IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received February 25, 2019, accepted March 22, 2019, date of publication March 27, 2019, date of current version April 13, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2907739

Short-Term Abnormal Passenger Flow Prediction Based on the Fusion of SVR and LSTM

JIANYUAN GUO[1], ZHEN XIE[1], YONG QIN[2], LIMIN JIA[2], AND YAGUAN WANG[1]
[1]School of Traffic and Transportation, Beijing Jiaotong University, Beijing 100044, China
[2]State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, Beijing 100044, China

Corresponding authors: Jianyuan Guo (jyguo@bjtu.edu.cn), Yong Qin (yqin@bjtu.edu.cn), and Limin Jia (lmjia@bjtu.edu.cn)

This work was supported in part by the National Key Research and Development Program of China under Grant 2016YFB1200402, and in part by the National Engineering Laboratory for System Safety and Operation Assurance of Urban Rail Transit.

ABSTRACT Passenger flow prediction is important for the operation of urban rail transit. The prediction of abnormal passenger flow is difficult due to rare similar history data. A model based on the fusion of support vector regression (SVR) and long short-term memory (LSTM) neural network is proposed. The inputs of the model are the abnormal features, which consist of the recent real volume series and the predicted volume series based on the periodic features. A two-stage training method is designed to train the LSTM model, which can reflect the large fluctuations of abnormal flow more timely and approximately. A combination method based on the real-time prediction errors is proposed, on which the outputs of SVR and LSTM are combined into the final outputs of the prediction model. The results of the experiments show that the SVR-LSTM model more accurately reflects the abnormal fluctuations of passenger flow, which performs well and yields greater forecast accuracy than the individual models.

INDEX TERMS Short-term passenger flow prediction, urban rail transit, support vector regression (SVR), long short-term memory (LSTM).

I. INTRODUCTION

Currently, urban rail transit systems play a key role in public transportation for large cities because they are rapid, punctual and green. Fluctuation of passenger flow has been a key factor for operation processes such as timetable optimization [1], train scheduling [2], train regulation and passenger flow control [3], [4]. Thus, obtaining the information about fluctuation of passenger flow is necessary for the efficiency of operating urban rail transit systems.

The fluctuation in passenger flow is periodic and random. In this paper, a significant random fluctuation in volume is called abnormal passenger flow. The abnormal passenger flow may be inflow or outflow, occur in work day or holiday, during peak hours or flat peak hours, even at midnight. In particular, an increasing abnormal passenger flow can cause the assembling of passengers and affect the urban rail transit system and the traffic connected to the subway station, thereby affecting the efficiency and safety of travelers. Accurate passenger flow prediction is helpful for timely and scientific train scheduling and passenger organization. With precise passenger flow prediction, a passenger information system can warn travelers about potential station and train congestion. Therefore, short-term abnormal passenger flow prediction is very important for operations in urban rail transit systems.

With the application of automated fare collection (AFC) systems, the records of passengers entering and exiting urban rail transit systems are collected, which provide rich data for passenger flow prediction. Based on the rich history data, some researchers have investigated short-term passenger flow prediction for subway stations. Wei and Chen [5] forecasted short-term metro passenger flow via empirical mode decomposition and back propagation (BP) neural networks. Yang and Hou [6], Zhou and Zhang [7], and Sun *et al.* [8] predicted rail transit passenger flow using wavelets and a least squares support vector machine (LS-SVM). Sun *et al.* [9] predicted the passenger flow of subway transfer stations based on a nonparametric regression model (k-Nearest Neighbors). Ding *et al.* [10] predicted short-term subway ridership using gradient boosting decision trees. Li *et al.* [11] forecasted short-term subway alighting passenger flow under special event scenarios using multiscale radial basis function networks. These studies contributed to short-term passenger flow prediction. However, short-term abnormal passenger flow prediction has received much less

The associate editor coordinating the review of this manuscript and approving it for publication was Jiahu Qin.

attention in the literatures. Among these researchers, only Li *et al.* [11] focused on abnormal passenger flow prediction and for passenger outflow merely.

For an urban rail transit system, abnormal passenger flow rarely occurs on a similar scale and in the same location in urban rail transit networks. More importantly, the development of abnormal passenger flow is very uncertain, which reduces the reference value of history data. Thus, short-term abnormal passenger flow prediction is difficult for rare similar samples and uncertain development.

The new method for utilizing the history data and the utility of real-time data should be strengthened to respond to the occurrence and development of the abnormal passenger flow. Therefore, the problem of estimating abnormal passenger flow is investigated in this paper.

According to the related studies and our practical experiments, the support vector regression (SVR) model is good at passenger flow prediction for its advantage in nonlinear regression based on similar history samples. Abnormal features, which consist of the recent real volume series and the predicted volume series based on periodic features, are used as the inputs of SVR, so that SVR can reflect the frequent fluctuations of passenger flow.

However, the SVR models can not reflect the abnormal flow for rare similar samples. Therefore, a two-stage training method is designed to train the long short-term memory (LSTM) neural networks to reflect the rarely large random fluctuations. The inputs of the LSTM model are also the abnormal features. The training at the first stage uses history data and the training at the second stage uses the recent samples which contain the real-time dada. The LSTM training method can utilize the real-time data sufficiently and reflect the abnormal fluctuation timely.

To take advantages of SVR and LSTM, the fusion model of SVR and LSTM are proposed in this paper. The outputs of SVR and LSTM are combined based on the real-time prediction errors as the outputs of the SVR-LSTM model. The SVR-LSTM model utilizes both the history data and the real-time data and predicts the short-term abnormal passenger flow more accurately than the individual models.

The remainder of this paper is structured as follows. The following section reviews the related literature. Then, the SVR-LSTM model is constructed, and experiments based on real data are conducted. Finally, conclusions are drawn in the last section.

II. RELATED LITERATURE REVIEW

A. METHOD OF PASSENGER FLOW PREDICTION

A wide range of statistical and machine learning prediction models have been applied to predict passenger flow. Moreover, studies on traffic flow prediction are helpful for studying passenger flow prediction. Therefore, we review both passenger flow and traffic flow prediction in this section.

Statistical methods have long been used to predict traffic flow. The typical statistical methods include the autoregressive integrated moving average (ARIMA) method [12], [13] and the B-spline method [14]. Machine learning methods have more recently been widely used to predict passenger flow and traffic variables. These methods include deep learning based on feed-forward neural networks for short-term traffic flow prediction [15], multiscale radial basis function networks for short-term subway passenger flow [11], deep belief networks for traffic matrix prediction [16], LSTM neural networks for traffic speed prediction [17], SVR for travel-time prediction [18], [19], LS-SVM for passenger flow prediction of transit rail stations [6]–[8], [20], adaptive multi-kernel SVM for short-term traffic flow prediction [21], nonparametric regression models for passenger flow prediction [9], [22], gradient boosting decision trees for ridership prediction [10], and hierarchical temporal memory (HTM) and LSTM for short-term arterial traffic flow prediction [23].

Researchers have recently begun to investigate strategies to combine predictors for passenger flow and traffic forecasting to increase accuracy. The use of a Kalman filter neural network was recommended to forecast short-term traffic flow for a medium-sized network [24]. Two novel neural network structures were proposed and integrated for short-term railway passenger demand forecasting [25]. An aggregation approach based on the moving average, exponential smoothing, ARIMA, and neural network models was proposed for traffic flow prediction [26]. Regression analysis, neural networks, and ARIMA models were used to predict transit ridership [27]. A hybrid short-term demand forecasting approach was developed by combining the ensemble empirical mode decomposition and gray SVM models [28]. Kalman filters were utilized to implement a stochastic seasonal ARIMA plus generalized autoregressive conditional heteroscedasticity structure for stochastic short-term traffic flow rate prediction [29]. A stacked auto-encoder model was employed to learn generic traffic flow features, and a logistic regression layer was applied for prediction [30]. Three single predictors, namely, ARIMA, Kalman filter and back propagation neural network, were designed and incorporated linearly to forecast short-term traffic flow [31]. A hybrid modeling approach that combines artificial neural networks and a simple statistical approach was used to forecast urban traffic flow [32]. A hybrid model of stacked auto-encoders and deep neural networks was applied and evaluated in a case study of passenger flow prediction for four bus rapid transit stations [33]. A fusion convolutional long short-term memory (LSTM) network was applied to the short-term forecasting of taxi passenger demand [34]. A combination of a convolutional neural network and recurrent neural network was used to mine spatial and temporal features [35]. Neural networks, SVR and random forests were selected as individual predictors, and a k-nearest neighbors fusion-based method was used for short-term traffic forecasting [36].

B. FEATURES CONSIDERED IN PREDICTION MODELS

Feature selection is important for ensuring accurate prediction. The various features used in passenger flow prediction models are reviewed in Table 1.

TABLE 1. Features Used for Passenger Flow Prediction.

Studies	Modes	Data interval	Features
Ke, Zheng [34]	Taxi	Hourly	Peak hours, off-peak hours, and sleep hours
Liu and Chen [33]	Bus	Hourly	Week, hour and holiday, the scenario type and direction, and the previous average value
Li, Wang [11]	Subway	15 minutes	The recent alighting passengers and recent boarding passengers in related stations
Kim and Shin [37]	Air	Monthly	The number of air passengers and the summed normalized absolute frequencies
Ding, Wang [10]	Subway	15 minutes	Time of day, day of month, day of week, recent adjacent bus stops alighting and boarding numbers
Sun, Leng [8]	Subway	15 minutes	Time series
Ma, Xing [38]	Bus	30 minutes	Real-time observation, weekly, daily and hourly relevant pattern time series
Jiang, Zhang [28]	Railway	Daily	Time series
Sun, Zhang [9]	Subway	6, 10 minutes	Trend of serial data
Yang [6]	Subway	5, 15, 60 minutes	Time series
Wei and Chen [5]	Subway	15 minutes	Day of the week, time period of the day, weekday or weekend
Chiang, Russell [27]	Bus	Daily	Food stamps, operating funds, route fares, monthly seasonality, time series
Tsai, Lee [25]	Railway	Daily	Trend, day of week, month of year, vacation

According to Table 1, the features of passenger flow prediction focus on periodic features (such as weekly, daily and hourly), time-series features, some statistical features (such as the previous average and the summed normalized absolute frequencies), and some features from real-time data (such as real-time observation, recent alighting passengers and recent boarding passengers).

III. MODELING

A. STRUCTURE OF THE SVR-LSTM MODEL

SVR is developed according to the basic SVM. The basic idea of SVM is to map the training data from the input space into a higher dimensional feature space via function and then construct a separating hyperplane with a maximum margin in the feature space. The idea of the regression problem is to determine a function that can accurately approximate future values. SVM and SVR have been extensively employed in prediction models [6]–[8], [18], [20], [21].

LSTM is an important type of recurrent neural network model, which was proposed by Hochreiter and Schmidhuber [39] and improved by Ger and Schmidhuber [40]. LSTM and its simple form (gated recurrent unit) have recently been applied in prediction models [17], [23], [34], [35].

The structure of the SVR-LSTM model proposed in this paper is shown in Figure 1. The periodic features are input into SVR (named SVR1) to compute a steady passenger flow volume series, which is referred to as the steady series. The recently observed real volume is used as the temporal series. The steady series and temporal series constitute the abnormal features that are input into SVR (named SVR2) and LSTM. The outputs of SVR2 and LSTM are $\hat{y}_1$ and $\hat{y}_2$, respectively. The combination of $\hat{y}_1$ and $\hat{y}_2$ is the final result of the SVR-LSTM model.

The training and prediction flow chart of the SVR-LSTM model is shown in Figure 2. $S_i(i \in \{1,2\})$is a sample set which contains the abnormal features. S_1 is used to train SVR2 and LSTM. S_2 is used to train LSTM and predict

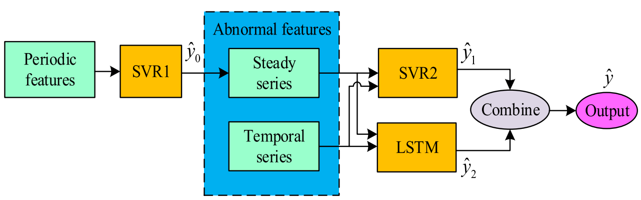


FIGURE 1. Framework of the SVR-LSTM model.

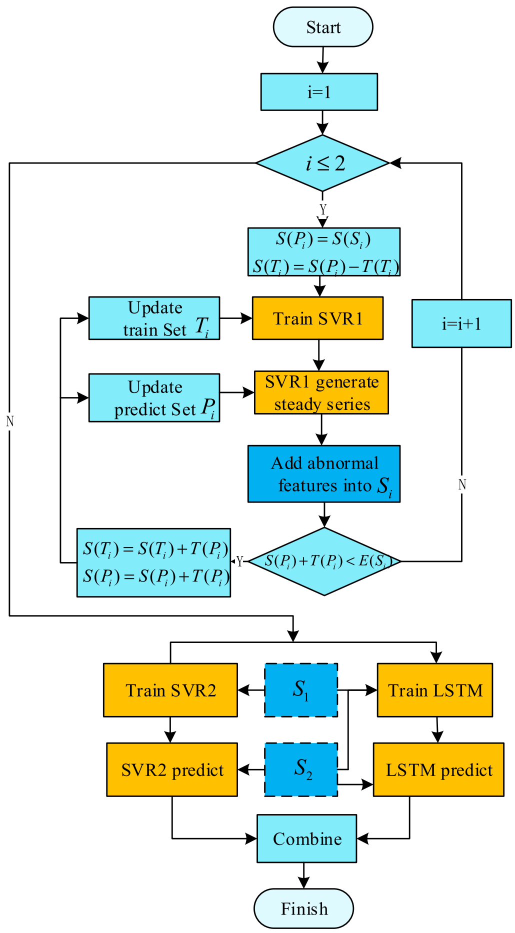


FIGURE 2. Training and prediction flow chart of the SVR-LSTM model.

based on SVR2 and LSTM. T_i is the sample set used to train SVR1 and P_i is the sample set used to predict and obtain the steady series for abnormal features in sample set S_i. $S(T_i)$, $S(P_i)$and $S(S_i)$ are start times of T_i,$P_i$and S_i respectively. $T(P_i)$ and $T(T_i)$are time spans of T_i and $P_i$respectively. $E(S_i)$ is the end time of S_i. According to Figure 2, SVR1 is trained based on T_i, generates the steady series based on P_i. Then the abnormal features containing the steady series and the temporal series are added to S_i. These operations are repeated until the samples for S_i are enough.

B. OBTAINING ABNORMAL FEATURES

The abnormal features are obtained by SVR1, as shown in Figure 1. Denote t as an element in the time series. $y(t)$ is the passenger flow volume at time t. The periodic features that are input into SVR1 are given by equation 1, where $h(t)$ indicates whether t is a holiday, $w(t)$ represents the day of the week, and $d(t)$represents the time of day.

$$[h(t), w(t), d(t)] \tag{1}$$

The output of SVR1 is given by equation 2, which is computed by the SVR model [18].

$$\hat{y}_0(t) = f(h(t), w(t), d(t)) \tag{2}$$

The abnormal feature at time t can be expressed by equation 3.

$$x(t) = [\hat{y}_0(t-L), \ldots, \hat{y}_0(t-offset), \hat{y}_0(t), y(t-L), \ldots y(t-offset)] \tag{3}$$

In equation 3, $\hat{y}_0$ is the output of SVR1, y is the observed real passenger flow volume, and *offset* is a coefficient greater than or equal to 1 that is set according to the prediction needs. L is a coefficient greater than the offset, which can be adjusted according to the experimental results.

C. LSTM

Referencing Ger and Schmidhuber [40], the structure of LSTM can be described as shown in Figure 3.

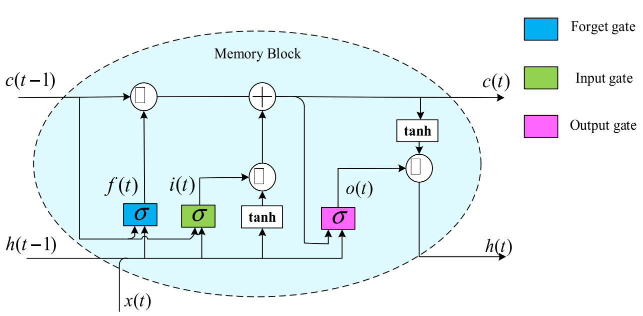


FIGURE 3. Structure of LSTM.

The LSTM prediction output can be computed by equations 4 to 11, where $W$and b are coefficients.

$$i(t) = \sigma(W_{ih}h(t-1) + W_{ix}x(t) + W_{ix}c(t-1) + b_i) \tag{4}$$

$$f(t) = \sigma(W_{fh}h(t-1) + W_{fx}x(t) + W_{fx}c(t-1) + b_f) \tag{5}$$

$$c(t) = f(t) \odot c(t-1) + i(t) \odot \tanh(W_{ch}h(t-1) + W_{cx}x(t) + b_c) \tag{6}$$

$$o(t) = \sigma(W_{oh}h(t-1) + W_{ox}x(t) + W_{oc}c(t) + b_o) \tag{7}$$

$$h(t) = o(t) \odot \tanh(c(t)) \tag{8}$$

$$\hat{y}_2(t) = W_{yh}h(t) + b_y \tag{9}$$

where $x(t)$ is the input of the model at time t; W are the weight matrices; b are the bias vectors; $i(t), f(t), o(t)$ are the activation functions of the input gate, forget gate and output gate at time t; $c(t)$ is the state of the memory cell at time t, $h(t)$ is the output of the memory block at time t; $\odot$ represents the scalar product of two vectors; $\sigma(x)$ is the standard logistics sigmoid expressed in equation 10; and tanh is the function expressed in equation 11.

$$\sigma(x) = \frac{1}{1+e^{-x}} \tag{10}$$

$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \tag{11}$$

The common objective function is to minimize the sum of square errors. The advantages of LSTM is that it uses gated neurons to capture both the short-term memories and the long-term memories and to avoid the gradient vanishing/exploding problem [35].

D. TWO-STAGE TRAINING METHOD FOR LSTM

According to the periodical fluctuation feature of passenger flow in transit networks, the samples are divided into sequences for different days. Denote the set of history samples that contain some sequences as S in equation 12, where n represents the number of days in S. As shown in equation 13, for sequence s_j, m samples exist, and s_{jk} contains $x(t)$ and $y(t)$.

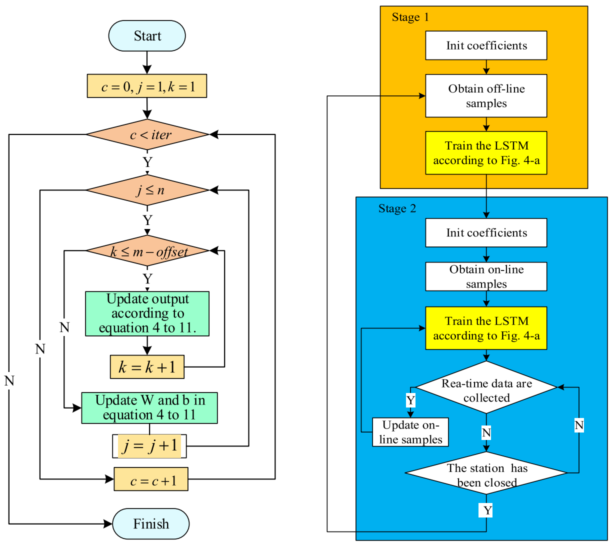
$$S = [s_1, \ldots, s_j, \ldots, s_n]^T \tag{12}$$

$$s_j = [s_{j1}, \ldots, s_{jk}, \ldots, s_{im}] \tag{13}$$

We divide all training samples into two types: off-line samples (S_1 in Figure 2) and on-line-samples (S_2 in Figure 2). The off-line samples do not contain samples from the current day, whereas the on-line samples include the most recent samples.

Based on the definition of sequence, off-line samples and on-line samples, the flow of the training LSTM is designed as shown in Figure 4.

The basic training flow is shown in Figure 4-a, and the two-stage training is shown in Figure 4-b. c and *iter* are the coefficients that control the iterations, and the definitions of m and n are same to those in equations 12 and 13. The real-time data are inserted into on-line samples when they are collected. The on-line samples will be inserted into off-line samples when the operation of a station stops and the station has been closed at midnight. Commonly, n in the second stage equals a small number, such as 1. After the training in the second stage, LSTM can be used to predict the short-term passenger flows.



(a) (b)

FIGURE 4. Training Flow of LSTM. (a) Basic training. (b) Two-stage training.

E. COMBINATION OF SVR AND LSTM

As shown in Figure 1, the final result of SVR-LSTM is the combination of the outputs of SVR2 and LSTM. The combination method is designed based on the real-time prediction errors.

The combination is computed by equation 14, where a is a coefficient ($0 \le a \le 1$), and $f(\hat{y}_1(t))$ is used to compute the degree of abnormal volume, where $f(\hat{y}_1(t))$ greater than zero indicates significant abnormal flow.

$$\hat{y}(t) = \begin{cases} \hat{y}_1(t), & f(\hat{y}_1(\mathrm{t})) < 0 \\ (1-a)\hat{y}_1(t) + a\hat{y}_2(t), & f(\hat{y}_1(\mathrm{t})) \ge 0 \end{cases} \tag{14}$$

$f(\hat{y}_1(t))$ can be computed by equation 15, where $e(\hat{y}_1(t))$ represents the degree of absolute error of SVR2, $g(\hat{y}_1(t))$ represents the degree of relative error of SVR2, and $\pi(\hat{y}_1(t))$ represents the trend of recent error of SVR2. ε,$\delta$and η are coefficients.

$$f(y_1(t)) = \min(e(\hat{y}_1(t)) - \varepsilon, g(\hat{y}_1(t)) - \delta, \pi(\hat{y}_1(t)) - \eta) \tag{15}$$

$e(\hat{y}_1(t))$, $g(\hat{y}_1(t))$ and $\pi(\hat{y}_1(t))$are computed by equations 16 to 18.

$$e(y_1(t)) = \sum_{i=offset}^{L} |y_1(t-i) - \widehat{y}_1(t-i)| \tag{16}$$

$$g(y_1(t)) = \sum_{i=offset}^{L} \frac{|y_1(t-i) - \widehat{y}_1(t-i)|}{y_1(t-i)} \tag{17}$$

$$\pi(y_1(t)) = \prod_{i=offset}^{L} (y_1(t-i) - \widehat{y}_1(t-i)) \tag{18}$$

a in equation 14 can be computed by equation 19.

$$a = \frac{g(\hat{y}_1)}{|g(\hat{y}_1)| + |g(\hat{y}_2)|} \tag{19}$$

IV. EXPERIMENTS

A. DATA DESCRIPTION

Real data from Yangji station in Guangzhou, China are used to assess the performance of the model in this paper. Yangji station is a transfer station where lines 1 and 5 cross. Here, Yangji1 and Yangji5 are used to represent the stations on lines 1 and 5, respectively. The data cover the inflow and outflow of passengers in 2017, which are counted by the records of the AFC system. The time step used for experiments is 15 minutes. There are about 73 samples for inflow and outflow respectively during an operation day (5:45-24:00) in a station.

The three cases shown in Table 2 are considered. The fluctuations of passenger flow including the three cases are shown in Figure 5, where the red lines are abnormal passenger flows.

The test samples of SVR2 and LSTM are the samples in these three cases. The details of training samples of SVR2 and LSTM are shown in Table 3. $T(P_i)$ and $T(T_i)$ in Figure 2 (the time spans of training set and prediction set of SVR1) are 10 days and 1 day respectively.

B. EXPERIMENTAL METRICS

The Mean Absolute Percentage Error (MAPE), the Root Mean Square Error (RMSE) and the Mean Absolute Error (MAE), are employed as the experimental metrics. They are defined as equations 20 to 22.

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\frac{\left|y_i - \widehat{y}_i\right|}{y_i} \times 100\% \tag{20}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}\left|y_i - \widehat{y}_i\right|^2}{n}} \tag{21}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|y_i - \widehat{y}_i\right| \tag{22}$$

The predictors are better when the values of the metrics are smaller. The MAPE is the most important metrics in our paper, which is supported by the equation 17. When there are inconsistent for the metrics, the predictor with the smaller MAPE is the better one. For example, if predictor A has smaller MAPE and bigger RMSE than predictor B, predictor A will be more optimal than predictor B.

C. PARAMETER CONFIGURATION

The parameters setting of the model in this paper are shown in Table 4.

L in abnormal feature, gamma and c in SVR and *iter* in LSTM are set based on the samples during January and February. ε in combination is set according to the value span of samples andδ in combination is set according to the prediction accuracy of SVR2 in normal conditions. Moreover, m and n in LSTM are set respectively according to the number

TABLE 2. Details of the Three Cases.

	Station	Date	Type of day	Type of flow	Feature of the flow
Case 1	Yangji5	June 2	work day	Outflow	flow increasing during morning peak hours
Case 2	Yangji5	August 31	work day	Outflow	flow increasing during flat peak hours
Case 3	Yangji1	December 31	holiday	Inflow	flow increasing during midnight

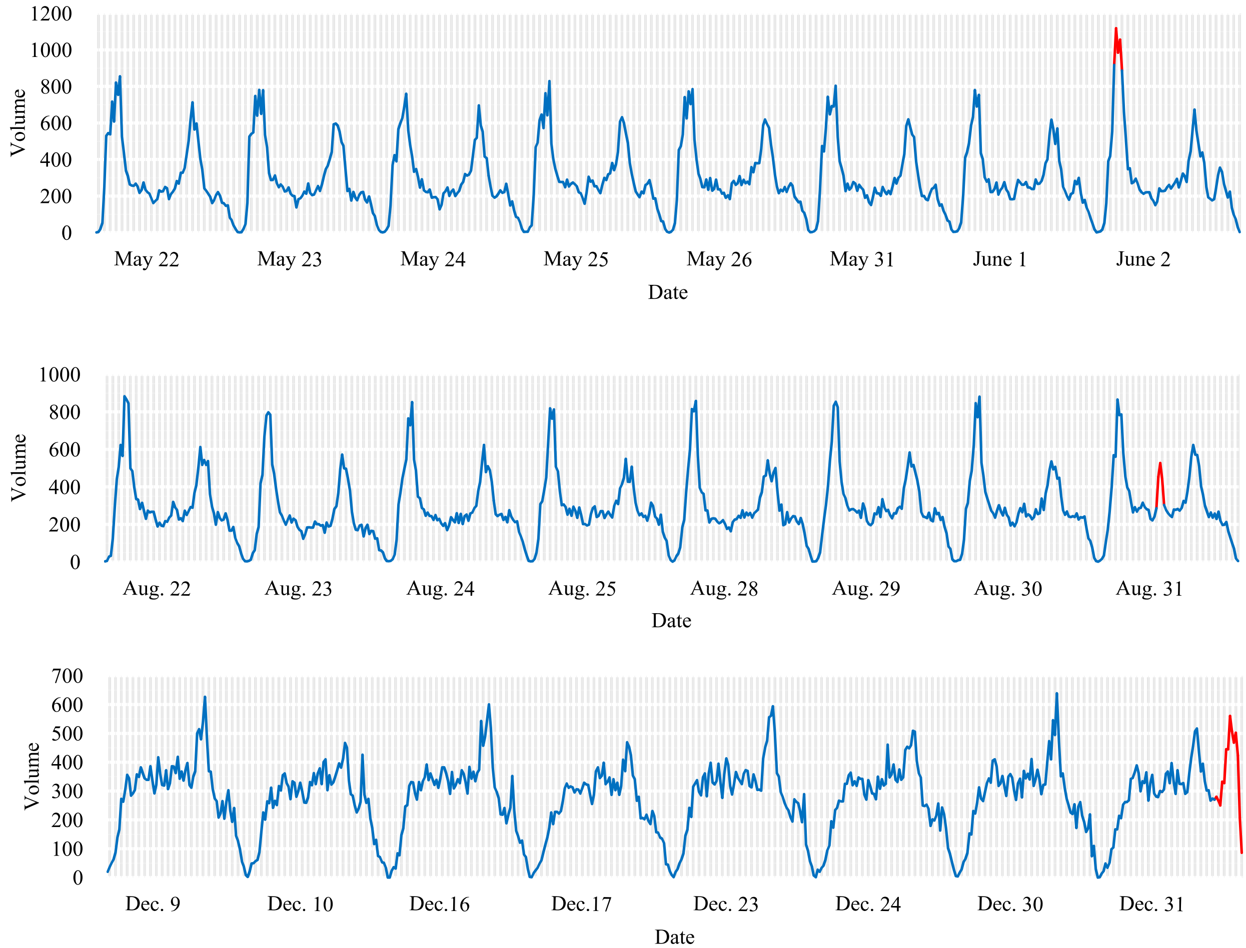


FIGURE 5. Fluctuations of passenger flow including the three cases.

TABLE 3. Details of the Training Samples for Three Cases.

	Samples for training SVR2	Samples for training stage 1 of LSTM	Samples for training stage 2 of LSTM
Case 1	Outflow in workday during March 1 to June 1	Outflow in workday during March 1 to June 1	Outflow in June 2
Case 2	Outflow in workday during March 1 to August 30	Outflow in workday during March 1 to August 30	Outflow in August 31
Case 3	Inflow in holiday during March 1 to December 30	Inflow in holiday during March 1 to December 30	Inflow in December 31

of days in sample sets and the number of samples in a day, which are different for three cases based on Table 3.

D. RESULTS

To validate the result of the model proposed in this paper, the experiments based on the typical statistic method of ARIMA was carried. Moreover, a model referencing [36] was realized and experimented, which fused the predictors of BP, SVR and RF based on KNN method (Fusion-KNN) and was good at short-term traffic prediction under abnormal conditions.

The results of SVR1, SVR2, LSTM, SVR-LSTM, ARIMA and Fusion-KNN are compared as shown in Table 5. One-step-ahead represents predicting the volume in the next 15 minutes (offset=1). Two-step-ahead represents predicting the volume during the next 15 to 30 minutes (offset=2).

TABLE 4. The Parameters Setting.

key parts in model	Value of coefficients
Abnormal feature	L =2
SVR1	gamma=0.01, c=50
SVR2	gamma=0.04, c=1
LSTM-stage 1	$iter$ =3000
LSTM-stage 2	$iter$ =1000
Combination	ε =100, δ =0.2/offset, η =0

The results show that the SVR-LSTM model proposed in this paper performs best in different cases.

The one-step-ahead prediction results of SVR1, SVR2, LSTM, SVR-LSTM, ARIMA and Fusion-KNN are compared with the real volume in Figure 6 to Figure 8.

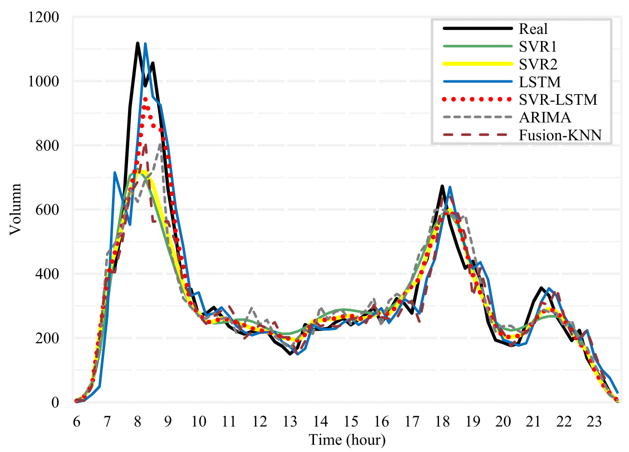


FIGURE 6. One-step-ahead prediction results of passenger flow in case 1.

According to Figures 6 to 8, the trend in SVR1 is the smoothest. SVR2 reflects more abnormal fluctuations than SVR1, whereas SVR2 does not respond to large random fluctuations, possibly because no similar fluctuation is present in the historic data sampleo0s.

LSTM reflect the abnormal fluctuations more timely and approximately than SVR. However, two defects of LSTM are significant: LSTM always excessively responds to slight fluctuations, and the error is large while the abnormality is disappearing. SVR-LSTM combines the advantages of SVR2 and LSTM and can reflect random fluctuations more accurately.

In addition, ARIMA can't reflect the abnormal fluctuation and has poorer performance on holiday. The model of Fusion-KNN can reflect the significant fluctuations, but is not better than LSTM in abnormal conditions in the three cases.

Furthermore, the absolute residuals of the SVR-LSTM model and real volume are shown in Figure 9 to Figure 11, where Absolute Residual1 indicates the absolute residuals of the one-step-ahead predictions and Absolute Residual2 represents the absolute residuals of the two-step-ahead predictions.

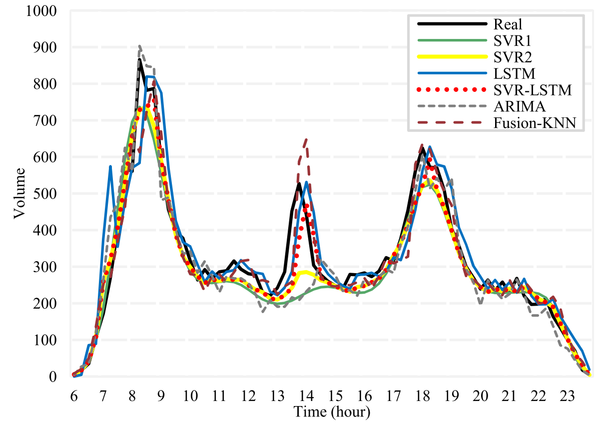


FIGURE 7. One-step-ahead prediction results of passenger flow in case 2.

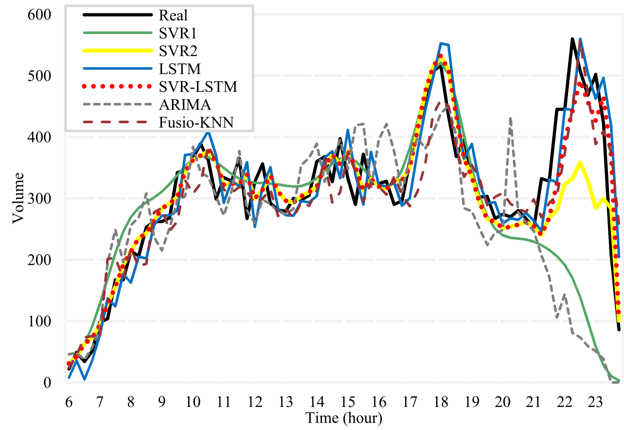


FIGURE 8. One-step-ahead prediction results of passenger flow in case 3.

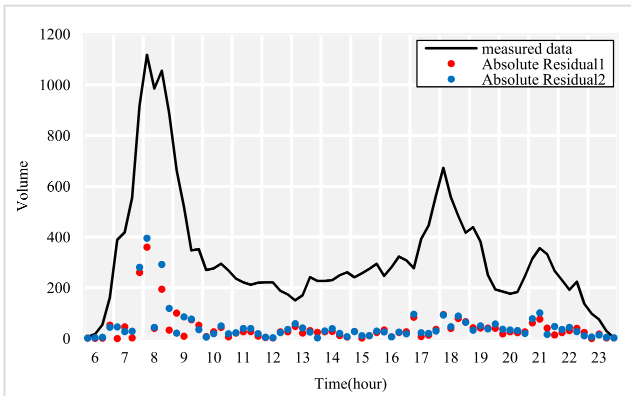


FIGURE 9. Absolute residuals in case 1.

According to Figure 9 to Figure 11, most of the absolute residuals are small in all three cases, and no significant fluctuation trend is observed for the absolute residuals.

E. DISCUSSION

According to the experiments, the SVR-LSTM model can capture the abnormal flow more timely and approximately. The SVR-LSTM model is more advantageous in abnormal passenger flow prediction than other algorithms compared.

TABLE 5. Comparison of Various Predictors.

		Case 1			Case 2			Case 3		
		MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE
One-step-ahead	SVR1	17.53	100.17	57.47	12.98	69.29	42.66	23.22	120.92	66.78
	SVR2	13.67	94.34	50.78	11.03	57.99	36.17	13.03	61.83	39.32
	LSTM	33.87	88.88	57.97	27.59	76.99	50.08	16.46	53.96	38.82
	SVR-LSTM	**12.59**	**68.54**	**40.06**	**10.45**	**49.51**	**32.58**	**11.53**	**43.09**	**30.48**
	ARIMA	16.52	97.29	54.10	15.13	69.35	43.05	25.80	139.84	81.13
	Fusion-KNN	19.83	105.71	58.89	15.11	59.67	37.50	18.03	56.86	42.50
Two-step-ahead	SVR1	17.53	100.17	57.47	12.98	69.29	42.66	23.22	120.92	66.78
	SVR2	15.83	101.68	57.14	**12.75**	64.25	40.34	16.31	85.78	49.76
	LSTM	31.20	**76.55**	55.04	23.39	77.86	51.81	16.87	57.46	42.16
	SVR-LSTM	**14.84**	80.12	**47.58**	**12.75**	**58.16**	**39.13**	**13.95**	**52.96**	**34.57**
	ARIMA	16.76	97.91	54.51	15.78	68.79	43.15	27.11	140.84	83.49
	Fusion-KNN	34.08	130.16	79.81	20.67	84.92	53.60	26.51	80.29	60.47

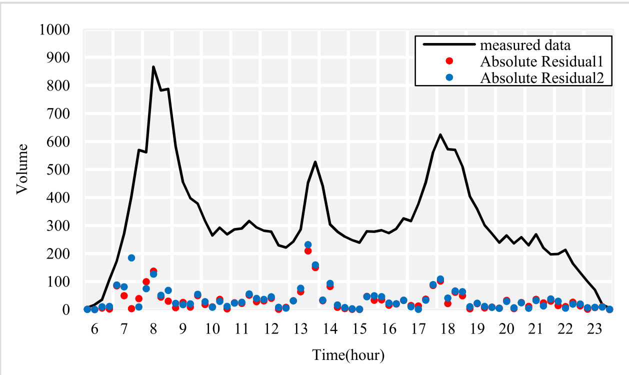


FIGURE 10. Absolute residuals in case 2.

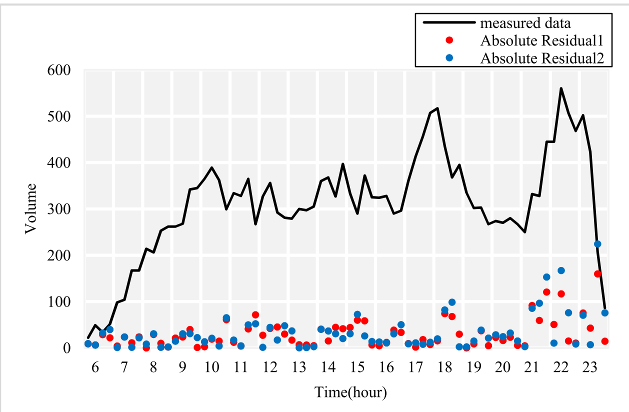


FIGURE 11. Absolute residuals in case 3.

However, it is should be noted that the abnormal features and combination conditions in SVR-LSTM model depend on the timely supply of real time data. The large transmission delay of real time data, such as more than 15 minutes, will not support the application of the model.

In addition, the prediction accuracy of SVR-LSTM at two-step-ahead is worse than that at one-step-ahead. It means the prediction accuracy of SVR-LSTM will decrease with the increase of value of offset. If the prediction time is more than half of an hour, the SVR-LSTM model proposed in this paper will be not applicable and the more information about normal condition should be obtained for prediction. Moreover, the traditional model, such as SVR, ARIMA and LSTM, may be better at prediction after half of an hour under normal condition than the SVR-LSTM model.

Furthermore, when the SVR-LSTM model are applied, the method of combination can be changed to adapt to different importance of metrics, and more optimal method of parameter configuration can be used to improve the prediction accuracy.

V. CONCLUSION

A model combining SVR and LSTM is proposed in this paper to predict abnormal passenger flow of stations in urban rail transit networks. The steady series are computed by SVR1, and the steady series and temporal series are input into SVR2 and LSTM as abnormal features. A two-stage training method is designed to train the LSTM model, which utilizes both the real-time samples and the history samples to reflect the large fluctuations of abnormal flow more timely and approximately. The combination of the outputs of SVR2 and LSTM is the final result of the model, which considers the real-time prediction errors. Real data from Yangji station in Guangzhou, China are used to assess the performance of the model. The experimental cases include work days and holidays; inflow and outflow; and passenger flow increases during peak hours, flat peak hours and midnight. The results show that the model proposed in this paper can accurately predict abnormal passenger flow based on the combination of SVR and LSTM, whose inputs are the abnormal features.

Additional factors, such as significant events, temporary organization measures, and the status of traffic connected to the urban rail transit station, will be considered in future studies.

REFERENCES

[1] H. Sun, J. Wu, H. Ma, X. Yang, and Z. Gao, "A bi-objective timetable optimization model for urban rail transit based on the time-dependent passenger volume," *IEEE Trans. Intell. Transp. Syst.*, vol. 20, no. 2, pp. 604–615, Feb. 2018. doi: 10.1109/TITS.2018.2818182.

[2] Y. Wang, T. Tang, B. Ning, T. J. J. van den Boom, and B. D. Schutter, "Passenger-demands-oriented train scheduling for an urban rail transit network," *Transp. Res. C, Emerg. Technol.*, vol. 60, pp. 1–23, 2015. doi: 10.1016/j.trc.2015.07.012.

[3] S. Li, M. M. Dessouky, L. Yang, and Z. Gao, "Joint optimal train regulation and passenger flow control strategy for high-frequency metro lines," *Transp. Res. B, Methodol.*, vol. 99, pp. 113–137, May 2017. doi: 10.1016/j.trb.2017.01.010.

[4] J. Shi, L. Yang, J. Yang, and Z. Gao, "Service-oriented train timetabling with collaborative passenger flow control on an oversaturated metro line: An integer linear optimization approach," *Transp. Res. B, Methodol.*, vol. 110, pp. 26–59, Apr. 2018. doi: 10.1016/j.trb.2018.02.003.

[5] Y. Wei and M.-C. Chen, "Forecasting the short-term metro passenger flow with empirical mode decomposition and neural networks," *Transp. Res. C, Emerg. Technol.*, vol. 21, no. 1, pp. 148–162, 2012. doi: 10.1016/j.trc.2011.06.009.

[6] J. Yang and Z. Hou, "A wavelet analysis based LS-SVM rail transit passenger flow prediction method," *China Railway Sci.*, vol. 34, no. 3, pp. 122–127, 2013.

[7] J. Z. Zhou and D. Y. Zhang, "Direct ridership forecast model of urban rail transit stations based on spatial weighted LS-SVM," *J. China Railway Soc.*, vol. 36, no. 1, pp. 1–7, 2014.

[8] Y. Sun, B. Leng, and W. Guan, "A novel wavelet-SVM short-time passenger flow prediction in Beijing subway system," *Neurocomputing*, vol. 166, pp. 109–121, Oct. 2015. doi: 10.1016/j.neucom.2015.03.085.

[9] Y. Sun, G. Zhang, and H. Yin, "Passenger flow prediction of subway transfer stations based on nonparametric regression model," *Discrete Dyn. Nature Soc.*, vol. 2014, Apr. 2014, Art. no. 397154. doi: 10.1155/2014/397154.

[10] C. Ding, D. Wang, X. Ma, and H. Li, "Predicting short-term subway ridership and prioritizing its influential factors using gradient boosting decision trees," *Sustainability*, vol. 8, no. 11, p. 1100, 2016. doi: 10.3390/su8111100.

[11] Y. Li, X. Wang, S. Sun, X. Ma, and G. Lu, "Forecasting short-term subway passenger flow under special events scenarios using multiscale radial basis function networks," *Transp. Res. C, Emerg. Technol.*, vol. 77, pp. 306–328, Apr. 2017. doi: 10.1016/j.trc.2017.02.005.

[12] C.-F. Chen, Y.-H. Chang, and Y.-W. Chang, "Seasonal ARIMA forecasting of inbound air travel arrivals to Taiwan," *Transportmetrica*, vol. 5, no. 2, pp. 125–140, 2009. doi: 10.1080/18128600802591210.

[13] M. M. Hamed, H. R. Al-Masaeid, and Z. M. B. Said, "Short-term prediction of traffic volume in urban arterials," *J. Transp. Eng.*, vol. 121, no. 3, pp. 249–254, 1995.

[14] F. Crawford, D. P. Watling, and R. D. Connors, "A statistical method for estimating predictable differences between daily traffic flow profiles," *Transp. Res. B, Methodol.*, vol. 95, pp. 196–213, Jan. 2017. doi: 10.1016/j.trb.2016.11.004.

[15] N. G. Polson and V. O. Sokolov, "Deep learning for short-term traffic flow prediction," *Transp. Res. C, Emerg. Technol.*, vol. 79, pp. 1–17, Jun. 2017. doi: 10.1016/j.trc.2017.02.024.

[16] L. Nie, D. Jiang, L. Guo, and S. Yu, "Traffic matrix prediction and estimation based on deep learning in large-scale IP backbone networks," *J. Netw. Comput. Appl.*, vol. 76, pp. 16–22, Dec. 2016. doi: 10.1016/j.jnca.2016.10.006.

[17] 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," *Transp. Res. C, Emerg. Technol.*, vol. 54, pp. 187–197, May 2015. doi: 10.1016/j.trc.2015.03.014.

[18] C.-H. Wu, J.-M. Ho, and D. T. Lee, "Travel-time prediction with support vector regression," *IEEE Trans. Intell. Transp. Syst.*, vol. 5, no. 4, pp. 276–281, Dec. 2004. doi: 10.1109/tits.2004.837813.

[19] Q. Chen, W. Li, and J. Zhao, "The use of LS-SVM for short-term passenger flow prediction," *Transport*, vol. 26, no. 1, pp. 5–10, 2011.

[20] S. Liu and E. Yao, "Holiday passenger flow forecasting based on the modified least-square support vector machine for the metro system," *J. Transp. Eng. A, Syst.*, vol. 143, no. 2, 2017, Art. no. 04016005. doi: 10.1061/jtepbs.0000010.

[21] X. Feng, X. Ling, H. Zheng, Z. Chen, and Y. Xu, "Adaptive multi-kernel SVM with spatial-temporal correlation for short-term traffic flow prediction," *IEEE Trans. Intell. Transp. Syst.*, to be published. doi: 10.1109/tits.2018.2854913.

[22] D. Xia, B. Wang, H. Li, Y. Li, and Z. Zhang, "A distributed spatial-temporal weighted model on MapReduce for short-term traffic flow forecasting," *Neurocomputing*, vol. 179, pp. 246–263, Feb. 2016. doi: 10.1016/j.neucom.2015.12.013.

[23] J. Mackenzie, J. F. Roddick, and R. Zito, "An evaluation of HTM and LSTM for short-term arterial traffic flow prediction," *IEEE Trans. Intell. Transp. Syst.*, to be published. doi: 10.1109/tits.2018.2843349.

[24] H. Chen and S. Grant-Muller, "Use of sequential learning for short-term traffic flow forecasting," *Transp. Res. C, Emerg. Technol.*, vol. 9, pp. 319–336, Oct. 2001.

[25] T.-H. Tsai, C.-K. Lee, and C.-H. Wei, "Neural network based temporal feature models for short-term railway passenger demand forecasting," *Expert Syst. Appl.*, vol. 36, no. 2, pp. 3728–3736, Mar. 2009. doi: 10.1016/j.eswa.2008.02.071.

[26] M.-C. Tan, S. C. Wong, J.-M. Xu, Z.-R. Guan, and P. Zhang, "An aggregation approach to short-term traffic flow prediction," *IEEE Trans. Intell. Transp. Syst.*, vol. 10, no. 1, pp. 60–69, Mar. 2009. doi: 10.1109/tits.2008.2011693.

[27] W.-C. Chiang, R. A. Russell, and T. L. Urban, "Forecasting ridership for a metropolitan transit authority," *Transp. Res. A, Policy Pract.*, vol. 45, no. 7, pp. 696–705, 2011. doi: 10.1016/j.tra.2011.04.016.

[28] X. Jiang, L. Zhang, and X. Chen, "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," *Transport. Res. C, Emerg. Technol.*, vol. 44, no. 4, pp. 110–127, 2014. doi: 10.1016/j.trc.2014.03.016.

[29] J. Guo, W. Huang, and B. M. Williams, "Adaptive Kalman filter approach for stochastic short-term traffic flow rate prediction and uncertainty quantification," *Transp. Res. C, Emerg. Technol.*, vol. 43, pp. 50–64, Jun. 2014. doi: 10.1016/j.trc.2014.02.006.

[30] Y. Lv, Y. Duan, W. Kang, Z. Li, and F.-Y. Wang, "Traffic flow prediction with big data: A deep learning approach," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 2, pp. 865–873, Apr. 2014. doi: 10.1109/tits.2014.2345663.

[31] J. Wang, W. Deng, and Y. Guo, "New Bayesian combination method for short-term traffic flow forecasting," *Transp. Res. C, Emerg. Technol.*, vol. 43, pp. 79–94, Jun. 2014. doi: 10.1016/j.trc.2014.02.005.

[32] F. Moretti, S. Pizzuti, S. Panzieri, and M. Annunziato, "Urban traffic flow forecasting through statistical and neural network bagging ensemble hybrid modeling," *Neurocomputing*, vol. 167, pp. 3–7, Nov. 2015. doi: 10.1016/j.neucom.2014.08.100.

[33] L. Liu and R.-C. Chen, "A novel passenger flow prediction model using deep learning methods," *Transp. Res. C, Emerg. Technol.*, vol. 84, pp. 74–91, Nov. 2017. doi: 10.1016/j.trc.2017.08.001.

[34] J. Ke, H. Zheng, H. Yang, and X. M. Chen, "Short-term forecasting of passenger demand under on-demand ride services: A spatio-temporal deep learning approach," *J. Transp. Res. C, Emerg. Technol.*, vol. 85, pp. 591–608, Dec. 2017. doi: 10.1016/j.trc.2017.10.016.

[35] Y. Wu, H. Tan, L. Qin, B. Ran, and Z. Jiang, "A hybrid deep learning based traffic flow prediction method and its understanding," *Transp. Res. C, Emerg. Technol.*, vol. 90, pp. 166–180, May 2018. doi: 10.1016/j.trc.2018.03.001.

[36] F. Guo, J. W. Polak, and R. Krishnan, "Predictor fusion for short-term traffic forecasting," *Transp. Res. C, Emerg. Technol.*, vol. 92, pp. 90–100, Jul. 2018. doi: 10.1016/j.trc.2018.04.025.

[37] S. Kim and D. H. Shin, "Forecasting short-term air passenger demand using big data from search engine queries," *Autom. Construct.*, vol. 70, pp. 98–108, Oct. 2016. doi: 10.1016/j.autcon.2016.06.009.

[38] Z. Ma, J. Xing, M. Mesbah, and L. Ferreira, "Predicting short-term bus passenger demand using a pattern hybrid approach," *Transp. Res. C, Emerg. Technol.*, vol. 39, pp. 148–163, Feb. 2014. doi: 10.1016/j.trc.2013.12.008.

[39] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997. doi: 10.1162/neco.1997.9.8.1735.

[40] F. A. Gers and J. Schmidhuber, "Recurrent nets that time and count," in *Proc. IEEE-INNS-ENNS Int. Joint Conf. Neural Netw.*, Jul. 2000, pp. 189–194.

JIANYUAN GUO received the Ph.D. degree in system engineering from Beijing Jiaotong University, China, in 2016, where she is currently an Associate Professor with the School of Traffic and Transportation. Her research interests include passenger flow prediction and organization in urban rail transit networks.

ZHEN XIE received the bachelor's degree in urban rail transit from Beijing Jiaotong University, China, in 2018, where she is currently pursuing the M.S. degree in planning and management of traffic and transportation with the School of Traffic and Transportation. Her research interests include passenger flow prediction and organization in urban rail transit networks.

YONG QIN received the Ph.D. degree from the China Academy of Railway Sciences, Beijing, China, in 1999. He is currently with the State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University. His current research interests include intelligent transportation systems, railway operation safety and reliability, rail network operation management, and traffic model.

LIMIN JIA received the Ph.D. degree from the China Academy of Railway Sciences, Beijing, China, in 1991. He is currently with the State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University. His current research interests include safety science and engineering, control science and engineering, and transportation engineering.

YAGUAN WANG received the B.S. degree from Shijiazhuang Tiedao University, Shijiazhuang, China, in 2016. She is currently pursuing the Ph.D. degree in safety science and engineering with the School of Traffic and Transportation, Beijing Jiaotong University, Beijing. Her research interests include passenger flow prediction and early warning.