Passenger Flow Prediction of High Speed Railway Based on LSTM Deep Neural Network

Jie Li [a,b], Ping Huang [a], Yuxiang Yang [a] ,Qiyuan Peng[a,1]
[a] School of Transportation and Logistics, Southwest Jiaotong University
High-tech Zone West Campus, 611756, Chengdu, China
[b] Civil, Buildings and Environmental Engineering Department, SAPIENZA Università di Roma
Via Eudossiana 18, 00184 Roma, Italy
[1] E-mail: qiyuan-peng@swjtu.cn, Phone: +86 13808061287

Abstract
The paper presents the characteristics of the departing passenger flow in different stations based on the real-record passenger flow data of Wuhan-Guangzhou high speed railway, from January, 2010 to December, 2015. The passenger dataset is framed for the long short-term memory (LSTM) model, considering the expectation input format of LSTM layers and the characteristics of the data. The Keras model in Python is used to fit LSTM model with tuning and regulating all the parameters necessary in the model. Then the fitted LSTM model is applied to forecast the short-term departing passenger flow of Wuhan-Guangzhou high speed railway. The influence of important parameters in the LSTM model on the prediction accuracy is analysed, and the comparison with other representative passenger flow forecast models is conducted. The results show that the LSTM model can get the valid information in a long passenger flow time series and achieve a better performance than other models. The passenger flow prediction errors valued by MAPE are 7.36%, 7.33%, 8.03%, respectively for Chenzhou station, Hengyang station and Shaoguan station. The parameters in the LSTM model such as the number of hidden units, the batch size and the input historical data length have a great influence on the prediction accuracy.

Keywords
High speed railway, Passenger flow prediction, Long short-term memory model, Deep learning, Time series

1 Introduction

The HSR has developed significantly in China due to its efficiency in transporting large numbers of passengers within short travel times. The short-term forecasting of high-speed rail passenger flow is one of the most critical issues because passenger demand provides a reference for seat allocation, ticket booking and train routing. The daily-based passenger demand in the near future is essential for the railway revenue management.

Short-term passenger flow prediction has a long history, and many successful models have been developed for this issue. These models can be generally divided into three categories: parametric approaches, nonparametric approaches and hybrid models.

In general, parametric approaches are model-based methods, whose structure is predetermined based on certain theoretical assumptions and the model parameters can be computed with empirical data. A variety of parametric models have been applied on traffic forecasting, such as the grey forecasting model, exponential smoothing model (Kyungdo (1995) and Tan (2009)), Kalman filtering models((Chen (2001), Chien (2003),Wang(2006) and Van (2008)), state space model (Liu (2006)) and so on. The most widely used parametric method is Autoregressive Moving Average (ARIMA) model, which assumes the traffic condition is a stationary process. ARIMA performs well and is effectively in modelling linear and stationary time series. A number of ARIMA based time series models have been proposed for traffic prediction (Moreira (2013), Williams (2003), Smith (2002), Williams (2001) and Chandra (2009)).However, the parametric approaches cannot work well on stochastic and nonlinear data, thus the nonparametric methods are developed to forecast the traffic flow with stochastic and nonlinear.

In the nonparametric approaches, the parameters and the structure of the nonparametric approaches are uncertain. The non-parametric models used for traffic forecasting include support vector regression (Wu (2004), Zhang (2009), Asif (2014) and Zhang (2007)), neural networks (Çetiner (2010) and Tsai (2009)), Kalman filtering(Van Lint (2008) and Wang (2007)), Gaussian maximum likelihood (Tang (2003)) and so on. SVM is an artificial intelligence method based on the structural risk minimization principle and has the potential to overcome the problems of nonlinearity, small samples, high dimension, local minima and over-fitting. Neural networks are capable of handling multi-dimensional data with flexible model structure, strong learning ability as well as adaptability. The Neural networks has been applied in many researches (Karlaftis (2011), and Ma (2015)).However the neural networks have drawbacks of the potential of over fitting, the requirement of large train samples and the cost of long training time.

Third, hybrid models have been proposed for a better performance in passenger performance. Zhang (2014) proposed a hybrid EEMD-GSVM model and applied the model to forecast the short-term passenger flows of three typical origin–destination pairs in terms of travel distances. Wei (2012) forecasted metro passenger flows with a hybrid of EMD and neural networks that generated higher forecasting accuracy and stability than the seasonal ARIMA. Zhu (2007) presented a hybrid method based on EMD and SVM for short-term electronic load forecasting. Li (2014) proposed an ensemble learning framework to appropriately combine estimation results from multiple macroscopic traffic flow models. Khashei (2012) proposed a new hybrid model of the autoregressive integrated moving average (ARIMA) and probabilistic neural network (PNN) to yield more accurate results than traditional ARIMA models.

Although numerous passenger flows forecast models have been developed, the short-term forecast of HSR passenger flow is still challenging because daily passenger flows are highly oscillated, nonlinear and non-stationary. In addition, most HSR lines in China are still under development, while passenger flows of opened HSR lines can be influenced by unstable demands such as holidays.

Currently, deep learning has been successfully applied in many fields and achieved reasonable results (Srivastava (2015), Donahue (2017) and Polson (2017)). Ma (2017) proposed a deep convolutional neural network for large-scale traffic network speed prediction. Yu (2017) designed a spatiotemporal recurrent convolutional network for predicting network-wide traffic speeds. Meanwhile, big data has revolutionized the transportation industry over the past several years. These two hot topics have inspired us to reconsider the traditional issue of passenger flow prediction. In this paper, a HSR passenger flow forecasting model based on LSTM is proposed.

The passenger flow sequence of HSR is nonlinear time series. The interaction in the passenger time series should be considered to forecast the short-term passenger flow. Most of the current passenger flow prediction model cannot take advantage of the effective information in the passenger flow time series. LSTM is one kind of deep neural network and the model is fitted based on the big data of passenger flow. LSTM can capture the nonlinearity and randomness of traffic flow more effectively, as well as overcome the issue of back-propagated error decay through memory blocks, and thus shows superior capability for time series prediction with long temporal dependency. In this paper, daily ticket data on the Beijing- Guangzhou HSR was collected from January, 2010 to December, 2015. The proposed LSTM passenger forecast model is applied to forecast the passenger flow of Beijing- Guangzhou HSR.

The remainder of this paper is organized as follows: Section 1 provides a general overview of the existing approaches of traffic flow forecasting and the application. The long short-term memory neural network architecture is present and the passenger prediction model based on LSTM is introduced in section 2. In addition, the performance of the LSTM is evaluated, compared to other models such as Support Vector Machine (SVM), K-Nearest Neighbor (KNN) and Random Forest (RF). In section 4 the effect of parameters in the LSTM on the prediction performance is analysed. Finally, conclusion and future envisions are discussed in section 5.

2 Passenger Flow Prediction model Based on LSTM

2.1 Structure of the Memory Unit of LSTM

Recurrent neural network (RNN) is a powerful deep neural network which can deal with sequence data using the internal memory. The architecture of RNN is illustrated in Figure 1. RNN contains input layer X , hidden layer S and output layer O . U , V , W are weight vectors. At the time t the hidden layer S_t and the output O_t can be calculated as Equation (1) and Equation (2).

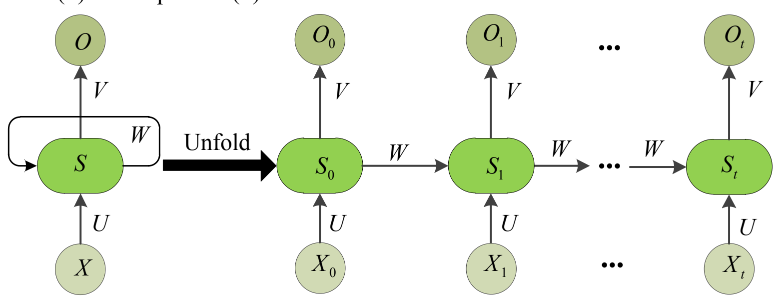

Figure 1: Standard RNN architecture

$$S_t = f\left(UX_t + WX_{t-1}\right) \tag{1}$$

$$O_t = g\left(VS_t\right) \tag{2}$$

Thus the hidden vector S_t at time t is determined by the input vector at time t and hidden vector at the previous time $t-1$ while the output O_t is determined by the historic input $X_t, X_{t-1}, X_{t-2} \ldots X_1$.

In principle, RNN can map the whole historical input data to each output, relying on the key point that the recurrent connections allow the memory of previous input to affect the network's output. However, in standard RNN architecture, the given weight vector in the hidden layer plays an important role on the network output, which can lead to either decays or blows up exponentially as it cycles around the recurrent connections in the networks for too many times. This effect is often considered as the vanishing gradient problem. Thus, RNN is incapable of learning from long time lags, or saying long-term dependencies (Bengio (2002)).

To address the problem, a LSTM is proposed to work well on modelling long-term time series. The difference between standard LSTM architecture and the RNN architecture is the hidden layer, which enhance the LSTM to avoid vanishing gradient problem. LSTM is a special kind of RNN. By treating the hidden layer as a memory unit, LSTM network can get the valid information in a long passenger flow time series. The typical architecture of LSTM memory unit is shown in Figure 2.

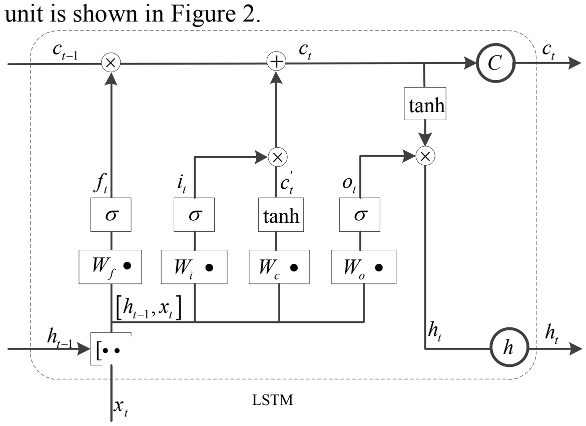

Figure 2: The structure of LSTM unit cell

There is a memory cell in the unit, denoted by C . Moreover, the LSTM memory unit contains three gates, namely input gate i_t, forget gate f_t and output gate o_t .The state of the memory cell at time t is indicated by c_t ,the input of every gate contains the preprocessed data x_t and the previous output of the LSTM unit, called h_{t-1} .Based on the information flow in the structure of memory unit, the update of the memory cell' state can be summarized as Equation (3) to Equation (8).

$$f_t = \sigma\left(W_f \bullet [h_{t-1}, x_t] + b_f\right) \tag{3}$$

$$i_t = \sigma\left(W_i \bullet [h_{t-1}, x_t] + b_i\right) \tag{4}$$

$$c_t^{'} = \tanh\left(W_c \bullet [h_{t-1}, x_t] + b_c\right) \tag{5}$$

$$c_t = f_t \circ c_{t-1} + i_t \circ c_t{'} \tag{6}$$

$$o_t = \sigma\left(W_o \bullet [h_{t-1}, x_t] + b_o\right) \tag{7}$$

$$h_t = o_t \circ \tanh(c_t) \tag{8}$$

i_t, f_t and o_t are the output of different gates, c_t is the new state of memory cell and h_t is the final output of the LSTM unit. W_f , W_i , W_c , W_o are coefficient matrixes, b_f , b_i , b_c , b_o represent the offset vectors, σ is the weight of the sigmoid function, and *tanh* is the hyperbolic tangent activation function. Via the function of the different gates, LSTM memory units can capture the complex correlation features within time series in both short and long term, which is a remarkable improvement compared with RNN.

2.2 LSTM Network for Passenger Flow Prediction

In the proposed LSTM model, the daily historical passenger flow data at one station can be viewed as a time sequence $P=\{p_1, p_2, ..., p_i, ..., p_n\}$. In general the passenger flow is non-stationary time series with increase and periodicity trend. The non-stationary of the time series affects the prediction accuracy of the LSTM model. Stationary data is easier to model and will very likely result in more skillful forecasts. A standard way to remove a trend is by differencing the data. That is the observation from the previous time step (t-1) is subtracted from the current observation (t). This removes the trend and we are left with difference series or the changes to the observations from one time step to the next.

Thus the passenger dataset should be framed for the long short-term memory (LSTM) model, considering the expectation input format of LSTM layers and the characteristics of the data.

The passenger flow $P=\{p_1, p_2, ..., p_i, ..., p_n\}$ can be transformed from time series to stationary by two steps, one is rolling window smoothing with M order and the other step is the differencing process. The rolling window smoothing process can remove the periodicity in the time series, after which the passenger data can be denoted as $P^W=\{p_1^w, p_2^w, ..., p_{n-M+1}^w\}$. The differencing process can remove the increase trend in the time series, after which the passenger data can be denoted as $P^D=\{p_1^d, p_2^d, ..., p_{n-M}^d\}$. Below are functions calculating the rolling window smoothing and differenced series.

$$p_t^w = (p_t + p_{t+1} + ... + p_{t+M-1}) / \mathrm{M} = \frac{1}{M}\sum_{i=t}^{t+M-1} p_i \tag{9}$$

$$p_t^d = p_{t+1}^w - p_t^w \tag{10}$$

Like other neural networks, LSTM expect data to be within the scale of the activation function used by the network. The default activation function for LSTMs is the hyperbolic tangent (*tanh*), which outputs values between -1 and 1. This is the preferred range for the time series data.

We can transform the dataset to the range [-1, 1] using the MinMaxScaler class. The function below inverts this operation. Again, we must invert the scale on forecasts to return the values back to the original scale so that the results can be interpreted and a comparable error score can be calculated.

$$x_{n,t+1} = \frac{p_n^d - \min(P_n^D)}{\max(P_n^D) - \min(P_n^D)} \tag{11}$$

The LSTM model in Keras assumes that your data is divided into input (X) and output (Y) components. Suppose we need to predict the passenger flow $P_{out}=\{p_{t+1}, p_{t+2}, ...p_{t+n}\}$ of time duration $T=\{t+1, t+2, ...t+n\}$ using the of m historical time steps passenger flow $P_{in}=\{p_{t-m+1}, p_{t-m+2}, ...p_t\}$, we can concatenate these two series together to create data frame $\{X_{in}, Y_{out}\}$ for supervised learning. Let us denote the input of LSTM model $X_{in}=\{X_1, X_2, ...X_j ... X_{n-M-L-F+1}\}$, $X_j=\{x_j, x_{j+1}, ...x_{j+L-1}\}$, $x_j \in X$, $|X_j| = L$ .The output of LSTM model $Y_{out}=\{Y_1, Y_2, ...Y_j ... Y_{n-M-L-F+1}\}$, $Y_j=\{x_{j+L}, x_{j+L+1}, ...x_{j+L-1+F}\}$,

$x_{j+L} \in X$ and $|Y_j| = F$ .The $Dataframe = \{X_{in}, Y_{out}\}$ Should be divided into training datasets $Datatrain = (X_{Train}, Y_{Train})$ and test dataset $Datatest = (X_{Test}, Y_{Test})$. The training dataset are used to fit the model and the test dataset is used to evaluate the performance of the fitted model.

Given that the training dataset is defined as X inputs and Y outputs. Let us denote the input passenger time series as $X = \{x_1, x_2, ..., x_m\}$, hidden state of memory cells as $H = \{h_1, h_2, ..., h_m\}$ and output passenger prediction time series as $Y = \{y_1, y_2, ..., y_m\}$, LSTM works the computation as Equation (12) to Equation (13).

$$h_t = H(W_{hx}x_t + W_{hh}h_{t-1} + b_h) \tag{12}$$

$$p_t = W_{hy}y_{t-1} + b_y \tag{13}$$

The objective of the passenger flow prediction is to minimize the difference between the actual passenger flow and the predicted passenger flow. The square loss function given by the following formula is used as the objective function, in which y_t represents the actual passenger flow and p_t represents the predicted passenger flow.

$$e = \sum_{t=1}^{n} (y_t - p_t)^2 \tag{14}$$

In order to minimize training error and meanwhile avoid local minimal points, Adam optimizer, a modification of stochastic gradient descent (SGD) optimizer with adaptive learning rates, is applied for back propagation through time (BPTT).

The prediction accuracy of short-term traffic flow can be assessed by two commonly used metrics, i.e., Mean Absolute Percentage Error (MAPE) which evaluates the relative error and Root Mean Square Error (RMSE) which evaluates the absolute error. They are defined by Equation (15) and Equation (16).

$$MAPE = \frac{1}{n} \sum_{i=1}^{n} \frac{|p_{Prediction,i} - p_{Test,i}| \times 100\%}{p_{Test,i}} \tag{15}$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (p_{Test,i} - p_{Prediction,i})^2} \tag{16}$$

The flowchart of short-term passenger prediction based on LSTM is shown as Figure 3.

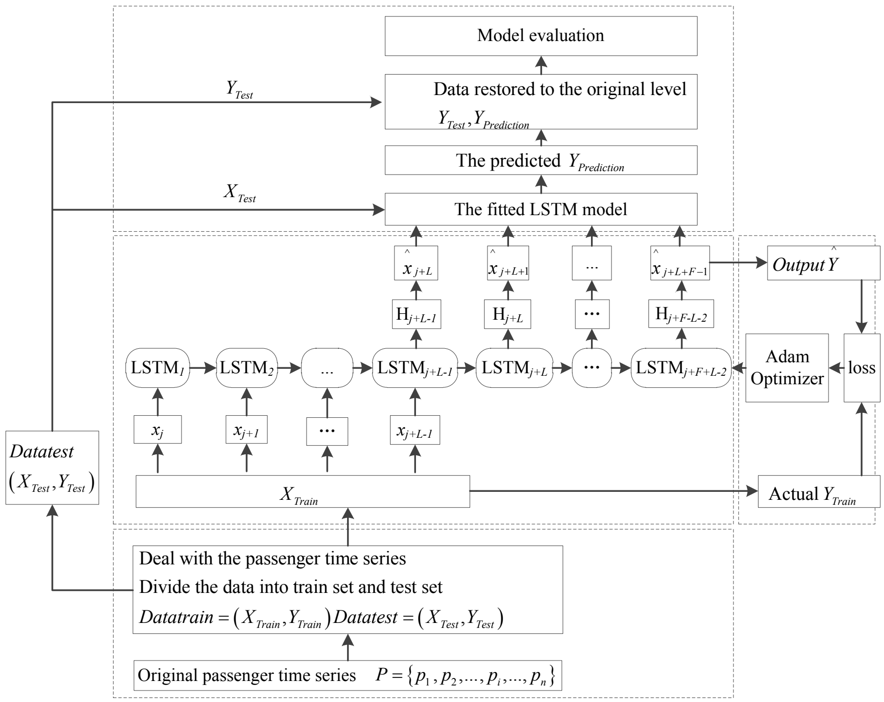

Figure 3: The flowchart of short-term passenger prediction model based on LSTM

3 Experiments and Results

3.1 Dataset Description

The passenger flow departing from Chenzhou station, Hengyang station and Shaoguan station, which locates on Wuhan-Guangzhou HSR, are taken as examples to demonstrate the efficiency of the LSTM based passenger prediction model.

The passenger volume data are collected every day from the booking tickets system, from 1st January, 2010 to 30th December, 2015, 2174 days in total. Part of the original dataset is shown in Figure 4. There is a big difference among the number of passengers departing form the three stations. Thus the performance of the LSTM Model on different grand passenger volume can be evaluated.

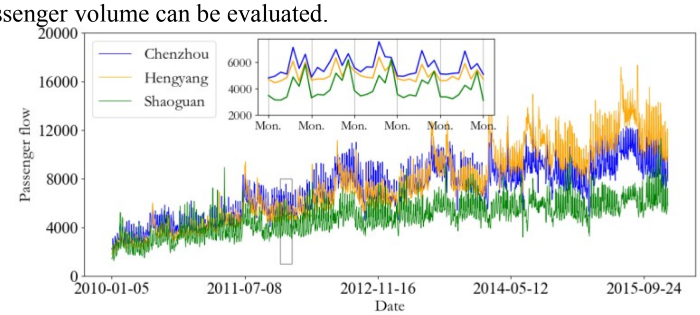

Figure 4: The passenger flow time series of the stations

The passenger volume increased during vacation, which may affect the prediction precision of LSTM model. Thus the passenger data of vacation are removed from the time series, and then 1673 days left.

The passenger time series of each station are shown in Figure 4. The passengers series present an increase trend as well as a significant cyclical with a period of 7 days. The passenger peak days appears on Friday and Sunday while the passenger trough appears on Sunday. The passenger data should be transformed from time series to LSTM data, just follow the data process in part 2. The prepared data for LSTM is shown in Figure 5.

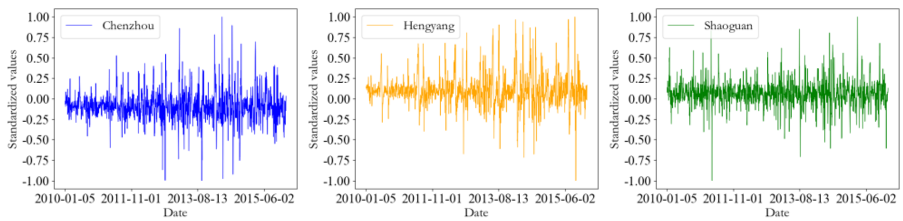

Figure 5: The standardized passenger flow time series

The passenger flow of Chenzhou station, Hengyang station and Shaoguan station are taken as examples to demonstrate the efficiency of the LSTM based passenger prediction model. Since the passenger volume presents a significant periodicity of 7 days, the objection of the model is to predict the passenger volume in the following days by means of the data of the previous seven days. To demonstrate the efficiency of the LSTM model as well as simplify the passenger prediction problem, LSTM model is just applied to forecast the passenger volume in the next day.

The passenger data is divided into two parts, the first 80% of the data is used to train the model, and the last 20% of the data is used to test the prediction accuracy of the model. To validate the efficiency of the proposed LSTM network, the performance is compared with some conventional forecast models, include ARIMA, SVM, RF, KNN.

Some key parameters should be determined for the short-term passenger flow prediction based on LSTM-RNN, including the size of input layer, the number of hidden layers, and the number of hidden units in each of hidden layer, the batch size and the size of output layer. The input historical data length is equal to the size of input layer, which is defined as 7 in the experiment. The number of hidden layers is assigned as 1,2,4,6,8 and the number of units in each hidden layer is assigned as 5,10,20,50,75,100. The size of output layer is 1, indicating the passenger flow in the next step will be forecasted. Grid search method is used to obtain the optimal parameters. The candidate values of the parameters are shown in Table 1. We performed a grid search over this parameter in order to find the size that leads to the best results. The grid search results of the optimal model parameters and the prediction precision are listed in Table 2.

Table 1: The parameters and hyper parameters of LSTM model

parameters	**Values**
Input historical data length	7
Output size	1
Epoch	1000
Optimizer	Adam
Learning rate	0.0001
Dropout	0.3
Loss function	Mean_Squared_Error
Activation function	Tanh
hyper parameters	**Values**
Batch size	1,2,4,6,8,10,12,14,16
Hidden Unit	5,10,20,50,75,100
Architecture	Input layer→LSTM layer→LSTM layer→Dropout layer→Fully connected layer→Output layer

Table 2: The optimal parameters of LSTM model for different stations

Station	**Batch size**	**Hidden unit**	**MAPE**	**RMSE**
Chenzhou	1	10	7.21%	759.582
Hengyang	1	10	7.28%	800.227
Shaoguan	1	10	7.79%	562.000

3.2 Prediction Performance Analysis

In this section, we use the same experimental setup and fit the model for 1000 training epochs. A line plot of the series of RMSE scores on the train and test sets after each training epoch is also created, which is shown in Figure 6. The result clearly shows a downward trend in RMSE over the training epochs for the experimental runs of the three stations. The lines for the all the train case shows a sharp decrease before 200 epochs and then become more horizontal, but still generally show a downward trend, although at a lower rate of change. The lines for the test case of Chenzhou station, Hengyang station and Shaoguan station show a downward trend respectively before 500 epochs, 50 epochs and 400 epochs, and then the lines become more horizontal.

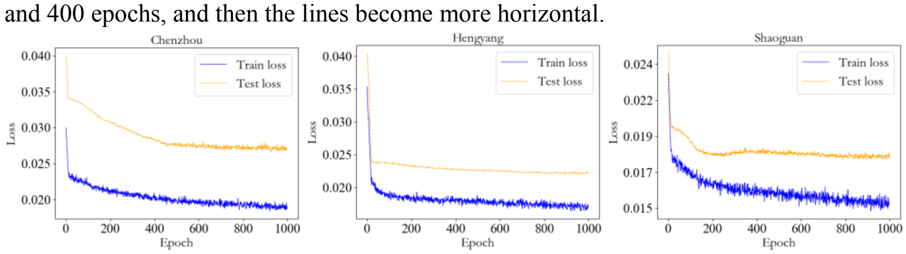

Figure 6: Diagnostic results with 1000 Epochs

Figure 7 shows the passenger flow comparison between observation values and the prediction values obtained by the LSTM. The prediction results are fairly good for Chenzhou, Hengyang and Shaoguan station, whose MAPE are 7.26%,7.33%,8.03% respectively.

In general the LSTM model is well capable of predicting the passenger volume trend.

The prediction accurate is high as the passenger flow shows a regular trend. However, when dramatic changes in the passenger flow are observed, the prediction accurate is low.

The distribution of MAPE over the predicted values is shown in Figure 7. We could find that most of the MAPE are located in （0,10%）. Specifically, 52.7%, 57.5%, 57.5% of the MAPE is less than 5%, and 81.9%, 79.2%, 82.4% of the MAPE is less than 10%, respectively for Chenzhou, Hengyang and Shaoguan station.

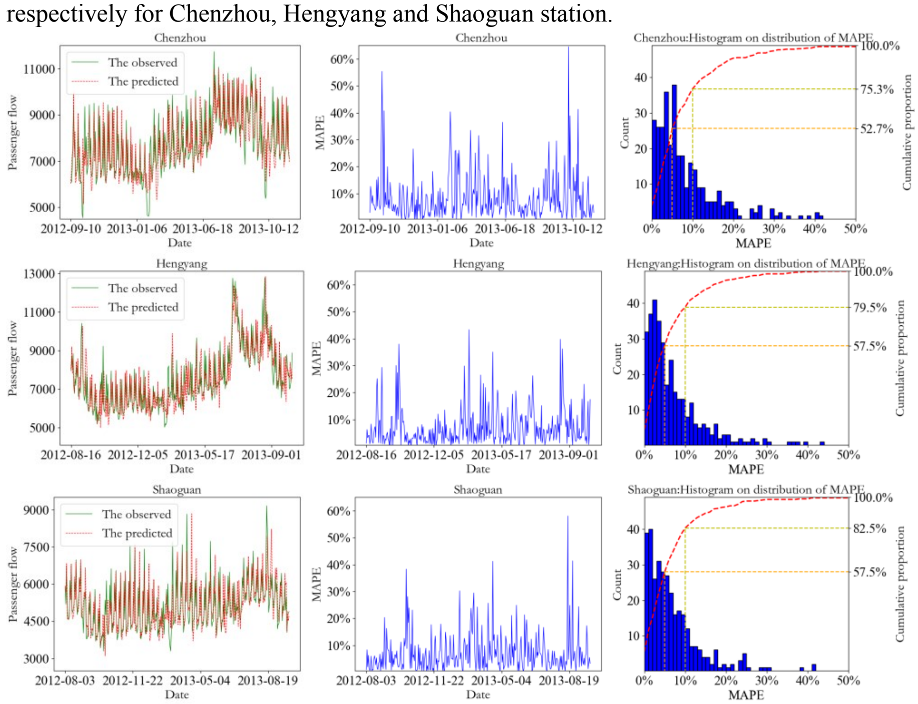

Figure 7: The prediction passenger flow and MAPE

To validate the efficiency of the proposed LSTM network, the performance is compared with some conventional forecast approaches; include ARIMA, SVM, RF, KNN. Each prediction method is tested for 10 times to avoid the randomness. The experimental results are shown in Table 3.As we can see from Table 3, compared to other methods, the MAPE of LSTM are the lowest. For the RMSE, the GB method performed best for the passenger volume prediction at Hengyang station while the RMSE of LSTM at other stations is the lowest.

Table 3: Prediction results of different models

Model	Chenzhou		Hengyang		Shaoguan	
	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE
LSTM	**778.74**	**7.26%**	801.46	**7.33%**	**584.29**	**8.03%**
RF	861.45	8.14%	844.41	7.89%	628.60	8.96%
GB	831.37	7.80%	**801.15**	7.36%	602.41	8.44%
KNN	860. 74	8.11%	805.26	7.56%	625.07	8.51%
SVM	796.76	7.55%	816.51	7.34%	591.31	8.21%

For a further analysis of the prediction efficiency and the stability of different prediction models, RMSE and MAPE distributions of each model is shown on a box and whisker plot in Figure 8.

The red line shows the median and the box shows the 25th and 75th percentiles, or the middle 50% of the MAPE. The values of the red line give an idea of the average expected performance of a configuration whereas the box gives an idea of the range of possible best and worst case examples that might be expected.

Looking at just the median RMSE scores, the results suggest that the choice of LSTM to predict the passenger volume is better than the other models since the median RMSE scores of LSTM for every station are the lowest and the average expected performance of LSTM is good. In terms of the stability of different prediction models, the comparison of the boxes suggest that the performance of the FR model is unstable since the gap between the 25th and 75th percentiles MAPE scores is large. The performance of LSTM model is relative stable while GB, KNN, SVM model is very stable.

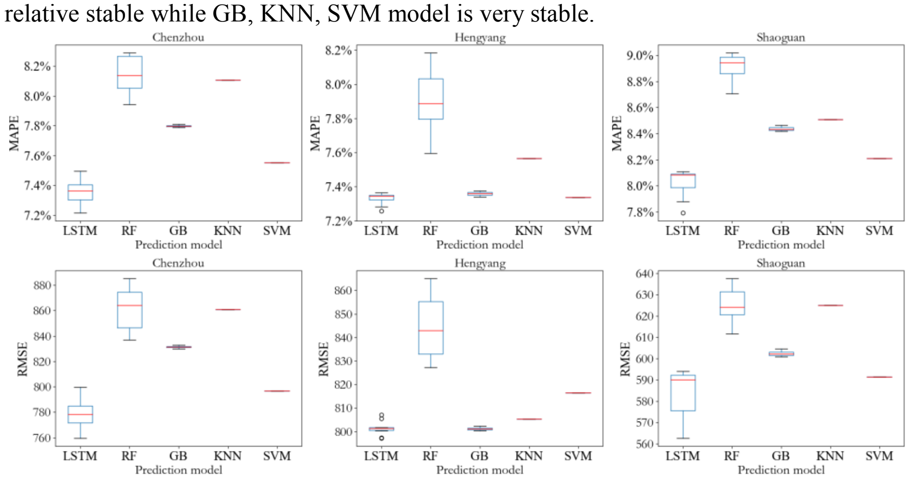

Figure 8: Boxplot of prediction RMSE and MAPE of different models

To sum up the above analysis, LSTM RNN model is capable of memorizing long historical data and achieving higher prediction accuracy even if the model is quite simple. Therefore, the proposed model is effective in short-term traffic flow prediction.

4 Analysis on the Influence of Model Parameters on Prediction Accuracy

4.1 The Number of Hidden Units

The number of hidden units in each of hidden layer affects the learning ability of the network. Generally, more neurons would be able to learn more structure from the problem at the cost of longer training time. More learning capacity also creates the problem of potentially over fitting the training data.

The effect of hidden units on the prediction results is investigated by assigning the number of units as 5, 10,20,50,75,100. We can objectively compare the impact of increasing the number of neurons while keeping all other network configurations fixed. We will use a batch size of 1 and 1000 training epochs

In order to alleviate the influence of random initialization for the model, we repeat each experiment 30 times and compare the average test RMSE performance with the number of neurons, the result is shown in Figure 9.

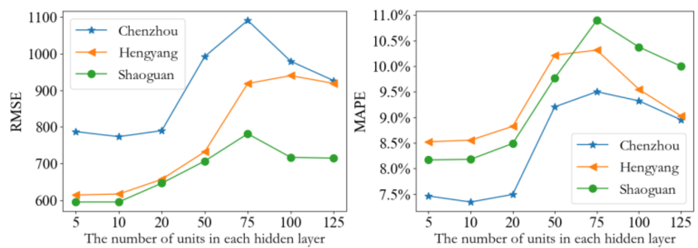

Figure 9: The distribution of prediction MAPE and RMSE of models with different hidden units

From Figure 9, we can see that the MAPE and RMSE remain stable when the number of the number of hidden units is less than 20, and the values of which are low. As the number of hidden units is more than 20 and less than 75, the MAPE and RMSE rise up with the increase of the number of hidden units. The MAEPs and MSEs of Hengyang station reach to the highest as the hidden units is 100, while the MAPE of the other stations decrease. As the number of hidden unit is 100, the LSTM model perhaps show an acceleration of over fitting.

Specially, diagnostic with1000 epochs and various neuron of Hengyang station are taken as an example to demonstrate the effect of the neurons on the LSTM. As the number of neurons is 5 and 10, both the line of train loss and test loss show horizontal. The results suggest a good, but not great, general performance. It shows a rapid decrease in test RMSE as the neurons is 10, which means the learning capacity of the network is improved as the number of the neurons increase from 5 to 10.

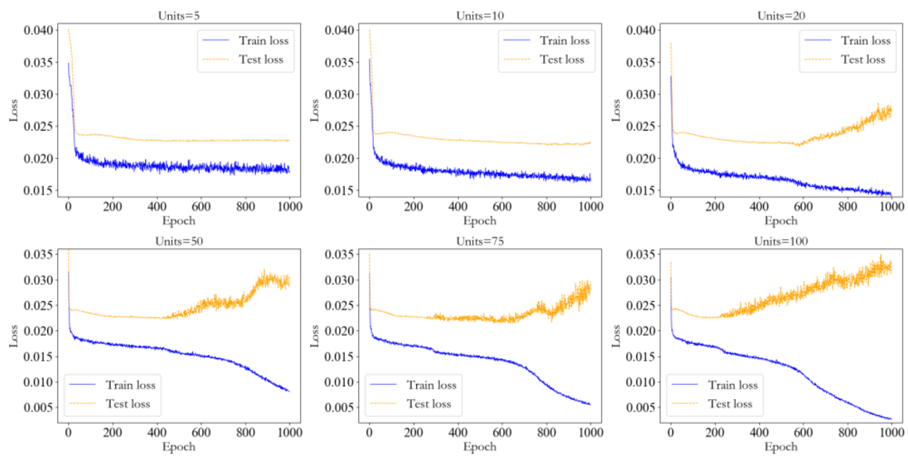

Figure 10: Diagnostic results of models with different hidden units

The diagnostic results of models with different hidden units are shown in Figure 10. As the number of the neurons is 20, 50, 75, diagnostic results shows a rapid decrease in test RMSE to about epoch 500-600. Meanwhile, the training dataset shows a continued decrease to the final epoch. These are significant signs of over fitting of the training dataset. When the number of the neurons is 100, the inflection point in the training dataset seems to be happening sooner than the 20, 50, 75 neurons experiment, perhaps at epoch 300-400.

It can be proved that more neurons can enhance the learning ability of LSTM network. However, too many neurons may lead to an over fitting of the training dataset. These increases in the number of neurons may benefit from additional changes to slowing down the rate of learning, such as the use of regularization methods like dropout, decrease to the batch size, and decrease to the number of training epochs.

4.2 The Batch Size

Batch size is an important parameter in the LSTM configures, which limits the number of samples to be shown to the network before a weight update can be performed. Thus batch size controls how often to update the weights of the LSTM network. This same limitation is then imposed when making predictions with the fit model.

In this section, we will explore the effect of varying the batch size. In this study the batch size used are 1,2,3,4,5,6,7.We will hold the number of training epochs constant at 1000.As with training epochs, we can objectively compare the performance of the network given different batch sizes. Each configuration was run 10 times and summary statistics calculated on the final results.

A box and whisker plot of the prediction MAPE and MSE were created to help graphically compare the distributions, shown in Figure 11. The green line shows the average performance while the box shows the variability of the performance of the LSTM with different batch size.

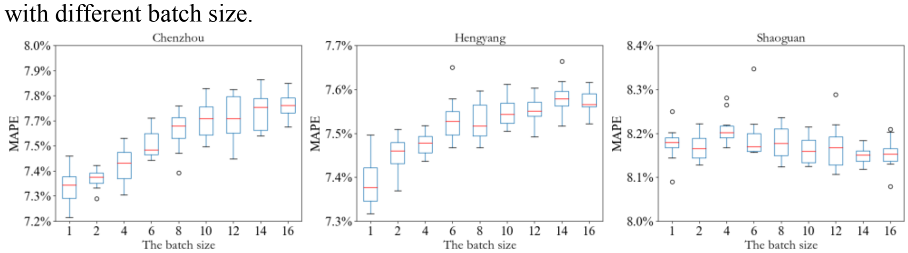

Figure 11: The distribution of prediction MAPE and RMSE of models with different batch size

In terms of the average performance, the median MAPE of Chenzhou station and Hengyang station showed an upward trend as the batch size increase from 1 to 16. The lowest low median MAPE are 7.36% and 7.39% as the batch size is 1, respectively for Chenzhou station and Hengyang station. For the Shaoguan station, the median MAPE fluctuates with the varying of the batch size, without obvious increase or decrease trend, indicating that the prediction accuracy of Shaoguan LSTM model for Shaoguan station is less affected by batch size.

The variability of the performance, the batch size has an influence on the stability of the LSTM model, since the variability of the box varies with the batch size. However the trend is not clear.

Tuning the batch size in a neural network is a tradeoff of average performance and variability of that performance. The ideal result should have a low mean error with low variability, meaning that it is generally good and reproducible. And the batch size should be decided according to the Data characteristics.

4.3 The Input Historical Data Length

The excellent performance of LSTM for short term traffic flow prediction mainly benefits from the memory ability of LSTM. For purpose of verifying the ability of LSTM to memorize long historical data, the performances of each model with different historical data length are compared. The input historical data length ranges from 7 to 35 with the interval of 7.Note that the input historical data length is always equal to the input size of each model. The five models' MAPE and RMSE are illustrated in Figure 12.

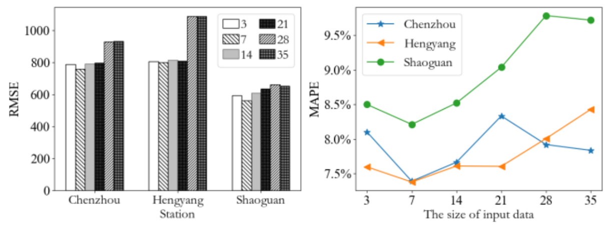

Figure 12: The distribution of prediction MAPE and RMSE of models with different input historical data length

There is a general trend of decreasing RMSE and MAPE as the number of historical data length increases from 3 to 7. As the historical data length increases from 3 to 7, the RMSE and MAPE for the passenger prediction at Chenzhou and Hengyang station rise up. For the passenger prediction at Shaoguan station, RMSE and MAPE increase as the historical data length increases from 7 to 21 and decrease as the historical data length increases from 21 to 35. The experiment results suggest a network configuration with historical data length of 7 having the best performance, the MAPE of which are 7.39%, 7.38%, 8.21%, Respectively for Chenzhou, Hengyang, Shaoguan station. It means that for one day prediction interval, the passenger flow in the past 7 days has a great impact on the current passenger flow, corresponding with the significant periodicity of 7 days presented by the passenger flow.

Specially, Diagnostic with1000 Epochs and various historical data length of Hengyang station are taken as an example to demonstrate the effect of the input of historical data length on the LSTM, and the result is shown in Figure 13.

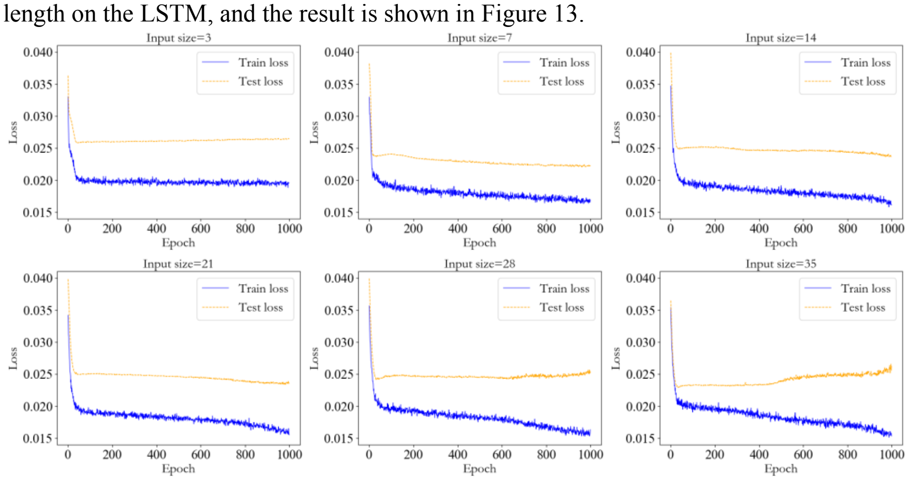

Figure 13: Diagnostic Results of models with different input historical data length

As the historical data length is 3, 7, 14, 21, both the lines of train loss and test loss show decrease trend and then keep horizontal, and the performance of these model seems to be reasonable. It shows a rapid decrease in test RMSE as the historical data length is 7 and the final RMSE score is 0.0222.

As the historical data length increases to 28 and 35, diagnostic results shows a rapid decrease in test RMSE to about epoch 500 and then rise up slightly until to the final epoch 1000. Meanwhile, the training dataset shows a continued decrease to the final epoch. There is a potential possibility of over fitting the training dataset.

To sum up, the performance of LSTM is effected by the input size of the data. LSTM can learning and memorize the complex interaction in the passenger time series and then predict the following passenger volume. For the passenger prediction in the experiment, the LSTM model is not well capable of getting the characteristics of the passenger time series as the input historical data length is too short. Meanwhile, as the input historical length is long, the limited learning ability cannot get the enough valid information contained in the data. In addition, and the longer the sequence data is, the more Interference noise information it contains, may lead a low prediction precision. Therefore, it is proper to model long-term dependencies and determine the optimal size of input data dynamically for the desirable results of short-term traffic flow prediction.

5 Conclusions

The paper analysis on the passenger flow characteristic of Wuhan-Guangzhou High Speed rail and proposes a passenger flow prediction method based on LSTM deep neural network. The results showed that:

(1) The LSTM passenger prediction model can cope with the correlation within long-term passenger time series and predict the trend of passenger flow accurately. The average prediction error MAPE of Chenzhou, Hengyang and Shaoguan stations are 7.36%, 7.33% and 8.03%, respectively. LSTM model is more effective and reliable than the other models, including RF, SVM, KNN, and GB models, while the stability of LSTM model is poor.

(2) The number of hidden units in each of hidden layer has a great influence on the prediction accuracy. While the number of hidden units is low, a slight increase of the hidden units of LSTM model can improve the convergence speed and prediction accuracy. As the number of hidden units in the LSTM model increase to a high level, the LSTM model may show an over-fitting state. In the experiment, the LSTM work show a better performance as the number of hidden units is set as 5 or 10.

(3) The input historical data length and the batch size have a great influence on the prediction accuracy of the LSTM model. When the historical data length is 7 and the batch size is 1, the passenger prediction accuracy is higher, which means the passenger flow in the past 7 days has a great influence on the following passenger flow.

Acknowledgement

This work was supported by the National Nature Science Foundation of China [grant number 71871188 and U1834209] and National Key R&D Program of China [grant number 2017YFB1200700]. We acknowledge the support of SAPIENZA Università di Roma and the China Scholarship Council. We are grateful for the contributions made by our project partners.

References

Asif, M.T., Dauwels, J., Goh, C.Y., et al., 2014. "Spatiotemporal patterns in large-scale traffic speed prediction". *IEEE Transactions on Intelligent Transportation Systems,* vol.15(2),pp.794–804.

Bengio, Y., Simard, P., Frasconi, P., 2002. "Learning long-term dependencies with gradient descent is difficult". *IEEE Transactions on Neural Networks,* vol.5(2),pp.157-166.

Çetiner, B.G., Sari, M., Borat, O., 2010. "A neural network based traffic flow prediction model". *Mathematical and Computational Applications*. vol.15 (2), pp.269-278.

Chandra, S.R., Al-Deek, H., 2009. "Prediction of freeway traffic speeds and volumes using vector autoregressive models". *Journal of Intelligent Transportation Systems*,vol.13 (2), pp.53-72.

Chen, H., Grant-Muller, S., 2001. "Use of sequential learning for short-term traffic flow forecasting". *Transportation Research Part C: Emerging Technologies,* vol.9(5), pp.319-336.

Chien, S.I.J., Liu, X., Ozbay, K., 2003. "Predicting travel times for the South Jersey real-time motorist information system". *Transportation Research Record,* vol.1855(1), pp.32-40.

Donahue, J., Hendricks, L A., Rohrbach M., et al. 2017. "Long-Term Recurrent Convolutional Networks for Visual Recognition and Description". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.39, pp.677-691.

Jiang, X., Zhang, L., and Chen, X.,2014. "Short-term forecasting of high-speed rail demand: a hybrid approach combining ensemble empirical mode decomposition and gray support vector machine with real-world applications in china". *Transportation Research Part C: Emerging Technologies,* vol. 44, pp.110-127.

Karlaftis, M., Vlahogianni, E., 2011. "Statistical methods versus neural networks in transportation research: differences, similarities and some insights". *Transportation Research Part C: Emerging Technologies,* vol. 19 (3),pp. 387–399.

Khashei, M , Bijari, M , Ardali, G A R ,2012. "Hybridization of autoregressive integrated moving average (ARIMA) with probabilistic neural networks (PNNs)". *Computers and Industrial Engineering,* vol.63(1),pp.37-45.

Kyungdoo, N., Thomas, S., 1995. "Forecasting international airline passenger traffic using neural networks". *Logist and Transportation Review,* vol.31(3),pp.239–251.

Li, L., Chen, X., Zhang, L., 2014. "Multimodel ensemble for freeway traffic state estimations". *IEEE Transactions on Intelligent Transportation Systems,*vol.15(3),pp.1323-1336.

Liu, H., Zuylen, H.V., Lint, H.V et al., 2006. "Predicting urban arterial travel time with state-space neural networks and Kalman filters". *Transportation Research Record,* vol.1968(1), pp.99-108.

Ma, X., Dai, Z., He, Z., et al., 2017. "Learning traffic as images: a deep convolutional neural network for large-scale transportation network speed prediction". *Sensors* vol.17(4),pp. 818.

Ma, X., Tao, Z., Wang, Y., et al. 2015. "Long short-term memory neural network for traffic speed prediction using remote microwave sensor data". *Research Part C: Emerging Technologies*, vol.54,pp. 187-197.

Moreira-Matias, L., Gama, J., Ferreira, M., et al.,2013. "Predicting taxi–passenger demand using streaming data". *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol.14, pp.1393–1402.

Polson, N G., Sokolov, V O., 2017."Deep learning for short-term traffic flow prediction". *Transportation Research Part C: Emerging Technologies,* vol.79,pp.1-17.
Smith, B.L., Williams, B.M., Oswald, R.K., 2002. "Comparison of parametric and nonparametric models for traffic flow forecasting". *Transportation Research Part C: Emerging Technologies,* vol.10(4),pp.303-321.
Srivastava, N., Mansimov, E., Salakhutdinov, R., 2015, "Unsupervised learning of video representations using LSTMs",In: *Proceedings of The International Conference on Machine Learning(ICML 2015),* Lille, France
Tan, M. , Wong, S. , Xu, J , et al, 2009. "An aggregation approach to short-term traffic flow prediction". IEEE Transactions on Intelligent Transportation Systems, vol.10,pp.60-69..
Tang, Y.F., Lam, W.H.K., Ng, P.L.P., 2003., "Comparison of four modeling techniques for short-term AADT forecasting in Hong Kong". *Journal of Transportation Engineering,* vol.129 (3),pp.271–277.
Tsai, T H., Lee, C K., Wei, C H., 2009, "Neural network based temporal feature models for short-term railway passenger demand forecasting". *Expert Systems with Applications,* vol.36(2), pp.3728-3736.
Van Lint, J.W.C., 2008. "Online learning solutions for freeway travel time prediction". *IEEE Transactions on Intelligent Transportation Systems,* vol.9 (1), pp.38–47.
Wang, Y., Papageorgiou, M., Messmer, A., 2007. "Real-time freeway traffic state estimation based on extended Kalman filter: a case study". *Transportation Science,*vol.41(2), pp.167–181.
Wang,Y.,Papageorgiou, M., Messmer, A., 2006. "RENASSANCE: a unified macroscopic model-based approach to real-time freeway network traffic surveillance". *Transportation Research Part C: Emerging Technologies,* vol. 14(3),pp. 190–212.
Wei,Y., Chen, M C., 2012."Forecasting the short-term metro passenger flow with empirical mode decomposition and neural networks". *Transportation Research Part C: Emerging Technologies,* vol. 21(1),pp.148-162.
Williams, B.M., 2001. "Multivariate vehicular traffic flow prediction: evaluation of ARIMAX modelling". *Transportation Research Record*,vol.1776, pp.194–200.
Williams, B.M., Hoel, L.A., 2003. "Modeling and forecasting vehicular traffic flow as a seasonal ARIMA process: a theoretical basis and empirical results". *Journal of transportation engineering,* vol.129 (6), pp.664-672.
Wu, C.H., Ho, J.M., Lee, D.T., 2004. "Travel-time prediction with support vector machine regression". *IEEE transactions on intelligent transportation systems*,vol. 5(4), pp.276-281.
Yu, H., Wu, Z., Wang, S., et al. 2017. "Spatiotemporal recurrent convolutional networks for traffic prediction in transportation networks". *Sensors,*vol.17(7), pp.1501.
Zhang, Y., Liu, Y., 2009. "Traffic forecasting using least squares support vector machines". Transportmetrica . vol.5 (3), pp.193–213
Zhang, Y., Xie, Y., 2007. "Forecasting of short-term freeway volume with v-support vector machines". *Transportation Research Record,* vol.2024(1), pp.92-99.
Zhu, Z., Sun, Y., Li, H., 2007. "Hybrid of EMD and SVMs for short-term load forecasting". In: *Proceedings of the IEEE International Conference on Control and Automation(ICCA 2007)*, Guangzhou, China.