WILEY | Hindawi

Research Article

ST-LSTM: A Deep Learning Approach Combined Spatio-Temporal Features for Short-Term Forecast in Rail Transit

Qicheng Tang, Mengning Yang [ID], and Ying Yang

School of Big Data & Software Engineering, Chongqing University, Chongqing, China

Correspondence should be addressed to Mengning Yang; mnyang@cqu.edu.cn

The short-term forecast of rail transit is one of the most essential issues in urban intelligent transportation system (ITS). Accurate forecast result can provide support for the forewarning of flow outburst and enables passengers to make an appropriate travel plan. Therefore, it is significant to develop a more accurate forecast model. Long short-term memory (LSTM) network has been proved to be effective on data with temporal features. However, it cannot process the correlation between time and space in rail transit. As a result, a novel forecast model combining spatio-temporal features based on LSTM network (ST-LSTM) is proposed. Different from other forecast methods, ST-LSTM network uses a new method to extract spatio-temporal features from the data and combines them together as the input. Compared with other conventional models, ST-LSTM network can achieve a better performance in experiments.

1. Introduction

With the development of urban scale, the short-term traffic forecast has become a core issue of ITS. Accurate short-term traffic forecast can provide technical support for the surveillance and the forewarning of passenger flow. Therefore, over the past few decades, many data analysis models have been proposed to promote the forecast accuracy. Among these models, LSTM network is widely recognized as the most suitable model to deal with traffic forecast. LSTM unit has three gates, namely, input gate, forget gate, and output gate, which can adjust the state of unit dynamically, so LSTM network is able to capture the features on longer time span. Therefore, LSTM network can provide a higher accuracy in traffic forecast because traffic data is usually collected according to time series.

In recent years, researchers pay more attention to the spatial features of traffic flow. It is widely acknowledged that traffic forecast is a problem with spatio-temporal complexity, i.e., the problem of spatial transportation in temporal dimension. In [1], Zheng Zhao et al. establish a network by connecting several LSTM units, which aimed to imitate the

structure of urban traffic. However, it failed to imitate the structure of large urban scale. Xiaobo Chen et al. proposed a new method to process spatial features by using sparse hybrid genetic algorithm [2]. Liu Qingchao et al. proposed a model based on manifold similarity to capture the spatial regularity from freeway data [3]. These two approaches are sensitive to the spatial features, but compared with LSTM network, they cannot process temporal features well.

In this paper, the object of study is the short-term forecast of rail transit. In the research, we find that differing from other transportation, rail transit has stations with fixed position, vehicles with uniform speed, and regular schedule. Because of these characteristics, the spatial correlation between stations can be transformed into the time cost. Based on this analysis, this paper proposes a new method to capture spatio-temporal features from rail transit data and input the features into a new model named spatio-temporal long short-term network (ST-LSTM), which is based on LSTM network. Compared with most existing methods, the proposed model has a better performance on accuracy and meets the real-time requirement.

2. Related Work

There are many methods that have been proposed to improve traffic forecast, including historical average and smoothing [4, 5], dynamic linear methods [6, 7], traffic theory-based methods [8, 9], and machine learning methods [10, 11]. These forecast approaches can be divided into two categories, namely, parametric approaches and nonparametric approaches. Autoregressive integrated moving average (ARIMA) model is widely recognized as a classic method in parametric approaches. As early as the 1970s, Levin and Tsao found that ARIMA model was the most statistically significant in traffic forecast [12]. Parametric approaches have favorable properties and capture regular variations very well. However, traffic data usually shows irregular variations. To solve this problem, researchers also paid attention to nonparametric approaches, such as nonparametric regression models [13], support vector machine (SVM) [14, 15], and recurrent neural network [16, 17]. Afterwards, recurrent neural network [18] was proposed to process temporal features, such as evolutionary neural network (ENN) [19], dynamic neural network (DNN) [20], and nonlinear autoregressive models with exogenous inputs (NARX) [21]. Among them, RNN is widely recognized as a suitable method to capture the temporal features of passenger flow. However, previous studies proved that RNNs failed to capture the long-term features because of vanishing gradient and exploding gradient. To solve these problems, long short-term memory neural network (LSTM NN) [22] was applied in the traffic forecast. In recent years, some approaches have been proposed to deal with the spatio-temporal complexity of traffic data, which are mentioned in Section 1.

Different from these methods, this paper proposes a new method to capture spatio-temporal features and a new network based on LSTM to forecast the exit passenger flow of rail transit. The remainder of this paper is as follows. Section 3 introduces the architecture of ST-LSTM network. Experiments based on the data of Chongqing rail transit are shown in Section 4. Section 5 is composed of the analysis of experiment result, and future work is at the end of this paper.

3. Methodology

Short-term forecast for rail transit is a problem with spatio-temporal complexity. Suppose the exit passenger flow of station j is needed to be predicted. The temporal features are the correlation between historical data and current data, i.e., the previous exit passenger flow of station j. These features can be extracted directly because the rail transit data is collected according to the temporal dimension. The spatial features are the transportation of passenger flow on geographic position; i.e., the summation of estimated passenger flows from the other stations. For every two stations, the spatial features include the volume and the cost of transportation between them. The volume of transportation is reflected by the passenger flow between two stations. In the proposed model, spatial correlation matrix (SCM) is integrated to calculate the volume of transportation. The cost of transportation is reflected by several factors, such as time cost, economic cost,

and distance. Among them, time cost is the most suitable factor to reflect the spatial correlation, which is mentioned in Section 1. Therefore, time cost matrix (TCM) is introduced to calculate the cost of transportation. The proposed model is based on the technologies, including passenger information system (PIS), features extraction method, and ST-LSTM network. The detail of the technologies will be explained in this section.

3.1. Passenger Information System. Sufficient data is the basis of accurate forecast, and PIS can provide us with comprehensive data. PIS is a huge and complex network. Various rail transit data can be collected in real time through the gate system, ticketing system, and vehicle scheduling system. With the development of data acquisition technology, PIS is able to provide sufficient support for short-term forecast. Based on the card records, the entrance and exit passenger flow of stations are calculated with a frequency of 10 min, which can be denoted by

$$i = \{cardid, otime, ostation, dtime, dstation, date\} \quad (1)$$

$$x\_in_{j,t} = \sum_{i \in M} \{1 \mid i.ostation = j, i.otime = t\} \quad (2)$$

$$x\_out_{j,t} = \sum_{i \in M} \{1 \mid i.dstation = j, i.dtime = t\} \quad (3)$$

where i is a card record. *cardid*, *otime*, *ostation*, *dtime*, *dstation*, and *date* are the attributes of i, which represent card identification, origin time, origin station, destination time, destination station, and date, respectively. M is the database of card records. $x\_in_{j,t}$ is the entrance passenger flow of station j in time t, and $x\_out_{j,t}$ is the exit passenger flow of station j in time t.

3.2. Feature Extraction Method. The extraction of spatio-temporal features is one of the core problems of the proposed model. The proposed model extracts temporal features and spatial features, respectively, and then put, them together into the ST-LSTM network. The temporal features can be extracted directly, because rail transit data is recorded according to the temporal dimension. To extract the spatial features, TCM matrix and SCM matrix are integrated into the method.

3.2.1. Time Cost Matrix. Time cost is the most suitable factor to reflect the spatial correlation between stations, so the time cost between all stations constitutes the TCM matrix. Due to the changes of schedule and passenger flow, TCM matrix is dynamic with time going on. Suppose there are m stations in the rail transit system; then the size of TCM matrix is $m \times m$, which can be denoted by

$$TCM_t = \begin{pmatrix} \Delta T_{1,1} & \Delta T_{1,2} & \cdots & \Delta T_{1,m} \\ \Delta T_{2,1} & \Delta T_{2,2} & \cdots & \Delta T_{2,m} \\ \cdots & \cdots & \Delta T_{k,j} & \cdots \\ \Delta T_{m,1} & \Delta T_{m,2} & \cdots & \Delta T_{m,m} \end{pmatrix} \quad (4)$$
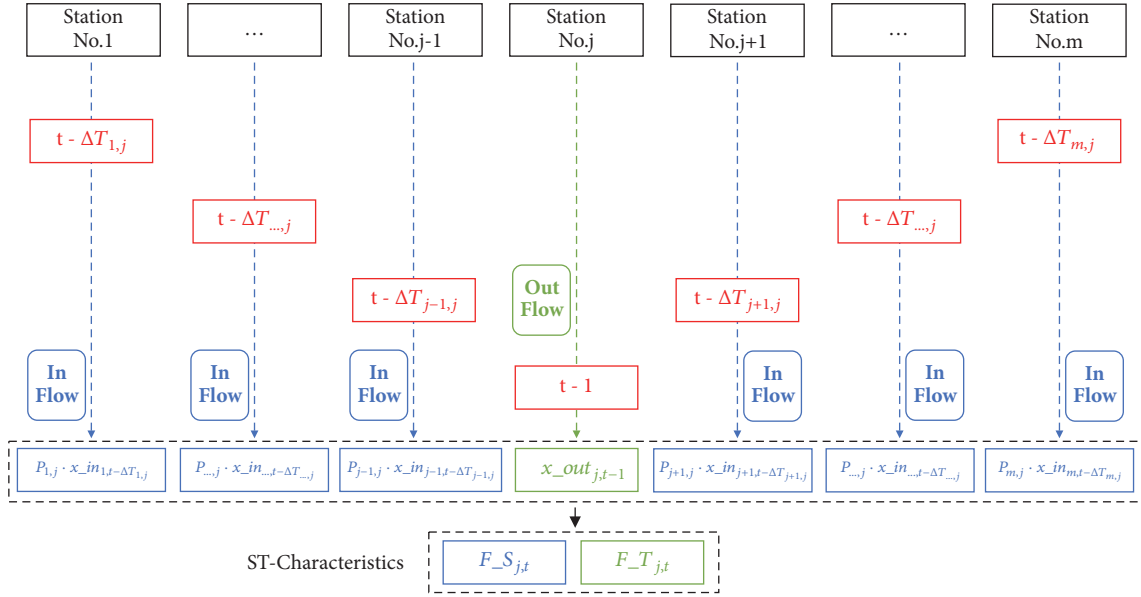
FIGURE 1: Design of the extraction of spatio-temporal features.

$$\Delta T_{k,j} = \overline{\Delta t_i}, \Delta t_i = i.dtime - i.otime$$

$$(i \in M, \ i.dtime \in T^*, \ i.ostation = k, \ i.dstation = j)$$
(5)

$$T^* = \{t + qw\}, \quad q = 0, 1, 2, \ldots, n$$
(6)

where TCM_t is the TCM matrix in time t. $\Delta T_{k,j}$ is the average of time cost between k and j in historical time series, where k is origin station and j is destination station. Δt_i is the time cost between two stations in record i. i is a card record in database M, which has been defined in Eq.(1). T^* is the historical time series of time t. w is a week and q is the number of weeks.

In the analysis, it is found that passenger flow varies by people's routine cycle, i.e., from Monday to Sunday. For example, the passenger flow in Thursday is similar to the one in last Thursday not yesterday. Therefore, to promote the extraction, $\Delta T_{k,j}$ in time t represents the average time cost in historical time series of time t. In Eq. (6), T^* is the historical time series, which consists of time t and the same period in several weeks ago of it. This method is also used in the calculation of spatial correlation matrix.

3.2.2. Spatial Correlation Matrix.

To forecast the exit passenger flow at station j in time t, passengers from station k in time $t - \Delta T_{k,j}$ have to be considered. The entrance passenger flow of station k in time $t - \Delta T_{k,j}$ ($x\_in_{k,t-\Delta T_{k,j}}$) is available. However, time t has not happened, so the proportion of passengers in $x\_in_{k,t-\Delta T_{k,j}}$, which set off to station j, is unavailable. To solve this contradiction, spatial factor is introduced in this paper. Spatial factor $P_{k,j}$ is the historical average probability of passengers between station k and j in entrance passenger flow. When forecasting the exit passenger flow at station j in time t, the spatial influence from station k can be calculated by multiplying the spatial factor $P_{k,j}$

and entrance passenger flow $x\_in_{k,t-\Delta T_{k,j}}$. The spatial factors between all stations constitute the spatial correlation matrix. There is an SCM matrix in each time, because the factors vary according to the time. Suppose there are m stations in the rail transit system; then the size of SCM matrix is $m \times m$, which is denoted by

$$SCM_t = \begin{pmatrix} P_{1,1} & P_{1,2} & \cdots & P_{1,m} \\ P_{2,1} & P_{2,2} & \cdots & P_{2,m} \\ \cdots & \cdots & P_{k,j} & \cdots \\ P_{m,1} & P_{m,2} & \cdots & P_{m,m} \end{pmatrix}$$
(7)

$$c_{k,j,t-\Delta T_{k,j}} = \sum_{i \in M} \{1 \mid i.otime = t - \Delta T_{k,j}, i.ostation$$

$$= k, i.dstation = j\}$$
(8)

$$P_{k,j} = \overline{\left(\frac{c_{k,j,t-\Delta T_{k,j}}}{x\_in_{k,t-\Delta T_{k,j}}}\right)}, \quad t \in T^*$$
(9)

$$T^* = \{t + qw\}, \quad q = 0, 1, 2, \ldots, n$$
(10)

where SCM_t is the SCM matrix of time t. $P_{k,j}$ is the spatial factor, where k is origin station and j is destination station. $\Delta T_{k,j}$ is the time cost from k to j. $c_{k,j,t-\Delta T_{k,j}}$ is the number of passengers from k to j, whose origin time is $t - \Delta T_{k,j}$. $x\_in_{k,t-\Delta T_{k,j}}$ is the entrance passenger flow of station k in time $t - \Delta T_{k,j}$. T^*, q, and w have been defined in Eq. (6).

3.2.3. Extraction of Spatio-Temporal Features.

The structure of extraction method is shown in Figure 1. To forecast the exit passenger flow at station j in time t, the temporal features is the exit passenger flow at station j in time $t - 1$. The spatial features are gathered by calculating the number of passengers,
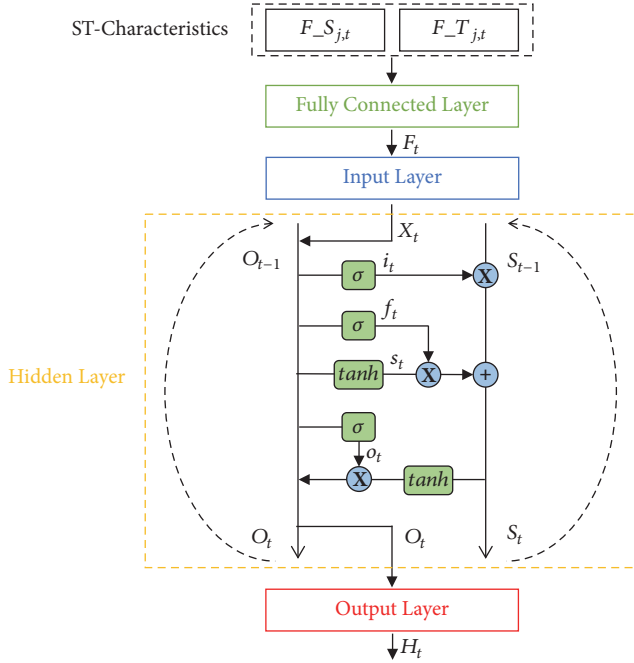
FIGURE 2: Structure of ST-LSTM network.

who will arrive at station j in time t and depart from other stations. The function of extraction can be set as

$$F\_T_{j,t} = x\_out_{j,t-1} \tag{11}$$

$$F\_S_{j,t} = \sum_{k \in N} P_{k,j} \bullet x\_in_{k,t-\Delta T_{k,j}} \tag{12}$$

where $F\_T_{j,t}$ is the temporal features of station j in time t. $x\_out_{j,t-1}$ is the exit passenger flow of station j in time $t - 1$. $F\_S_{j,t}$ is the spatial features of station j in time t. N is the set of stations and k is a station in it. $P_{k,j}$ is the spatial factor, which has been defined in Eq. (9). $x\_in_{k,t-\Delta T_{k,j}}$ is the entrance passenger flow of station k in time $t - \Delta T_{k,j}$.

3.3. Structure of ST-LSTM Network. Based on LSTM network, a fully connected layer is added to combine temporal features and spatial features in ST-LSTM network. The model will acquire the best mode of combination through the training.

The structure of ST-LSTM network is shown in Figure 2. The input of the model is spatio-temporal features $F\_T_{j,t}$ and $F\_S_{j,t}$, and the output is the forecast result H_t. There are four layers in this model, namely, fully connected layer, input layer, hidden layer, and output layer. The fully connected layer combines the features at first and conveys the result F_t to the input layer. The input of hidden layer X_t is calculated through the input layer. The hidden layer has three gates, namely, input gate i_t, forget gate f_t, and output gate o_t. Moreover, the state of the hidden layer is indicated by S_t. The inputs of every gate are X_t and the previous state S_{t-1}. The blue points in Figure 2 are confluences, which stand for multiplications, and dashed lines are the transmitting of the previous state. Based on the

information flow, the structure of ST-LSTM network can be summarized as

$$F_t = W_{fs}S_{j,t} + W_{ft}T_{j,t} + b_f \tag{13}$$

$$X_t = W_{hx}F_t + W_{hh}O_{t-1} + b_h \tag{14}$$

$$i_t = \sigma\left(W^{(i)}X_t + U^{(i)}S_{t-1}\right) \tag{15}$$

$$f_t = \sigma\left(W^{(f)}X_t + U^{(f)}S_{t-1}\right) \tag{16}$$

$$o_t = \sigma\left(W^{(o)}X_t + U^{(o)}S_{t-1}\right) \tag{17}$$

$$s_t = \tanh\left(W^{(c)}X_t + U^{(c)}S_{t-1}\right) \tag{18}$$

$$S_t = f_t \circ S_{t-1} + i_t \circ s_t \tag{19}$$

$$O_t = o_t \circ \tanh\left(S_t\right) \tag{20}$$

$$H_t = W_{oh}O_t + b_y \tag{21}$$

where F_t, X_t, O_t, and H_t are the output of different layers. i_t, f_t, o_t, and s_t are the intermediate variables of the hidden layer. S_t is the state of the hidden layer. W_{fs}, W_{ft}, W_{hx}, W_{hh}, W_{oh}, $W^{(i)}$, $W^{(f)}$, $W^{(o)}$, $W^{(c)}$, $U^{(i)}$, $U^{(f)}$, $U^{(o)}$, and $U^{(c)}$ are weight matrices. b_f, b_h, and b_y are bias vectors and σ is sigmoid function.

The cost function is activated after the forecast through the training. The proposed model is improved by reducing the output of cost function, which can be set as

$$loss = \left\|\hat{x}_{out,j,t} - x_{out,j,t}\right\|_2^2 \tag{22}$$

where $\hat{x}_{out,j,t}$ is the forecast of station j in time t and $x_{out,j,t}$ is the actual output.

3.4. Training Algorithm. The training algorithm contains two aspects. One is the extraction of spatio-temporal features, and the other is the training of ST-LSTM network. The key point of training is minimizing the output of cost function by adjusting the weight matrices and bias vectors. The training procedure can be stated as follows.

Step 1. Obtaining the Inputs and Labels. Capture the temporal features and the spatial features in each time t, which are the input of model. Collect the exit passenger flow in each time $t + 1$ as the labels.

Step 2. Initialization of the ST-LSTM Network. Initialize the weight matrices and bias vectors, including W_{fs}, W_{ft}, W_{hx}, W_{hh}, W_{oh}, $W^{(i)}$, $W^{(f)}$, $W^{(o)}$, $W^{(c)}$, $U^{(i)}$, $U^{(f)}$, $U^{(o)}$, $U^{(c)}$, b_f, b_h, and b_y.

Step 3. Fine-tuning the Whole Network. Fine-tune the whole network by adjusting the weight matrices and bias vectors in order to minimize the output of cost function. The process will be stopped until the output meets the qualification or the time of training reaches the limit.

TABLE 1: Performance of four models.

Model	ME	MAE	RMSE	MRE	Operation time
SARIMA	360.70	51.33	82.37	29.63%	75 min
PSO-SVR	153.45	18.89	27.48	21.46%	**8 min**
LSTM	183.59	22.82	32.29	26.45%	37 min
ST-LSTM	**145.72**	**18.42**	**25.89**	**20.77**%	38 min

4. Experiment

Based on the data of Chongqing rail transit, four models are contained in the experiment, namely Seasonal ARIMA (SARIMA) [23], Support Vector Regression Model combined with Particle Swarm Optimization (PSO-SVR) [24], LSTM network [25], and the proposed ST-LSTM network. The target of forecast is exit passenger flow with a frequency of 10 min. The four models will be trained and tested on 100 stations. The details of each model are as follows.

(1) SARIMA: The seasonal period 'S' is 100, due to the operating time being from 6.20 am to 23.00 pm (100×10 min). After the processing method in [23], ARIMA (2,1,0) $\times$ (0,1,1)100 is finally used.

(2) PSO-SVR: The time period is 100 (100×10 min per day), and the limit of parameter combination of SVR is from [10, 0.08] to [500, 0.3]. The final parameter combination will be selected by PSO in the training.

(3) LSTM network: The number of units is 10 and the time step is 100 (100×10 min per day).

(4) ST-LSTM network: The number of units is 10 and the time step is 100 (100×10 min per day).

4.1. Data Description. The data of card records are provided by Chongqing City Transportation Development & Investment Group Co., Ltd. Compared with other targets, such as Origin-Destination (OD) volume, exit passenger flow is more accurate and has less missing data. Therefore, we calculate the exit passenger flow from 01 March 2017 to 31 March 2017 based on the dataset. There are more than 46 million card records. After processing, 600 thousand data are calculated.

4.2. Evaluation. We use several criteria to compare the performance of four models. Maximum error (ME) and mean absolute error (MAE) are used to measure the accuracy of models. Root mean square error (RMSE) is sensitive to the stability of models. Mean relative error (MRE) is the most suitable to compare the performance of four models. The definitions of criteria are

$$\text{ME} = \max_{1 \le i \le n} \left\{ \left| \widehat{\varphi}_i - \varphi_i \right| \right\} \tag{23}$$

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^{n} \left| \widehat{\varphi}_i - \varphi_i \right| \tag{24}$$

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} \left(\widehat{\varphi}_i - \varphi_i \right)^2} \tag{25}$$

$$\text{MRE} = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{\widehat{\varphi}_i - \varphi_i}{\varphi_i} \right| \tag{26}$$

where $\widehat{\varphi}_i$ is the forecast data, while φ_i is the measured data.

4.3. Training and Testing. 5-fold cross validation is used to evaluate the models. In 5-fold cross validation, the data is divided into 5 subsets. Each subset is a testing set, and the rest of data is the training set. The experiments are repeated 5 times for each station. After the experiments, the performance of four models was collected. The experiments are conducted under a desktop computer with Intel i7 3.20 GHZ CPU and 16 GB memory.

4.4. Experiment Result. The experiment results of different algorithms are shown in Table 1 and the operation time is averaged on all stations. Compared with SARIMA, PSO-SVR, and LSTM network, the proposed ST-LSTM network achieved a better performance. From the view of ME and MAE, ST-LSTM network is more accurate than the other models. Moreover, from the view of RMSE, ST-LSTM network has a better stability. Therefore, the proposed ST-LSTM network is more suitable for the short-term forecast of rail transit.

5. Analysis of Result

When the models have been tested on 100 stations of Chongqing rail transit, we find that ST-LSTM network achieves a higher accuracy than the other models. However, the performance of ST-LSTM network fluctuates on different stations, which are shown in Table 2. Therefore, we analyze the experimental results based on the field investigation. The stations in Table 2 are sorted in descending order by passenger volume. Due to the lack of space, Table 2 just exhibits the performance on stations of top-10 and bottom-10 passenger volume.

5.1. Base Volume. In our research, we find that base volume is one of the influence factors of the forecast. The performance of two stations is chosen to shown in Figure 3, which are station No.321 and station No. 334 of Chongqing rail transit. Both of them are located in the residential district of Chongqing. However, station No. 334 only attains 5% of the base volume of station No.321 monthly. In the test, the MRE on station No.321 is 13.52%, while the MRE on station No.334 is 26.58%. We use these two different performances as samples to show the influence of base volume. The research suggests that the stations with higher base volume usually have more prominent regional features. As a result, passenger flows of these stations have stronger regularity and are more insensitive to the emergent factors. Therefore, short-term forecast on stations with low base volume is one of the difficulties in rail transit forecast.

TABLE 2: Forecast performance of ST-LSTM network on different stations.

| Station	ME	MAE	RMSE	MRE
321	366.96	71.08	99.12	13.52%
315	365.72	52.45	73.13	16.90%
318	280.70	32.45	47.31	11.13%
114	270.84	44.66	60.63	13.76%
322	215.52	30.41	41.12	12.10%
110	252.81	37.09	50.87	12.90%
606	238.99	33.44	48.46	11.97%
212	204.32	39.95	52.62	12.78%
108	215.70	32.76	47.10	13.94%
123	223.77	61.18	81.89	36.58%
. . .	. . .	. . .	. . .	. . .
335	69.81	6.58	10.10	29.34%
221	43.85	6.83	9.07	28.39%
305	58.25	9.55	12.93	32.25%
620	67.57	6.46	9.59	27.07%
621	45.15	6.69	8.91	30.47%
224	31.58	5.99	7.98	32.93%
219	36.12	5.40	7.40	31.86%
619	35.96	4.77	6.56	29.32%
118	23.63	4.41	5.72	29.88%
625	49.00	5.47	8.04	35.59%
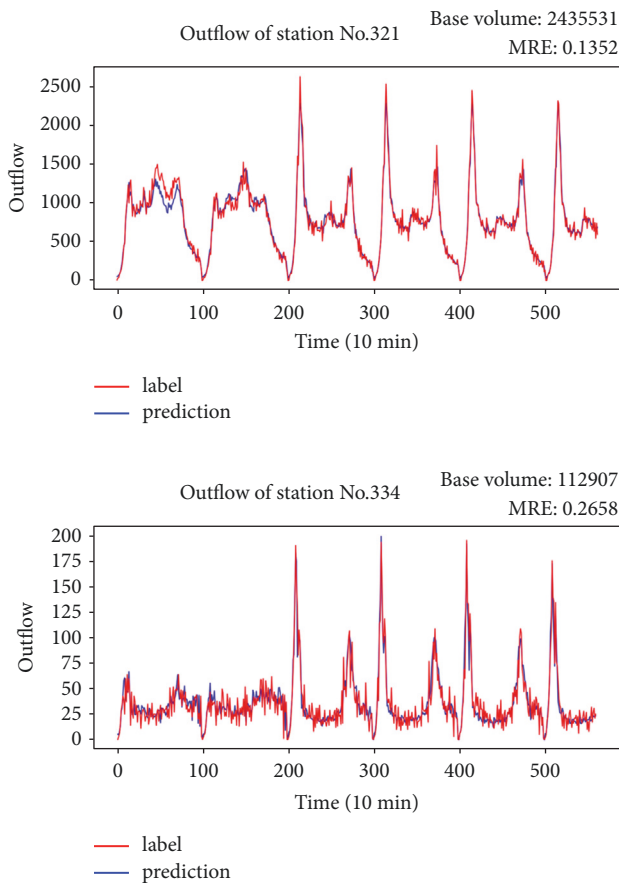
FIGURE 3: Influence of base volume in forecast.

5.2. Randomness. Except for the base volume, we discover that randomness is another influence factor of the forecast. As shown in Figure 4, the base volume of station No. 323 and station No.123 are both around 900 thousand monthly. However, the performance of forecast on two stations is quite different. In the test, the MRE on station No.323 is 16.73%, while the MRE on station No.123 is 38.37%. This phenomenon occurs on a few special stations, such as station No.123, which is located in the university town of Chongqing. Compared with the commuters, the undergraduates have more choice on the travel time. So the passenger flow of stations, which next to universities, has stronger randomness than others. Similarly, the passenger flow of stations, which next to railway stations or airports, is related to the flight schedule. Therefore, the randomness from the environment cannot be neglected on several stations. Short-term forecast on these stations is one of the difficulties in rail transit forecast.

6. Conclusion and Future Work

Short-term forecast for rail transit is an essential issue in ITS. We propose the ST-LSTM network, which combines the temporal features and spatial features. To extract spatial features, TCM matrix and SCM matrix are integrated into the method. Compared with other models, the proposed model is more suitable for rail transit forecast.

This study researches on prediction of exit passenger flow, but a model which also includes entrance passenger flow is more significant for the management. In addition, except the rail transit, ITS also contains bus system and taxi system. The

Base volume: 879544
MRE: 0.1673
Outflow of station No.323

Base volume: 899726
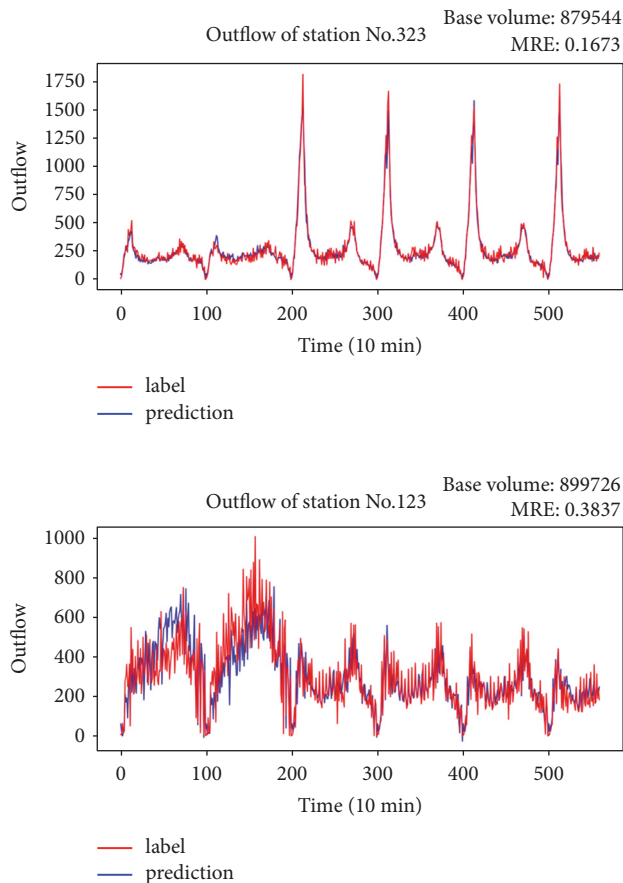MRE: 0.3837
Outflow of station No.123

Figure 4: Influence of randomness in forecast.

correlation between different public transportation is worth consideration. In the future, we will try to forecast other targets of rail transit and then consider the relation among different transportation. Finally, a comprehensive system for rail transit will be built to output a more accurate result of short-term forecast.

Data Availability

The data used in the experiments are freely available at https://drive.google.com/open?id=1RuH080U_9PHdh9B9VoOjNur-ODWnHAYaf and more data of Chongqing rail transit are available by contacting the corresponding author.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this article.

Acknowledgments

References

[1] Z. Zhao, W. Chen, X. Wu, P. C. Y. Chen, and J. Liu, "LSTM network: A deep learning approach for Short-term traffic forecast," *IET Intelligent Transport Systems*, vol. 11, no. 2, pp. 68–75, 2017.

[2] X. Chen, Z. Wei, X. Liu, Y. Cai, Z. Li, and F. Zhao, "Spatiotemporal variable and parameter selection using sparse hybrid genetic algorithm for traffic flow forecasting," *International Journal of Distributed Sensor Networks*, vol. 13, no. 6, pp. 1–14, 2017.

[3] Q. Liu, Y. Cai, H. Jiang, X. Chen, and J. Lu, "Traffic state spatial-temporal characteristic analysis and short-term forecasting based on manifold similarity," *IEEE Access*, vol. 6, pp. 9690–9702, 2017.

[4] K. Farokhi Sadabadi, M. Hamedi, and A. Haghani, "Evaluating moving average techniques in short-term travel time prediction using an AVI data set," in *Transportation Research Board 89th Annual Meeting*, vol. 2010.

[5] B. L. Smith and M. J. Demetsky, "Traffic flow forecasting: comparison of modeling approaches," *Journal of Transportation Engineering*, vol. 123, no. 4, pp. 261–266, 1997.

[6] W. Min and L. Wynter, "Real-time road traffic prediction with spatio-temporal correlations," *Transportation Research Part C: Emerging Technologies*, vol. 19, no. 4, pp. 606–616, 2011.

[7] X. Fei, C. C. Lu, and K. Liu, "A bayesian dynamic linear model approach for real-time short-term freeway travel time prediction," *Transportation Research Part C: Emerging Technologies*, vol. 19, no. 6, pp. 1306–1318, 2011.

[8] K. Nagel and M. Schreckenberg, "A cellular automaton model for freeway traffic," *Journal de Physique I*, vol. 2, no. 12, pp. 2221–2229, 1992.

[9] L. Li, X. Chen, Z. Li, and L. Zhang, "Freeway travel-time estimation based on temporal-spatial queueing model," *IEEE Transactions on Intelligent Transportation Systems*, vol. 14, no. 3, pp. 1536–1541, 2013.

[10] X. Zhang and J. A. Rice, "Short-term travel time prediction," *Transportation Research Part C: Emerging Technologies*, vol. 11, no. 3-4, pp. 187–210, 2003.

[11] Y. Wei and M. Chen, "Forecasting the short-term metro passenger flow with empirical mode decomposition and neural networks," *Transportation Research Part C: Emerging Technologies*, vol. 21, no. 1, pp. 148–162, 2012.

[12] M. Levin and Y. D. Tsao, "On forecasting freeway occupancies and volumes (abridgment)," *Transportation Research Record*, vol. 773, pp. 47–49, 1980.

[13] A. Rosenblad, "J. J. Faraway: Extending the linear model with R: generalized linear, mixed effects and nonparametric regression models," *Computational Statistics*, vol. 24, no. 2, pp. 369-370, 2009.

[14] Y. Zhang and Y. Liu, "Traffic forecasting using least squares support vector machines," *Transportmetrica*, vol. 5, no. 3, pp. 193–213, 2009.

[15] C.-H. Wu, J.-M. Ho, and D. T. Lee, "Travel-time prediction with support vector regression," *IEEE Transactions on Intelligent Transportation Systems*, vol. 125, no. 6, pp. 515–523, 2004.

[16] I. Okutani and Y. J. Stephanedes, "Dynamic prediction of traffic volume through Kalman filtering theory," *Transportation Research Part B: Methodological*, vol. 18, no. 1, pp. 1–11, 1984.

[17] H. Liu, H. van Zuylen, H. van Lint, and M. Salomons, "Predicting urban arterial travel time with state-space neural networks and Kalman filters," *Transportation Research Record*, no. 1968, pp. 99–108, 2006.

[18] P. Lingras, S. Sharma, and M. Zhong, "Prediction of recreational travel using genetically designed regression and time-delay neural network models," *Transportation Research Record*, vol. 13, no. 1, pp. 435–446, 2002.

[19] E. I. Vlahogianni, M. G. Karlaftis, and J. C. Golias, "Optimized and meta-optimized neural networks for short-term traffic flow prediction: a genetic approach," *Transportation Research Part C: Emerging Technologies*, vol. 13, no. 3, pp. 211–234, 2005.

[20] L. Shen, *Freeway Travel Time Estimation and Prediction using Dynamic Neural Networks (Ph. D. Dissertation)*, Florida International University, 2008.

[21] X. Zeng and Y. Zhang, "Development of recurrent neural network considering temporal-spatial input dynamics for freeway travel time modeling," *Computer-Aided Civil and Infrastructure Engineering*, vol. 28, no. 5, pp. 359–371, 2013.

[22] X. Ma, Z. Tao, Y. Wang, H. Yu, and Y. Wang, "Long short-term memory neural network for traffic speed prediction using remote microwave sensor data," *Transportation Research Part C: Emerging Technologies*, vol. 54, pp. 187–197, 2015.

[23] S. V. Kumar and L. Vanajakshi, "Short-term traffic flow prediction using seasonal ARIMA model with limited input data," *European Transport Research Review*, vol. 7, no. 3, 2015.

[24] W. Hu, L. Yan, K. Liu, and H. Wang, "A short-term traffic flow forecasting method based on the hybrid PSO-SVR," *Neural Processing Letters*, vol. 43, no. 1, pp. 155–172, 2016.

[25] C. Xu, J. Ji, and P. Liu, "The station-free sharing bike demand forecasting with a deep learning approach and large-scale datasets," *Transportation Research Part C: Emerging Technologies*, vol. 95, pp. 47–60, 2018.