence level (double test) of 0.01.

Table 2 it is obvious that all the influence factors selected in this study are significantly related to the freight volume of the Chang Jiang River. Therefore, using all these variables may provide a good forecast.

3 METHOD AND MODEL

3.1 Support Vector Machines

Support Vector Machines are based on statistical learning theory and are used for classification and regression problems (Vapnik, 1998).

For a given training data represented by \((x_1,y_1), \ldots, (x_n,y_n)\), where \(x\) is an input vector and \(y\) is a corresponding output vector, the regression function of the SVM can be expressed as:

\[
f(x) = w \cdot \varphi(x) + b
\]

where \(w\) is a weight vector, \(b\) is a bias, and \(\varphi(x)\) denotes a nonlinear transfer function, and the \(\varphi(x)\) is to nonlinearly map the input vectors to a high-dimensional feature space.

Vapnik (1995) introduced the convex optimization formula with a \(\xi\) insensitivity loss function as follows:

\[
\begin{aligned}
\min & \quad \frac{1}{2}w^2 + C \sum_{i=1}^{n} (\xi_i^+ + \xi_i^-) \\
n & \quad y_k - (w \cdot \varphi(x_k) + b) \leq \xi + \xi_k^- \\
\text{s.t.} & \quad (w \cdot \varphi(x_k) + b) - y_k \leq \xi + \xi_k^+, \quad k = 1, 2, \ldots, n \\
& \quad \xi_i^+, \xi_i^- \geq 0,
\end{aligned}
\]

where \(C\) is a positive regularization parameter that corresponds the weight parameters in the optimization problem for minimizing the empirical error. The variables \(\xi_i^+, \xi_i^-\) penalize training error by the loss function for the chosen error tolerance \(\xi\) (Debasmita et al., 2009). Using Lagrangian
multipliers the SVM optimization problem is usually solved in a dual form. By solving Eq. (1) the explicit function of Lagrangian multipliers or \( \alpha \) and \( \alpha^* \) is obtained as following:

\[
f(x, \alpha, \alpha^*) = \sum_{i=1}^{n} (\alpha_i - \alpha_i^*) \cdot K(x, x_i) + b
\]

(3)

where \( K(x, x_i) \) is the kernel function. It is used to map the input space which is non-linearly related to the output space onto some higher dimensional space. Figure 2 shows a schematic diagram of the SVM used in this study.

![Figure 2: A Schematic Structure of SVM Model (Yoon et al., 2011)](image)

Using SVM for regression forecast, the selection of kernel functions is very important. Choosing the appropriate kernel function helps to improve the training and prediction accuracy. This study uses the LIBSVM toolbox to model and chooses the radial basis function (RBF) as the kernel function, which is simple, efficient, adaptive and reliable (Roy et al., 2016).

### 3.2 Sub-period Forecast Model

Extracting the data of influence factors and freight volume of Chang Jiang River in the first season from 2011 to 2016 as the forecast data of the low-water period. More specifically, to choose the data in the first season from 2011 to 2015 as the training set and choose the data in the first season of 2016 as the validation set to develop the low-water period forecast model. For the convenience of show, marking 1 for the first season in 2011, 2 for the second season in 2011, 24 for the fourth season in 2016 and so on for the seasonal time series.

Based on the input variables of influence factors such as the gross domestic product (GDP), the added value of the secondary industry (SIV), the total retail sales of consumer goods (TRS), the total volume of waterway transport (WTV) and the turnover volume of waterway transport (WRT), the forecast results are given in Table 3.

<table>
<thead>
<tr>
<th>Seasonal Time Series</th>
<th>Observed Value (million tons)</th>
<th>Predicted Value (million tons)</th>
<th>Relative Error (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>21</td>
<td>727.87</td>
<td>738.60</td>
<td>1.47</td>
</tr>
</tbody>
</table>

Table 3 Results of Low-water Period Forecast

Extracting the data of influence factors and freight volume of Chang Jiang River in the second,
third and fourth season from 2011 to 2016 as the forecast data of the high-water period. More specifically, to choose the data in the second, third and fourth season from 2011 to 2015 as the training set and choose the data in the second, third and fourth season of 2016 as the validation set to develop the high-water period forecast model. The forecast results are given in Table 4.

### Table 4 Results of High-water Period Forecast

<table>
<thead>
<tr>
<th>Seasonal Time Series</th>
<th>Observed Value (million tons)</th>
<th>Predicted Value (million tons)</th>
<th>Relative Error (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>22</td>
<td>826.75</td>
<td>834.20</td>
<td>0.90</td>
</tr>
<tr>
<td>23</td>
<td>863.29</td>
<td>865.90</td>
<td>0.30</td>
</tr>
<tr>
<td>24</td>
<td>899.46</td>
<td>940.40</td>
<td>4.55</td>
</tr>
</tbody>
</table>

Selecting the all data of influence factors and freight volume of Chang Jiang River from 2011 to 2016 to develop full-period forecast model. More specifically, to choose the data from 2011 to 2015 as the training set and choose the data in 2016 as the validation set to develop the full-period forecast model. The forecast results are given in Table 5.

### Table 5 Results of Full-period Forecast

<table>
<thead>
<tr>
<th>Seasonal Time Series</th>
<th>Observed Value (million tons)</th>
<th>Predicted Value (million tons)</th>
<th>Relative Error (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>21</td>
<td>727.87</td>
<td>705.90</td>
<td>-3.02</td>
</tr>
<tr>
<td>22</td>
<td>826.75</td>
<td>846.30</td>
<td>2.36</td>
</tr>
<tr>
<td>23</td>
<td>863.29</td>
<td>877.10</td>
<td>1.60</td>
</tr>
<tr>
<td>24</td>
<td>899.46</td>
<td>962.80</td>
<td>7.04</td>
</tr>
</tbody>
</table>

Forecast results of different periods in Table 3, Table 4 and Table 5 show that the errors of sub-period and full-period freight volume forecast are small, which means the forecast model based on the influence factors and SVM regression method proposed in this paper is suitable for forecasting the freight volume of the Chang Jiang River trunk line.

### 3.3 Results

This section analyzes the sub-period forecast results and the full-period forecast results of the Chang Jiang River trunk line freight volume. The root mean square error (RMSE) and mean absolute percentage error (MAPE) are used as performance evaluation indexes (Yao et al., 2017). RMSE and MAPE can be expressed as:

\[
RMSE = \left( \frac{1}{n} \sum_{i=1}^{n} (\hat{y}_i - y_i)^2 \right)^{1/2}
\]  

(4)

\[
MAPE = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{\hat{y}_i - y_i}{y_i} \right| \times 100\%
\]

(5)

where \(\hat{y}_i, y_i\) denotes the predicted value and observed value, respectively. \(n\) denotes the number of the observations in validation set.

In order to more clearly compare the results of sub-period forecast and full-period forecast of Chang Jiang River freight volume, the results of low-water period forecast and high-water period forecast are combined to be referred to as sub-period forecast results. The comparative analysis results are given in Table 6 and Figure 3.
## Table 6 Results of Sub-period Forecast and Full-period Forecast

<table>
<thead>
<tr>
<th>Seasonal Time Series</th>
<th>Observed Value (million tons)</th>
<th>Sub-period Forecast</th>
<th>Full-period Forecast</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Predicted Value (million tons)</td>
<td>Relative Error (%)</td>
</tr>
<tr>
<td>21</td>
<td>727.87</td>
<td>738.60</td>
<td>1.47</td>
</tr>
<tr>
<td>22</td>
<td>826.75</td>
<td>834.20</td>
<td>0.90</td>
</tr>
<tr>
<td>23</td>
<td>863.29</td>
<td>865.90</td>
<td>0.30</td>
</tr>
<tr>
<td>24</td>
<td>899.46</td>
<td>940.40</td>
<td>4.55</td>
</tr>
<tr>
<td>MAPE (%)</td>
<td></td>
<td>1.81</td>
<td></td>
</tr>
<tr>
<td>RMSE (million tons)</td>
<td></td>
<td>21.53</td>
<td></td>
</tr>
</tbody>
</table>

Figure 3: Sub-period and Full-period Training Results and Forecast Results

According to Table 6, for the results of sub-period forecast, the MAPE and RMSE are 1.81% and 21.53 million tons, while these values for the full-period forecast are 3.51% and 35.59 million tons, respectively. Figure 3 shows the comparative plots between the results obtained from sub-period forecast and full-period forecast. The results suggest the accuracy of the sub-period model is significantly higher than the accuracy of the full-period model and the accuracy of former is almost twice the accuracy of latter. It is evident that the sub-period model based on the consideration of the cyclical water conditions of Chang Jiang River is more effective than the traditional full-period forecast model.

## 4 CONCLUSION

This study, with consideration of the actual cyclical water conditions in the Chang Jiang River, proposes a sub-period forecast method based on the support vector machines to forecast the freight volume of the Chang Jiang River trunk line. In the first stage, the analysis of influence factors of
freight volume was developed. While in the second stage, the SVM regression model was applied to forecast the sub-period freight volume of the Chang Jiang River trunk line. The results obtain small errors and this indicates an appropriate application of the support vector machines model to forecast freight volume of the Chang Jiang River trunk line. What’s more, the results indicate a greater accuracy with the sub-period forecasting method compared with the traditional full-period forecasting method.

This study provides a reliable method for the short-term freight volume forecast of the Chang Jiang River trunk line. Future direction of freight volume forecast would be the consideration of combination of the method proposed in this study and time series method. Besides, we should consider more influence factors and make the principal component analysis, and then develop the forecast model with the extracted principal components. This subject has the value of further study.

REFERENCES


