

XIN HUANG, Ph.D. candidate¹
E-mail: 201910101510@mail.scut.edu.cn

YIMIN WANG, Ph.D.^{1,2}
(Corresponding author)
E-mail: ctymwang@scut.edu.cn

PEIQUN LIN, Ph.D.¹
E-mail: pqlin@scut.edu.cn

HENG YU, Ph.D.¹
E-mail: ctyuheng@mail.scut.edu.cn

YUE LUO¹
E-mail: 201720107134@mail.scut.edu.cn

¹ School of Civil Engineering and Transportation
South China University of Technology
Guangzhou 510641, China

² State Key Lab of Subtropical Building Science
South China University of Technology

Traffic Management
Original Scientific Paper
Submitted: 31 Mar. 2020
Accepted: 14 Sep. 2020

FORECASTING THE ALL-WEATHER SHORT-TERM METRO PASSENGER FLOW BASED ON SEASONAL AND NONLINEAR LSSVM

ABSTRACT

Accurate metro ridership prediction can guide passengers in efficiently selecting their departure time and simultaneously help traffic operators develop a passenger organization strategy. However, short-term passenger flow prediction needs to consider many factors, and the results of the existing models for short-term subway passenger flow forecasting are often unsatisfactory. Along this line, we propose a parallel architecture, called the seasonal and nonlinear least squares support vector machine (SN-LSSVM), to extract the periodicity and nonlinearity characteristics of passenger flow. Various forecasting models, including auto-regressive integrated moving average, long short-term memory network, and support vector machine, are employed for evaluating the performance of the proposed architecture. Moreover, we first applied the method to the Tiyu Xilu station which is the most crowded station in the Guangzhou metro. The results indicate that the proposed model can effectively make all-weather and year-round passenger flow predictions, thus contributing to the management of the station.

KEYWORDS

seasonal and nonlinear least square support vector machine; short-term subway passenger flow prediction; multi-model fusion prediction; time series.

1. INTRODUCTION

In recent decades, metros with the characteristics of a large volume, high speed, low pollution, low resource consumption, low energy consumption, having a convenient and comfortable ride, and

being in line with the principles of sustainable development have become the first choice for major cities to solve traffic congestion and develop public transportation [1]. Passenger demand pressure is increasing, especially in developing cities. For instance, over 50% of public transportation passenger flow assignments of Guangzhou city were taken by the Guangzhou Metro, whose passenger flow intensity is the first in the country [2]. In metro-related research, short-term ridership forecasting plays a crucial role in improving the efficiency of metro systems, which has been an important part of intelligent transportation systems, for example, motivations and benefits that include alleviating station congestion, informing travelers about traffic conditions and providing real-time traffic monitoring and management. Thus, an increasing number of studies have been conducted to address the metro ridership pressure using short-term predictions, thus improving the metro service quality.

The subway passenger flow samples present complex nonlinear characteristics. The relationship between the change in subway passenger flow and the influencing factors is difficult to describe with a simple mathematical model [3]. In general, short-term passenger flow prediction methods in the existing literature primarily include three types: linear prediction methods (LPs) [4], nonlinear prediction methods (NLPs) [5-12], and combined model prediction methods (CMs) [13-19]. Motivated by concerns for the environment and

new developments in technology, there has been a resurgence of interest in nonlinear and combined model prediction methods. *Table 1* summarizes the previous studies on short-term passenger flow prediction in recent years. The findings can be briefly summarized as follows:

- 1) LPs including historical average prediction methods, time series prediction methods, and Kalman filter model prediction methods, etc. There is a general problem in LPs; namely, in the case of nonlinear and uncertain traffic flow, the performance of the model deteriorates, so it is not possible to predict events with strong randomness [10, 12].
- 2) NLPs including artificial neural network and support vector machine (SVM) etc., artificial neural network prediction methods, such as back propagation neural network (BPNN) [7], long short-term memory (LSTM) [11], gated recurrent units (GRU) [9], convolutional neural network (CNN) [18], are widely used in the field of intelligent transportation. Although the artificial neural network methods have a good predictive effect on the temporal data, there are some insurmountable shortcomings, like

overlearning, dimensionality disaster, and easily falling into local optimum [12]. Support vector machines (SVMs) have a good generalization performance, good robustness, high precision, and better prediction results than BPNNs and dynamic recurrent neural networks in short-term passenger flow forecasting [3]. However, SVMs run slowly when solving large sample problems [12].

- 3) CMs mainly combine two or more prediction methods. If each prediction method is combined with an effective form, the prediction sample information can be more fully utilized to achieve a higher prediction accuracy than a single prediction method – according to the excellent review paper by Bates JM et al. [20]. The CMs can be roughly classified into three categories – multi-features [18, 19], consider periodicity [14, 16, 17], and real-time prediction [10].

However, it is unfortunate that there is only one small study which specifically deals with short-term subway passenger flow prediction. Moreover, these models did not fully consider the key factors (*Table 1*) of short-term metro passenger flow

Table 1 – Summary of the previous studies on short-term passenger flow prediction

Year	Reference	Model	Single(S)/Multi (M) features	Consider periodicity	Subway passenger flow prediction	Short-term prediction
1997	Smith et al. [4]	LP	S	-	-	√
2006	Yang et al. [5]	NLP	S	-	-	√
2007	Vanajakshi [6]	NLP	S	-	-	√
2010	Boto-Giralda et al. [13]	CM	S	-	-	√
2012	Pan [14]	CM	S	√	-	√
2012	Wei et al. [15]	CM	S	-	√	√
2013	Bouzerdoum et al. [16]	CM	S	√	-	√
2013	Wang et al. [7]	NLP	S	-	-	√
2013	Yang [3]	CM	S	-	√	√
2013	Guo et al. [8]	NLP	S	-	-	√
2015	Sun et al. [17]	CM	S	√	√	-
2017	Guo et al. [9]	NLP	S	-	-	√
2018	Wang et al. [10]	NLP	S	-	√	√
2019	Kong et al. [11]	NLP	S	-	-	√
2019	Lu et al. [12]	NLP	S	-	-	√
2019	Tang et al. [18]	CM	M	-	√	√
2019	Ma et al. [19]	CM	M	-	√	√
-	This paper	CM	M	√	√	√

forecasting. Here, we utilize a wavelet transform method as well as SVM to reach a balance of solving problems speedily and precisely.

This article mainly aims to improve the prediction accuracy of metro passenger flow time series prediction models, whether the forecasting target is a working day or a holiday. Along this line, we propose a parallel architecture, called the seasonal and nonlinear least squares support vector machine (SN-LSSVM), which is comprised of a cycle-based least squares support vector machine (W-LSSVM) and a day-based least squares support vector machine (D-LSSVM), to extract the periodicity and nonlinearity characteristics of passenger flow, respectively. To summarize, primary contributions of this paper are as follows:

Firstly, a new nonlinear hybrid prediction model called the SN-LSSVM is proposed for short-term passenger flow forecasting. Moreover, the model is not limited to metro passenger flow but can be extended to other forecasting applications.

Secondly, we evaluate the performance of the proposed prediction model through extensively comparative experiments in terms of: (a) linear prediction model auto regressive integrated moving average (ARIMA); (b) supervised learning model SVM; (c) hybrid model Wavelet-SVM, and (d) neural network model LSTM.

Thirdly, with an empirical demonstration of the proposed model using the Guangzhou Metro AFC data, we conclude that the proposed method has higher accuracy in predicting the all-weather, and

year-round passenger flow. Thus, this paper provides a way for metro companies to better manage the subway station.

The remaining parts of this paper are organized as follows. Section 2 introduces detailed prediction methods proposed in this paper. Section 3 analyses the law of passenger flow at the Guangzhou Metro. Evaluation methods and extensive experiments are conducted in Section 4, and the conclusions and future work are discussed in Section 5.

2. METHODOLOGY

2.1 Data description

The experimental dataset was extracted from the automatic fare collection (AFC) system of the Tiyu Xilu Station in Guangzhou China of 2016 and 2017. It can be seen from *Figure 1* that the subway passenger flow during Spring Festival decreases significantly. The reason for this phenomenon is that more than half of the permanent residents in Guangzhou are migrants. Since the Spring Festival is the Chinese New Year, in the Spring Festival of 2017, 61.36% of people, according to statistics, returned to their hometown for the festival [21]. To eliminate the influence of the Spring Festival, passenger flow data during the Spring Festival are excluded, and 15 weeks of passenger flow data are selected for the experiment, as shown in *Table 2*. The type of working days indicates that the target date is a working day, as are the types of holidays 1 and holidays 2. Moreover, the passenger flow data are recorded with 60-minute intervals.

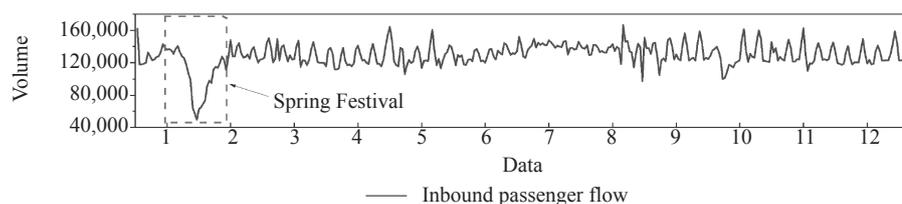


Figure 1 – Annual inbound passenger flow at the Tiyu Xilu Station of 2017

Table 2 – The experimental descriptions of the two datasets

Item/Type	Working days	Holidays1	Holidays2
Date of training	13 Feb 2017–21 May 2017 22 Feb 2016–22 May 2016	13 Feb 2017–23 Apr 2017 22 Feb 2016–24 Apr 2016	20 Feb 2017–30 Apr 2017 29 Feb 2016–1 May 2016
Date of testing	22 May 2017–28 May 2017 23 May 2016–29 May 2016	24 Apr 2017–30 Apr 2017 25 Apr 2016–1 May 2016	1 May 2017–7 May 2017 2 May 2016–8 May 2016
Prediction time horizon	06:00–24:00		
Size of dataset	18×28×7=3528	18×21×7=2646	18×21×2646

2.2 Notation and problem statement

For the convenience of the readers, the critical notations used throughout the paper are summarized in Table 3.

Given a t -size sequence of the observed passenger volume, define traffic data profile of one day as $d_j = \{x_j^1, x_j^2, \dots, x_j^t\}^T$ where t is the recording point (if the record interval is 60 minutes, $t=18$), and x_j^t is

Table 3 – Notations

Sets	D	Set of daily passenger flow, $j=1,2,\dots,n$
	W	Set of weekly passenger flow, $i=1,2,\dots,n$
	d_j	Passenger flow time series of j -th day
	w_i	Passenger flow time series of i -th week
	wL_q^i	Low-frequency signal of q -th decomposition of w_i
	wH_q^i	High-frequency signal of q -th decomposition of w_i
	wL_7	Predicted low-frequency signal
	wH_q	Predicted high-frequency signal of q -th decomposition
	d_{j+1}	Predicted daily passenger flow time series
	w_{i+1}	Predicted weekly passenger flow time series
	d_{j+1}^*	Daily passenger flow time series extracted from w_{i+1}
	P	Predicted passenger flow time series
	P	Real passenger flow time series
	P_r	Predicted passenger flow time series of r -th group
	P_r	Real passenger flow time series of r -th group
	Parameters	m
n		Number of the training set of daily passenger flow
q		Wavelet transform stage
l_q		Number of samples in the passenger flow time series of the q -th wavelet transform
h_0		Low-pass filters
h_1		High-pass filters
H^*		Dual operators of h_0
G^*		Dual operators of h_1
a_k		Coefficients of the low-pass filter
b_k		Coefficients of the high-pass filter
$\varphi(x)$		Nonlinear transformation
ω		Normal vector
b		Offset vector
γ		Adjustable parameters
ε_i		Slack variable
x_i		Input vector
y_i		Target value
K		Kernel function
t		Recording point
α_1		Weights of W-LSSVM
α_2	Weights of D-LSSVM	
N	Mean absolute percentage error	

the sum of transaction count during t -th time slot. Moreover, define traffic data profile of one week as $w_i = \{x_i^1, x_i^2, \dots, x_i^t\}^T$ (if the record intervals is 60 minutes, $t=126$). Two sets are extracted from the original time series as $W = \{w_1, w_2, \dots, w_i\}^T$ ($i=1, 2, \dots, m$) and $D = \{d_1, d_2, \dots, d_j\}^T$ ($j=1, 2, \dots, n$). W is a set of passenger flow, which contains the periodic characteristics of the time series. D is a set of passenger flow, which contains the nonlinear characteristics of the time series.

2.3 Overview of the forecasting framework

The SN-LSSVM combines the complementary advantages of the W-LSSVM model and the D-LSSVM model, and overcomes their shortcomings respectively. The primary characteristics of the proposed method are as follows:

- Consideration of the periodicity and nonlinearity of passenger flow, and employment of the LS-SVM model, which is more suitable for the nonlinear passenger flow.
- Great reduction of the learning size and the avoidance of slow training speed.

- The use of the passenger flow time series collected only from one metro station, which consists of two advantages: a) we can obtain a satisfactory prediction results even without the road network data, weather data and other factors; b) based on the proposed method to consider other factors (weather and passenger flow of nearby stations), we may have a more stable prediction method.

The prediction method can be described as follows (as shown in Figure 2):

Stage 1: Traffic data profiling. Extract two pieces of passenger flow data: the set of W and the set of D .

Stage 2: Time series decomposition. The W is regarded as a signal sequence, which is decomposed into a group of high-frequency signals (large fluctuations in the time series curve) and a low-frequency signal (small fluctuations in the time series curve) via wavelet transform (WT).

Stage 3: Predictive modelling. The LS-SVM is embedded in the framework to train the decomposed W and the D , and then the predicted high- and low-frequency signals and d_{j+1} are obtained.

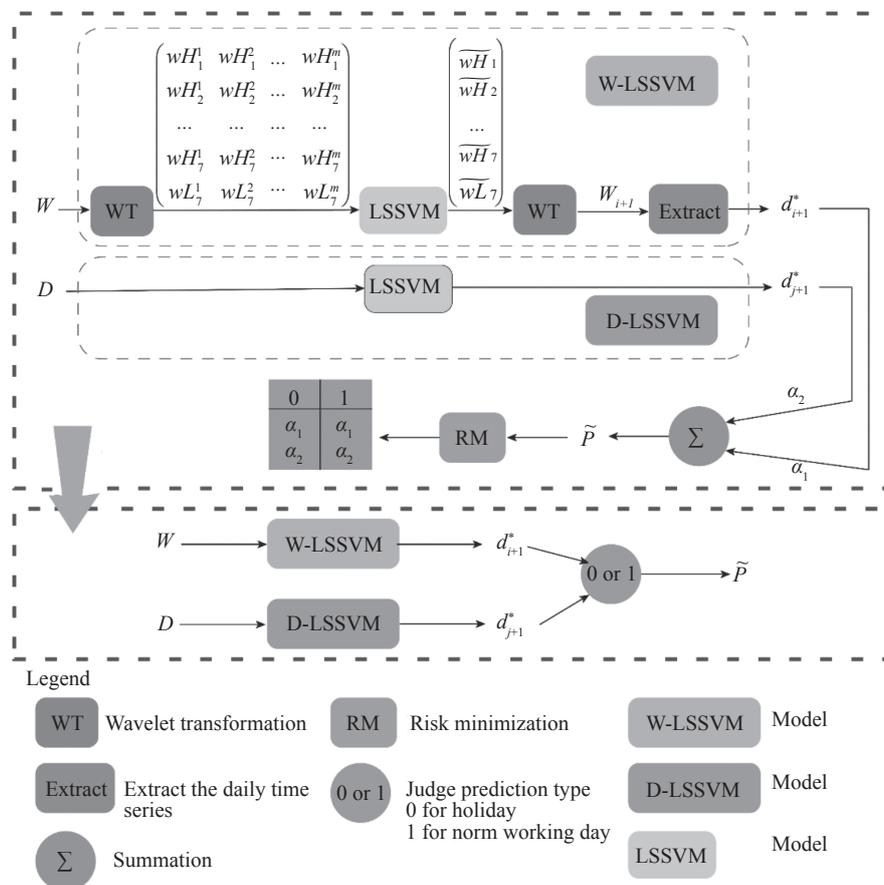


Figure 2 – SN-LSSVM model flow chart

Stage 4: Signals reconstruction. The predicted high- and low-frequency signals are synthesized via wavelet reconstruction technology to obtain the prediction results of w_{i+1} .

Stage 5: Output results. The weights of the two models are calculated using the empirical method (it will be discussed in detail in subsection 2.6). Ultimately, the results of the SN-LSSVM are obtained by merging d_{j+1} and d_{j+1}^* , where $d_{i+1}^* \in w_{i+1}$.

2.4 Decomposing the original passenger flow

Passenger flow data can be considered a signal sequence, whereas the WT can handle nonlinear and non-stationary data. In this paper, the WT is used to analyze and extract the characteristics of W , because W is much more complicated than D . This subsection briefly reviews the WT algorithm from Yang [3] and Sun [17], and interested readers can refer to it for detailed algorithmic descriptions and theoretical property analysis.

The WT can refine the passenger flow (signal) function through scaling and translation by using multiple scales, and finally achieve high- and low-frequency subdivisions. The LS-SVM is used to predict the high- and low-frequency information of passenger flow respectively, which greatly reduces its training and learning scale compared with direct prediction, thus avoiding problems such as slow convergence speed.

Db3 wavelet is one of the WTs and is used to decompose the w_i , then a high-frequency signal and a low-frequency signal wH_1^i can be respectively obtained (Figure 3a). Considering wL_1^i as a new passenger flow time series to be decomposed, the use of Db3 wavelet is continued to obtain a high-frequency signal wH_2^i and a low-frequency signal wL_2^i . By analogy, a low-frequency signal wL_q^i and q groups of high-frequency signals (wH_1^i, \dots, wH_q^i) whose frequency is essentially stable are finally obtained.

In this paper, $q=7$ is used, that is, after 7 steps of decomposition, 8 groups of decomposed sequences ($wH_1^i, \dots, wH_7^i, wL_7^i$) are given. Among these sequences, wL_7^i characterizes feature of w_i , and wH_1^i, \dots, wH_7^i show the subtle fluctuation.

The number of samples of w_i is l_0 , and the number of samples after the j -th decomposition is l_j . After each decomposition, the length of the new signal is shortened by half, that is, $l_j = \frac{l_0}{2^j}$. The process of the WT can be expressed by the following formula:

$$wL_j^i(n) = \sum_{k=1}^{l_{j-1}} h_0(k-2n)wL_{j-1}^i(k) \tag{1}$$

$$wH_j^i(n) = \sum_{k=1}^{l_{j-1}} h_1(k-2n)wL_{j-1}^i(k) \tag{2}$$

where the coefficients $h_0(k-2n)$ and $h_1(k-2n)$ are two columns of conjugate filter coefficients determined by wavelet functions, called low-pass filters and high-pass filters, respectively. Their values are determined by the db3 wavelet and the number of coefficients.

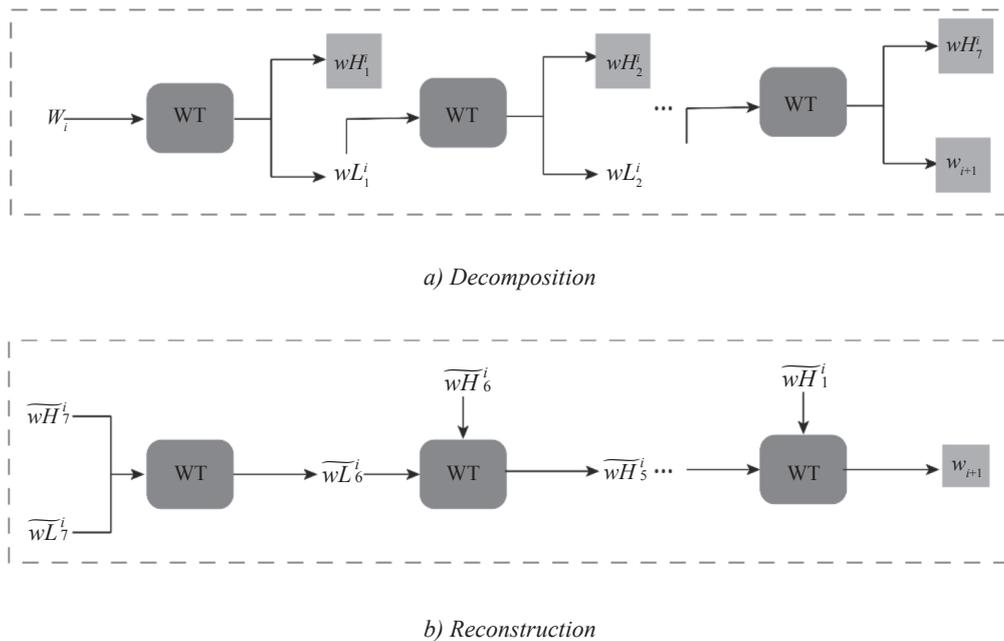


Figure 3 – The diagram of the original passenger flow time series

$$h_0(t) = \sum_{k=0}^{n-1} a_k h_0(2t-k) \tag{3}$$

$$h_1(t) = \sum_{k=0}^{n-1} b_k h_1(2t-k) \tag{4}$$

where (a_0, a_1, \dots, a_k) are the coefficients of the low-pass filter, and (b_0, b_1, \dots, b_k) are the coefficients of the high-pass filter. For a db3 wavelet, n is 6.

2.5 Predicting and reconstructing

The SVM is a technology proposed by Sain and Vapnik [22], whose basic idea is as follows: if the sample data X is nonlinearly separable, the input vector X is mapped to the high-dimensional feature space H by the nonlinear transformation $\varphi(x)$, and the optimal regression function is obtained in H , so a linear regression in a high dimensional space corresponds to a nonlinear regression in a low dimensional space.

Basically, the LS-SVM uses the same principles as the SVM, but it simplifies the operation by an appropriate transformation, and improves the accuracy and convergence speed of the solution problem [22, 23]. The LS-SVM can be formulated as a minimization problem as follows:

$$\begin{aligned} \min \frac{1}{2} \|\omega\|^2 + \frac{1}{2} \gamma \sum_{i=1}^n \varepsilon_i^2 \\ \text{s.t. } y_i [\omega^T \varphi(x_i) + b] = 1 - \varepsilon_i \end{aligned} \tag{5}$$

where y_i is the target value and x_i is the input vector, $\omega^T \varphi(x_i)$ represents the inner product in the high-dimensional feature space H , ω and b are the normal vector and the offset vector, respectively, ε_i is the slack variable and γ is the adjustable parameters that controls the regression error. The regression of LS-SVM is:

$$y(x) = \sum_{i=1}^n \alpha_i K(x \cdot x_i) + b \tag{6}$$

where K is the kernel function, α_i are the Lagrange multipliers. In this work, the radial basis function (RBF) kernel is used [24], and parameters are learned by the particle swarm optimization (PSO) algorithm [25].

The decomposed sequences $(wH_1^i, \dots, wH_7^i, wL_7^i)$ and (d_1, \dots, d_j) are the input of the LS-SVM model. First of all, the low-frequency sequences at the 7th level of m periods $wL_7^1, wL_7^2, \dots, wL_7^m$ are used to forecast the 7th level wL_7 of the target $(m+1)$ period. The process can be described as follows:

1) Select $wL_7^1, wL_7^2, \dots, wL_7^m$ as a training data sample to train the LS-SVM model.

2) Use the trained LS-SVM model to predict $wL_7(t)$.
 3) Repeat step 1 and step 2 to predict $wH_1 \sim wH_7$ and D .

At this stage, the high-frequency signals $wH_1 \sim wH_7$ and the low-frequency signals wL_7 are subjected to wavelet reconstruction to obtain a predicted passenger flow time series w_{i+1} , as shown in Figure 3b.

The reconstruction formula of the passenger flow data is as follows:

$$wL_{p-1}(k) = H^* wL_p(n) + G^* wH_p(n) \tag{7}$$

where H^* and G^* are the dual operators of $h_0(k-2n)$ and $h_1(k-2n)$, respectively.

2.6 The hybrid SN-LSSVM Model

As previously described, the SN-LSSVM model is constructed using a D-LSSVM model and a W-LSSVM model. The passenger flow time series d_{j+1}^* of a certain day is obtained from the reconstructed passenger flow time series w_{i+1} , and the final predicted passenger flow is calculated as follows:

$$\bar{P} = \alpha_1 d_{i+1}^* + \alpha_2 d_{j+1} \tag{8}$$

where α_1 and α_2 are the weights of d_{j+1}^* and d_{j+1} .

To obtain the value of α_1 and α_2 , a function is defined to solve this optimization problem:

$$\begin{aligned} J(\alpha_{1i}, \alpha_{2i}) = \min N \\ \text{s.t. } \bar{P} = \alpha_{1i} \bar{D}^* + \alpha_{2i} \bar{D}, \quad i = 1, 2, \dots, m \\ P = \alpha_{1i} \bar{D}^* + \alpha_{2i} \bar{D} + \varepsilon_i^2 \\ N = \frac{1}{n} \sum_{j=1}^n \frac{|P_j - \bar{P}_j|}{\bar{P}_j} \cdot 100\% \end{aligned} \tag{9}$$

where P is real value, \bar{P} is the predicted value, ε_i^2 is the error variable, N is the mean absolute percentage error (MAPE), m is the number of experiments, and n denotes the number of the data points each day.

Furthermore, the following equation was defined:

$$\begin{aligned} \alpha_1 = \frac{\sum_{i=1}^m \alpha_{1i}}{\sum_{i=1}^m \alpha_{1i} + \sum_{i=1}^m \alpha_{2i}} \\ \alpha_2 = \frac{\sum_{i=1}^m \alpha_{2i}}{\sum_{i=1}^m \alpha_{1i} + \sum_{i=1}^m \alpha_{2i}} \\ \text{s.t. } \sum_{i=1}^m \alpha_{1i} + \sum_{i=1}^m \alpha_{2i} = m \end{aligned} \tag{10}$$

More details of the SN-LSSVM are shown in Algorithms 1-3.

Algorithm 1 – SN-LSSVM Approach

Input: d_{i+1} and d_{i+1}^*
Output: Forecasting value p^*

```

if  $\alpha_1$  and  $\alpha_2 \in \phi$ 
  for  $i = 0$  to  $n$ , do
    use Algorithm 2,3
     $(\alpha_{1i}, \alpha_{2i}, 0, 1) \leftarrow \text{RM}(d_{i+1}^*, d_{i+1}, p)$ 
  end for
   $(\alpha_1, \alpha_2, 0, 1) \leftarrow \text{smooth}(\alpha_{1i}, \alpha_{2i}, 0, 1)$ 
end if
use Algorithm 2,3
if  $p$  is holiday
   $(\alpha_1, \alpha_2) \leftarrow (\alpha_1, \alpha_2, 0)$ 
   $p^* \leftarrow \alpha_1 * d + \alpha_2 * d$ 
else
   $(\alpha_1, \alpha_2) \leftarrow (\alpha_1, \alpha_2, 1)$ 
   $p^* \leftarrow \alpha_1 * d + \alpha_2 * d$ 
end if

```

Algorithm 2 – W-LSSVM Approach

Input: Set of weekly passenger flow W
Output: Forecasting result d_{i+1}^*

```

 $C \leftarrow \text{wavedec}(W, 7, 'db3')$  // Decomposition
 $(C^*, p1) \leftarrow \text{map}(C, 0, 1)$  // Normalization
 $C^{**} \leftarrow \text{LSSVM}(C^*)$ 
 $C^{***} \leftarrow \text{map}('reverse', C^{**})$  // Anti-normalization
 $w_{i+1} \leftarrow \text{waverec}(C^{***}, 'db3')$  // Reconstruction
 $j \leftarrow 0$ 
for  $i = 0$  to 6, do
   $d_{i+1}^* \leftarrow w_{i+1} [j \dots j+17]$ 
   $j \leftarrow j+18$ 
next  $i$ 
Return  $d_{i+1}^*$ 

```

Algorithm 3 – D-LSSVM Approach

Input: Set of daily passenger flow D
Output: forecasting result d_{i+1}

```

 $(D^*, p1) \leftarrow \text{map}(D, 0, 1)$  // Normalization
 $D^{**} \leftarrow \text{LSSVM}(D^*)$ 
 $d_{i+1} \leftarrow \text{map}('reverse', D^{**})$  // Anti-normalization

```

3. ANALYSIS OF THE SUBWAY PASSENGER FLOW

The AFC system was able to achieve real-time data collection of metro passengers entering and exiting the station [26]. By simple statistics, the ridership data can be achieved for the required time interval, such as 5 min, 15 min, 30 min, and 60 min.

Figure 4 shows the passenger flow changes of the inbound, outbound and transfer within one day, respectively. The data comes from a working day (24 May 2017), a weekend (28 May 2017) and a holiday (1 May 2017) at the Tiyu Xilu Station.

The passenger flow of the Tiyu Xilu Station presents three states:

- 1) The concentrated passenger flow pattern on working days;
- 2) The loose passenger flow pattern on weekends;
- 3) The increase in the number of state 2 during festivals.

Figure 5 illustrates two different inbound passenger flow patterns at the Tiyu Xilu Station: *weekdays+festivals* and *weekdays+weekends*. Passenger flow at the station will change greatly on the weekends, showing a cycle of one week. That is, from Monday to Friday the traffic is often concentrated, and the passenger flow on the weekends is relatively light. In this case, forecasting the passenger traffic in the unit of weeks will achieve better prediction results. However, when there are holidays, large-scale events or sudden emergencies, the passenger flow will no longer show a cycle of one week. At that point, using forecasting in the unit of days can achieve better results. In other words, the former meets the requirements of a sufficient sample size, simultaneously considering the periodicity of passenger flow, but ignores the recent changes, while the latter takes recent changes into account in the predicted values, but lacks sufficient samples. The SN-LSSVM model can achieve a sufficient sample size, consideration of the periodicity, and recent changes in the predicted value.

4. EXPERIMENTS

4.1 Evaluation methods

In this paper, the metrics for performance evaluation include: (a) root mean square error (RMSE), for measuring the deviation between the predicted value and the actual value; (b) correlation coefficient R , for measuring the explanatory power of models by calculating the correlation coefficient between observations and predictions; and (c) mean arctangent absolute percentage error (MAAPE), for measuring the relative average deviation, defined as following:

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (\bar{y}_i - y_i)^2}{n}} \tag{11}$$

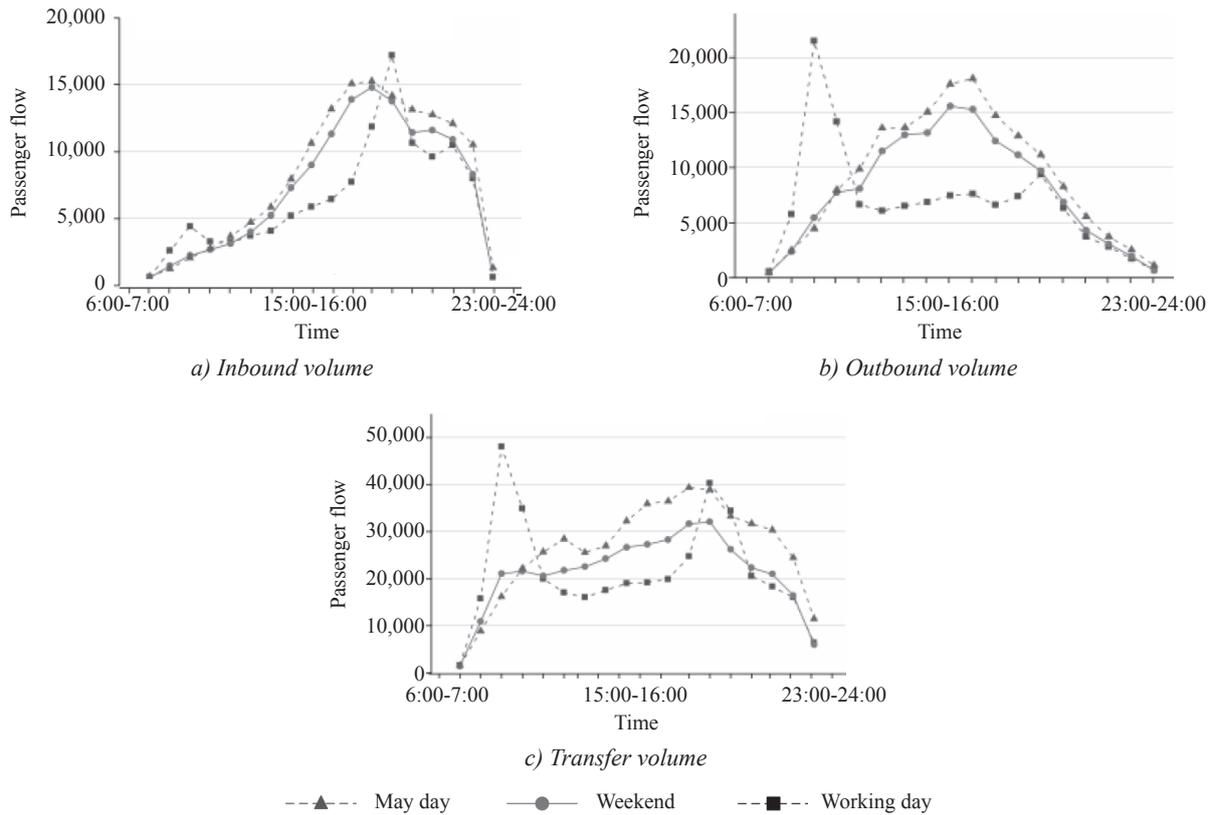


Figure 4 – The passenger flow on working days, weekends and festivals at the Tiyu Xilu station

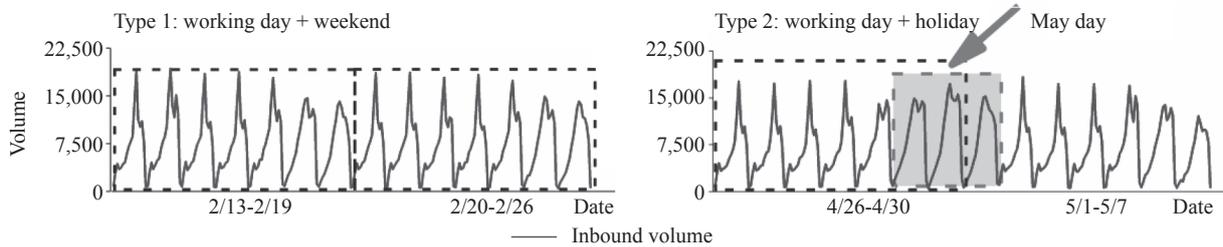


Figure 5 – Two different types of the original inbound passenger flow time series at the Tiyu Xilu Station

$$R(\bar{y}, y) = \frac{Cov(\bar{y}, y)}{\sqrt{Var[\bar{y}]Var[y]}} \quad (12)$$

$$MAAPE = \frac{1}{n} \sum_{i=1}^n \left(\arctan \left(\left| \frac{\bar{y}_i - y_i}{y_i} \right| \right) \right) \quad (13)$$

where n is the number of samples, \bar{y}_i, y_i are the predicted value and observation value respectively, and $Cov(\bar{y}, y)$ is the covariance of \bar{y} and y , $Var[\bar{y}]$ is the variance of $[\bar{y}]$, and $Var[y]$ is the variance of y .

4.2 Forecasting models and weights

In this work, we employ five typical forecasting methods: ARIMA [27], LSTM [11], D-LSSVM, W-LSSVM and SN-LSSVM. The experimental descriptions of the dataset are in Table 2 and the passenger flow on weekdays and holidays are predicted, respectively.

The forecasting models are noted as follows: ARIMA: ARIMA (5,1,3) model with Temporal features, including the *original time series*.

LSTM: LSTM model with Temporal features, including D matrices.

D-LSSVM: LSSVM model with Temporal features, including D matrices.

W-LSSVM: LSSVM model with Temporal features, including W matrices.

SN-LSSVM: LSSVM model with Temporal features, including D matrices and W matrices.

Figure 6 shows the high- and low-frequency signals decomposed from 13 February to 19 February 2017. As seen from the figure, the complex passenger flow time series is decomposed into simple high- and low-frequency curves, which not only reduces the calculation time, but also improves the stability of prediction results.

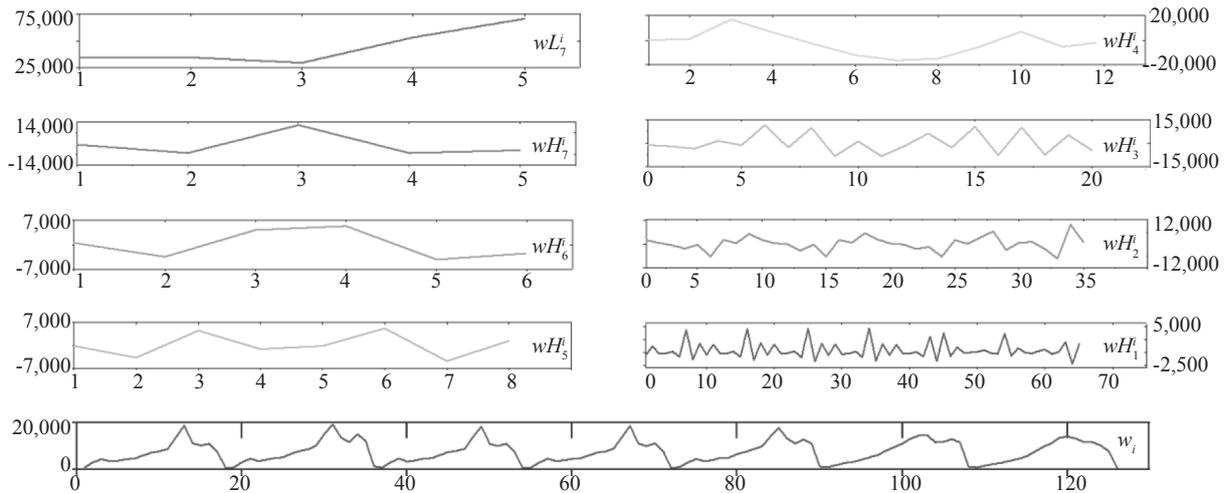


Figure 6 – The decomposition signals of the original inbound passenger flow data from 13 February to 19 February 2017

The prediction result of the SN-LSSVM is obtained by combining the results of the W-LSSVM and D-LSSVM. As discussed in Subsection 2.6, the weights of the W-LSSVM and D-LSSVM are obtained with the original passenger flow data at the Tiyu Xilu Station in 2016.

4.3 Evaluation of forecasting horizon

Forecasting horizon denotes the length of predicted future. In this section we evaluate the prediction performance in 1 day and 7 days in the future, that is $h=1,7$, meaning $t+h$ forecasting. In this paper, the D-LSSVM is a $t+1$ forecasting model and the W-LSSVM is a $t+7$ forecasting model.

Figure 7a is a comparison of the working days passenger flow data predicted by the W-LSSVM method with the actual passenger flow data. It can

be seen from the figure that the predicted passenger flow time series curve is essentially consistent with the actual passenger flow time series curve. However, as can be seen from Figure 7b and Figure 7c, when there is a holiday (May Day) in the forecast target, the predicted curve is not consistent with the actual curve. The underlying reason could be that the $t+7$ forecasting is unstable. A sudden change in the passenger flow, meanwhile, comes when there is a holiday.

Table 4 shows the weights (α_1, α_2) of the two methods, which range from 0 to 1; the larger the value, the higher the accuracy. Since the difference in the ridership pattern is between the previously described weekdays and holidays, the weights have two patterns. As can be seen from the table, in the SN-LSSVM model, the average weights of

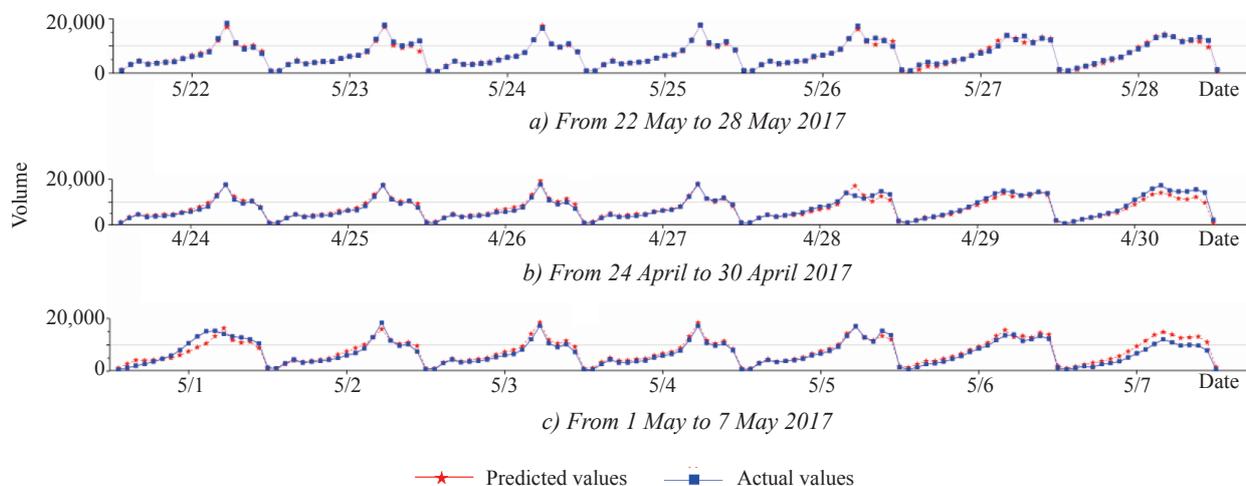


Figure 7 – A comparison between the actual values and the predicted values using the original inbound passenger flow time series via W-LSSVM at the Tiyu Xilu Station

Table 4 – Weights of W-LSSVM and D-LSSVM with the original time series at the Tiyu Xilu Station from 25 April to 29 May 2016

Date	W-LSSVM	D-LSSVM	Date	W-LSSVM	D-LSSVM
	α_{1i}	α_{2i}		α_{1i}	α_{2i}
25 April	0	1	6 May	0.496	0.504
26 April	0	1	7 May	0	1
27 April	0.128	0.872	8 May	0	1
28 April	0.846	0.154	23 May	0.94	0.06
*29 April	1	0	24 May	0.833	0.167
*30 April	0	1	25 May	1	0
*1 May	0.545	0.455	26 May	0.878	0.122
2 May	0	1	27 May	0	1
3 May	0	1	28 May	0	1
4 May	0	1	29 May	0.469	0.531
5 May	0.156	0.844			
Holidays	α_1	α_2	Working days	α_1	α_2
	0.18	0.82		0.41	0.59

Note: In this paper “working days” means all days in the year except holidays.

the W-LSSVM model for non-holiday predictions are much larger than that for holidays. This is because of the sudden change in the passenger flow, the nonlinearity of which can be addressed by the D-LSSVM model. In contrast, the working days ridership is mainly composed of the commuted passenger flow, which is more stable. Therefore, the weights of W-LSSVM model are similar to those during working days.

4.4 Forecasting results

It can be seen from Figure 8b that the prediction trends of the five prediction methods are essentially the same for the prediction of short-term subway passenger flow. However, Figure 8a shows that the prediction trends of the D-LSSVM, W-LSSVM, and SN-LSSVM are essentially the same, while the LSTM and ARIMA fail to fully grasp the real-time variation law of the passenger flow, and the prediction effect has an obvious deviation from the actual situation. Similarly, as can be seen from Figure 8c, for the short-term subway passenger flow forecasting during holidays, the prediction trends of the D-LSSVM, SN-LSSVM, and LSTM are essentially the same, while the others fail to fully grasp the real-time variation law of the passenger flow, and the prediction effect has an obvious deviation from

the actual situation. Generally, we conclude that the nonparametric regression model SVM outperforms some neural networks and ARIMA, which is consistent with previous studies as well [5, 6, 26].

4.5 Validation results of the novel SN-LSSVM prediction model

In order to facilitate the comparison of the five models, we have divided the indicators of the prediction results into hierarchies, and the principle of division is as follows:

- RMSE: $RX \in [0, 2000]$;
- R: $RX \in [90\%, 100\%]$;
- MAAPE: $RX \in [0, 0.16]$;

where RX is a qualified interval. Then, the RX s of the three indicators are normalized to $[0, 1]$. Specifically, the RX of R is $1 - [0, 1]$. If the indicators exceed the qualified interval, its value will be presented in black (Figure 9). Figure 9 shows the forecasting errors of different models for working days and holidays, respectively.

Firstly, it can be easily concluded that the predicted result of the ARIMA is unsatisfactory. Most of the indicators of the ARIMA are unqualified, indicating the linear modelling is unsuitable for forecasting nonlinear time series.

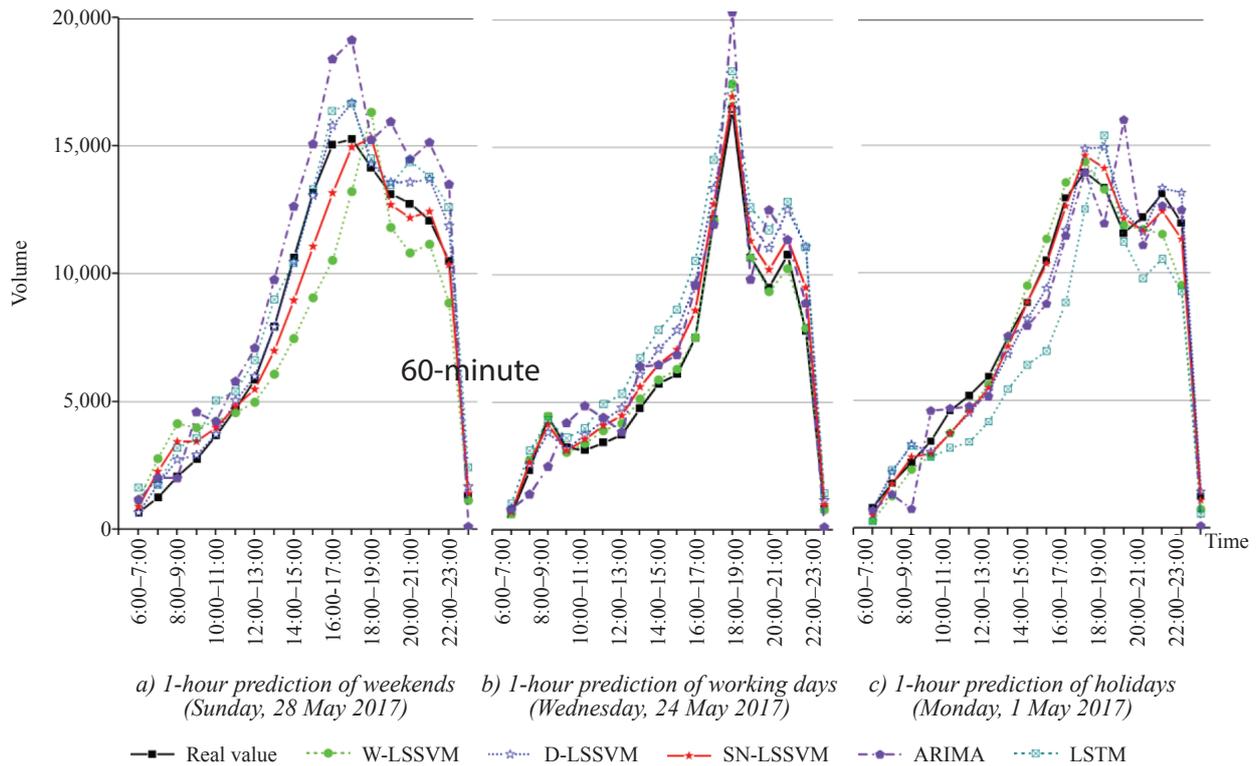


Figure 8 – Comparison between the actual values and the predicted values using the historical inbound passenger flow time series with 5 different models

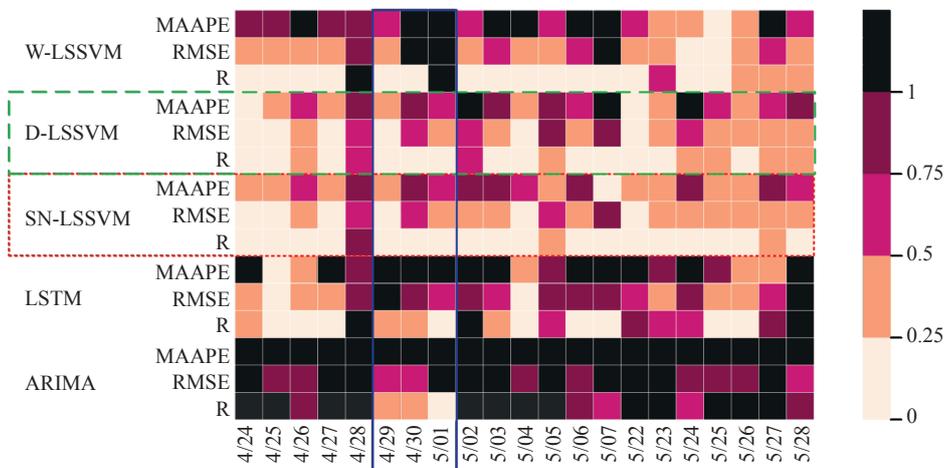


Figure 9 – Performances of the 5 different models

Secondly, RMSE and R of the LSTM generally meet the prediction accuracy requirements, while the MAAPE is not satisfactory. We can conclude that the LSTM also has large errors in predicting small passenger flow samples (in the morning and evening). Moreover, holidays have a larger impact on the prediction accuracy of the LSTM, and the value of the indicator is still affected after three weeks.

Thirdly, the D-LSSVM outperforms the W-LSSVM as a whole (up to 4.1% improvement in MAAPE and up to 5.3% improvement in RMSE).

Since the D-LSSVM has a smaller forecasting horizon compared to the W-LSSVM, we can conclude that short-term passenger flow forecasting could provide a fair prediction in the immediate future, which is typically up to one day ahead. Moreover, the W-LSSVM is suitable for the prediction during normal working days rather than holidays, due to the instability of passenger flow on holidays. Furthermore, the W-LSSVM provides a satisfactory predicted value for three weeks after May Day, indicating that the model is more robust.

Fourthly, there are some cases where the predicted value of the D-LSSVM fails, because the model is suitable for a small sample training and the training set of the model is insufficient.

Fifthly, all of the prediction results of SN-LSSVM fall into the qualified interval, which is superior to other models. Moreover, the SN-LSSVM has high indicators of the prediction results, especially for holiday forecasting. Since the SN-LSSVM leverages multivariate features (periodicity) for predictive learning, while the D-LSSVM considers nonlinearity of the time series only, we can conclude that additional time series features can be more effective for prediction.

In summary, under the different metro passenger flow modes of holidays and working days, the SN-LSSVM model can be maintained as a satisfying requirement, since its prediction accuracy is high, and it has greater stability and applicability compared to the state-of-the-art time series forecasting methods.

4.6 Efficiency analysis

Our approach consists of an offline training phase and an online forecasting phase. In the training phase, the major time cost depends on the training time cost that the predictive model uses. The computational complexity using the LSSVM with RBF kernel is $O(n^2 \cdot d)$ [28], where n is the training size and d denotes the number of features. To be specific, the average time cost of the SN-LSSVM training is approximately 5.5 minutes in our dataset (20 days \times 18+15 weeks \times 126 slots training samples), which means, the forecasting process can achieve real-time prediction.

5. CONCLUSION

In recent years, short-term traffic flow prediction has been developing, and it will undoubtedly keep improving. However, it is unfortunate that there is only one small study which specifically deals with short-term subway passenger flow prediction. Moreover, these models did not fully consider the key factors of short-term metro passenger flow forecasting. Thus, this article mainly aims to improve the prediction accuracy of metro passenger flow time series prediction models, whether the forecasting target is a working day or a holiday.

This paper establishes a novel SN-LSSVM hybrid model for short-term subway passenger flow prediction, which is composed of three stages:

traffic data profiling, passenger flow analysis, and predictive modelling. The proposed model has great potential for solving many other relevant prediction problems, which can extract the periodicity and non-linearity characteristics of the time series. Based on the research results, the following conclusions can be drawn: (1) Nonparametric nonlinear regression model LS-SVM is more suitable for capturing the volatility characteristics of the time series data, and is more flexible for nonlinear and high-dimensional AFC data than LSTM. (2) External characteristics of the time series such as periodicity or nonlinearity contribute to short-term passenger flow prediction; (3) A smaller forecasting horizon has a more stable predictive result than a larger one. (4) The SN-LSSVM model can be maintained as a satisfying requirement for forecasting year-around metro passenger flow. We first applied the method to the Tiyu Xilu station, which is the most crowded station of Guangzhou metro, and it contributed to the management of the station.

In future studies, several improvements should be considered. (a) external characteristics of the time series such as examples from previous years can be considered, (b) other external influential factors of ridership, such as spatio-temporal information in the metro network, can be added to the method if sufficient data are available [4], (c) another important issue is forecasting passenger flow under special circumstances, such as abnormal weather, large-scale events, and emergencies.

ACKNOWLEDGEMENTS

The authors would like to thank the Ministry of Science and Technology (P. R. China) for its funding and support (National Key R&D Program of China. No. 2016YFC0802500).

黄鑫, 汪益敏, 林培群, 于恒, 罗跃
基于多模型融合的全天候短期地铁客流预测

摘要

准确的地铁乘客量预测可以指导乘客有效地选择出发时间, 同时帮助交通运营商制定乘客组织策略。然而, 短期客流预测需要考虑许多因素, 现有的短期地铁客流预测模型的结果往往不能令人满意。在本文中, 我们专注于充分挖掘时间序列定律并提高预测的准确性。沿着这条主线, 我们提出了一种并行的模型结构以提取客流的周期性和非线性特征, 该模型称为季节性和非线性最小二乘支持向量机 (SN-LSSVM)。我们使用了包括整合移动平均自回归模型、长短期记忆神经网络、支持向量机等

预测模型评估所提出预测模型的性能。此外，我们将该算法在广州地铁拥有最庞大客流的体育西路站首次应用，实验表明该方法能有效预测全年，全天候的客流，为地铁站的组织管理贡献了力量。

关键词

小波分析；最小二乘支持向量机；短期地铁客流预测；多模型融合预测；时间序列。

REFERENCES

- [1] Urban Rail Transit Technical Working Committee of China Civil Engineering Society. Proposal for Urban Rail Transit Technology Development Outline (2010-2015). *Urban Rapid Transit Traffic*. 2010;23(06): 3-8. DOI: 10.3969/j.issn.1672-6073.2010.06.002
- [2] He Y, Sheng WW, Luo AH. Track extension, you and I travel more smoothly. *People's Daily*. February 3, 2019. Available from: http://paper.people.com.cn/rmrb/html/2019-02/13/nw.D110000renmrb_20190213_2-07.htm [Accessed 23rd March 2020].
- [3] Yang J. *Research on Metro Passenger Flow Short-time Prediction and Evacuation Simulation*. PhD thesis. Beijing Jiaotong University; 2013.
- [4] Smith BL, Demetsky MJ. Traffic flow forecasting: Comparison of modeling approaches. *Journal of Transportation Engineering*. 1997;123(4): 261. DOI: 10.1061/(ASCE)0733-947X(1997)123:4(261)
- [5] Yang ZS, Wang Y, Guan Q. Short-time traffic flow prediction method based on support vector machine method. *Journal of Jilin University (Engineering and Technology Edition)*. 2006;26(6): 881-884. DOI: 10.13229/j.cnki.jdxbgxb2006.06.010
- [6] Vanajakshi L, Rilett LR. Support Vector Machine Technique for the Short Term Prediction of Travel Time. In: *IEEE Intelligent Vehicles Symposium, 13-15 June 2007, Istanbul*. IEEE; 2007. p. 600-605. DOI: 10.1109/IVS.2007.4290181
- [7] Wang Y, Zheng D, Luo S, Zhan DM. The research of railway passenger flow prediction model based on BP neural network. *Advanced Materials Research*. 2013; 605-607: 2366-2369. DOI: 10.4028/www.scientific.net/AMR.605-607.2366
- [8] Guo SY, Li WQ, Bai W, Zhang D. Prediction of Short-term Passenger Flow on a Bus Stop Based on LSSVM. *Journal of Wuhan University of Technology (Transportation Science & Engineering)*. 2013;37(03): 603-607. DOI: 10.3963/j.issn.2095-3844.2013.03.037
- [9] Guo JY, Wang ZJ, Chen HW. On-line multi-step prediction of short term traffic flow based on GRU neural network. *Proceedings of the 2nd International Conference on Intelligent Information Processing*; 2017. p. 7-12. DOI: 10.1145/3144789.3144804
- [10] Wang XM, Zhang N, Zhang YL, Shi ZB. Forecasting of Short-Term Metro Ridership with Support Vector Machine Online Model. *Journal of Advanced Transportation*. 2018. DOI: 10.1155/2018/3189238
- [11] Kong WC, Dong ZY, Jia YW, Hill DJ. Short-Term Residential Load Forecasting Based on LSTM Recurrent Neural Network. *IEEE Transactions on Smart Grid*. 2019;10(1): 841-851. DOI: 10.1109/TSG.2017.2753802
- [12] Lu WX, Li C. Forecasting of Short-Time Tourist Flow Based on Improved PSO Algorithm Optimized LSSVM Model. *Computer Engineering and Applications*. 2019;55(18): 247-255. DOI: 10.3778/j.issn.1002-8331.1807-0063
- [13] Boto GD, Díaz PFJ, González OD, et al. Wavelet-based denoising for traffic volume time series forecasting with self-organizing neural networks. *Computer-Aided Civil and Infrastructure Engineering*. 2010;25(7): 530-545. DOI: 10.1111/j.1467-8667.2010.00668.x
- [14] Pan L. *Short-time Forecasting of High-speed Railway Passenger Flow Based on Ensemble Empirical Mode Decomposition-Gray Support vector Machine*. PhD thesis. Beijing Jiaotong University; 2012.
- [15] Wei Y, Chen MC. Forecasting the short-term metro passenger flow with empirical mode decomposition and neural networks. *Transportation Research Part C: Emerging Technologies*. 2012;21(1): 148-162. DOI: 10.1016/j.trc.2011.06.009
- [16] Bouzerdoum M, Mellit A, Massi PA. A hybrid model (SARIMA-SVM) for short-term power forecasting of a small-scale grid-connected photovoltaic plant. *Solar Energy*. 2013;98: 226-235. DOI: 10.1016/j.solener.2013.10.002
- [17] Sun YX, Leng B, Guan W. A novel wavelet-SVM short-time passenger flow prediction in Beijing subway system. *Neurocomputing*. 2015;166: 109-121. DOI: 10.1016/j.neucom.2015.03.085
- [18] Tang LY, Zhao Y, Cabrera J, Ma J. Forecasting Short-Term Passenger Flow: An Empirical Study on Shenzhen Metro. *IEEE Transactions on Intelligent Transportation Systems*. 2019;20(10): 3613-3622. DOI: 10.1109/TITS.2018.2879497
- [19] Ma XL, Zhang JY, Du BW, Ding C. Parallel Architecture of Convolutional Bi-Directional LSTM Neural Networks for Network-Wide Metro Ridership Prediction. *IEEE Transactions on Intelligent Transportation Systems*. 2019;20(6): 2278-2288. DOI: 10.1109/TITS.2018.2867042
- [20] Bates JM, Granger CWJ. The Combination of Forecasts. *Journal of the Operational Research Society*. 1969;20(4): 451-468. DOI: 10.1057/jors.1969.103
- [21] People's Network. *360 big data release Spring Festival 'empty city index' Beijing and Shanghai have not entered the top five*. Available from: http://www.sohu.com/a/125388616_114731 [Accessed 23rd March 2020].
- [22] Sain SR, Vapnik VN. The Nature of Statistical Learning Theory. *Technometrics*. 1996;38(4): 409. DOI: 10.1080/00401706.1996.10484565
- [23] Vapnik VN. An overview of statistical learning theory. *IEEE Transactions on Neural Networks*. 1999;10(5): 988-999. DOI: 10.1109/72.788640
- [24] Cherkassky V, Ma YQ. Practical selection of SVM parameters and noise estimation for SVM regression. *Neural Networks*. 2004;17(1): 113-126. DOI: 10.1016/S0893-6080(03)00169-2
- [25] Eberhart R, Kennedy J. A New optimizer using particle swarm theory. In: *Proceedings of the Sixth International*

- al Symposium on Micro Machine and Human Science, 4-6 October 1995, Nagoya, Japan.* IEEE; 1995. p. 39-43. DOI: 10.1109 / MHS.1995.494215
- [26] Zhu W, Wang W, Huang ZD. Estimating train choices of rail transit passengers with real timetable and automatic fare collection data. *Journal of Advanced Transportation.* 2017. DOI: 10.1155/2017/5824051
- [27] Zhang GP. Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing.* 2003;50: 159-175. DOI: 10.1016/S0925-2312(01)00702-0
- [28] Chapelle O. Training a support vector machine in the primal. *Neural Computation.* 2007;19(5): 1155-1178. DOI: 10.1162/neco.2007.19.5.1155