Received July 26, 2019, accepted September 8, 2019, date of publication September 12, 2019, date of current version September 26, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2941177

Short-Term Origin-Destination Forecasting in Urban Rail Transit Based on Attraction Degree

JINLEI ZHANG[1], FENG CHEN[1,2], ZIJIA WANG[1], AND HANXIAO LIU[3]

[1]School of Civil Engineering, Beijing Jiaotong University, Beijing 100044, China
[2]School of Highway, Chang'an University, Xi'an 710064, China
[3]Beijing General Municipal Engineering Design and Research Institute Company Ltd., Beijing 100082, China

Corresponding author: Feng Chen (fengchen@bjtu.edu.cn)

This work was supported in part by the Fundamental Research Funds for the Central Universities under Grant 2018YJS128, and in part by the National Natural Science Foundation of China under Grant 71871027.

ABSTRACT Origin-destination (OD) forecasting is a difficult task in urban rail transit because many random factors can influence outcomes such as numerous OD pairs with few or even no flows. Therefore, treating all OD pairs equally, which is the method adopted by all existing studies, may not only increase model complexity and computation, but also negatively influence forecasting results. Therefore, in this study, we propose an indicator called OD attraction degree (ODAD) to address this problem in the field of OD forecasting. First, we introduce the ODAD indicator and five ODAD levels to describe the attraction between OD pairs. Based on the ODAD, an OD matrix pre-processing method is presented to prepare data for the LSTM model. Second, we use the mature long short-term memory (LSTM) network model to examine the effects of the introduction of ODAD. The LSTM model's advantage of dealing with variable-length sequences by leveraging the masking layer is creatively utilized. Finally, nine cases under different time granularities and different ODAD levels are thoroughly studied to explore their optimal combination. Based on this analysis, we recommend a time granularity of 30 min and an ODAD level of "Low" for actual subway operation. In this case, the root mean squared error, mean absolute error, and weighted mean absolute percentage error are 2.31%, 0.66%, and 27.28%, respectively, for a network of nearly 300 subway stations. The introduction of ODAD can provide critical insights for subway operators to conduct short-term OD forecasting.

INDEX TERMS Long short-term memory (LSTM) network, origin-destination attraction degree, short-term origin-destination forecasting, urban rail transit, deep learning.

I. INTRODUCTION

With the rapid development of urban rail transit (URT), additional features such as improved service levels, punctuality, and real-time passenger flow monitoring have been implemented. To better monitor the real-time spatiotemporal distribution of passenger flow, short-term passenger flow forecasting (STPFF) should be conducted. STPFF in URT includes the forecasting of *tap-in* (card swipe/scan) passenger flows, origin-destination (OD) passenger flows, and sectional passenger flows. Of these three types of passenger flows, research on tap-in STPFF has been relatively more advanced and has achieved higher prediction precisions [1]–[6]. However, studies about OD STPFF and sectional STPFF remain relatively few in existing literatures. Additionally, OD STPFF plays a key role as a primary connection between tap-in STPFF and sectional STPFF. Therefore, this study focuses primarily on OD STPFF (forecasting the destination of passengers), namely solving the problem of determining where passengers are traveling, under the condition of knowing the real-time tap-in volume, that is, knowing where they are at a particular time [7].

The automatic fare collection (AFC) system of URT can record the card number, tap-in time, tap-in station name, tap-out time, and tap-out station name of a trip. The real-time tap-in volume can be easily obtained because passengers must swipe a card when they enter a station. Owing to travel duration, however, the destination of a trip cannot be determined until passengers exit the station. Therefore, the OD STPFF, namely OD estimation, is introduced for

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

forecasting the OD distribution when the real-time tap-in volume is known. The OD forecasting can provide real-time passenger flow distribution status to support decisions on network management tasks, as well as implement flow congestion control and anomaly detection [8]. Moreover, this is essential input information for real-time route guidance and dynamic traffic assignment [9].

OD forecasting has been used in various forecasting domains. However, different models are utilized in different areas because of different prior knowledge. We will present existing studies from several areas, in which computer science, road traffic, and URT are three main application fields.

In the area of computer science, a traffic matrix primarily refers to Internet Protocol. Models used for Internet Protocol traffic matrix estimation include the tomography method [10], tomogravity method [11], principal component analysis method [12], and mathematical optimization method [13]. With the development of machine learning, some intelligent algorithms, such as back propagation neural networks (BPNN) and long short-term memory (LSTM) recurrent neural networks (RNN) were introduced. Jiang *et al.* [14] and Zhou *et al.* [15] proposed a BPNN model to estimate large-scale Internet Protocol traffic matrices. To satisfy the equality constraint, an iterative proportional fitting procedure and expectation maximization algorithm were applied to their studies to adjust their respective output of BPNN. Nie *et al.* [8] built a deep belief network and logistic regression model to conduct Internet Protocol traffic matrix prediction. Some researchers proposed a deep architecture based on an LSTM RNN [16, 17]. They transformed the n-order traffic matrix into an n^2 vector and used the vector as the input of the LSTM network. This type of network is able to capture the spatiotemporal characteristics of the traffic matrix. Because computer network traffic has self-similarity, long-time dependence, and a highly nonlinear nature [18], similar to road and URT traffic flow, this research can provide critical insights for OD forecasting in URT.

There are many studies on OD forecasting of road traffic. Typically, the actual OD matrix is unknown, whereas the traffic counts or freeway and ramp flows can be easily observed. Some scholars seek to fully utilize this kind of information for OD forecasting. Hitherto, entropy maximization, maximum likelihood, generalized least squares, Bayesian inference estimation techniques, state space model, and a bi-level programming approach have been introduced. Several typical studies are listed as follows. Combining the survey and traffic count data, Bell [19], [20] built a generalized least squares (GLS) model to conduct OD estimation that took travel time distribution into consideration. Yang *et al.* [21], [22] established a bi-level programming approach based on the least squares method and equilibrium traffic assignment using link counts data. They also introduced a concept of *maximum possible relative error* to evaluate the reliability of the estimated OD matrix [23]. Lin and Chang [9] presented a generalized state space model based on the measured freeway and ramp flows to dynamically estimate the freeway OD matrix; the Kalman filtering algorithm is applied to solve this model. To capture the drivers' speed difference, this study also innovatively introduced a travel time distribution function. Recently, some researchers have started using convolution-based methods to conduct OD forecasting [24, 25], but the effect was not verified in wide fields. Although road OD forecasting has been thoroughly discussed in existing studies, there is limited similarity between road and URT application because of differences in prior knowledge. However, some modelling ideas and methods (e.g., GLS and state space model) can be used for reference.

In contrast to road traffic OD forecasting, research in URT forecasting has been relatively limited. Additionally, while the sectional flows cannot be observed, the historical actual OD matrix, which is completely different from road traffic, can be accurately extracted. This makes it easier to use intelligent algorithms in machine learning. Some of this research is listed as follows. Zhao *et al.* [26] used entry-only AFC data to estimate trip destinations of individuals and to further estimate the OD matrix in the Chicago transit system. However, this is different from transit systems that operate with both entry-control and exit-control AFC systems, whose historical OD matrix can be directly observed. Using the historical OD matrix and travel time distribution, Yao *et al.* [27] built a state space model and then used the Kalman filtering algorithm to solve the model. Further, he introduced a GLS model based on a moving-average strategy using the average OD flows in several previous time periods [28]. Based on Yao's study, Chen *et al.* [29] considered commuter volume in the state space model. All these studies seek to explain the inherent relationships between OD flows from the perspective of mathematical optimization. However, there is few studies utilizing deep learning based methods in OD forecasting in URT.

Moreover, there are some other limitations in OD STPFF in URT. First, OD flows have extremely complicated correlations, so a single mathematical optimization method may not fully capture the spatiotemporal characteristics. Second, existing models can only be applied reliably to small-scale networks. With the network expansion in URT, these models become so complicated that they may not meet the requirement of ''real-time''. Moreover, the number of OD flows is much greater than network nodes, specifically, n^2 times, where n is the number of nodes. In contrast, many OD pairs may have very little flow or even no flow. Therefore, treating all OD pairs equally, which is the method adopted by all existing studies, may not only increase model complexity and computation, but also negatively influence forecasting results.

To compensate for these shortcomings, we have introduced an OD attraction degree (ODAD) indicator. To test the effect of introduction of ODAD indicator when conducting OD forecasting in URT, we used research in the computer science field [16], [17] as a reference on the use of LSTM RNN. The main reason we selected LSTM is that it is a relatively mature and stable method in traffic forecasting and has been

TABLE 1. Example of raw AFC data for March 9, 2016.

Index	Card number	Entering station number	Exiting station number	Entering time	Exiting time	Entering station name	Exiting station name
1	74873***	121	203	20160309190900	20160309193524	Yonganli	Fuchengmen
2	19727***	643	210	20160309123200	20160309124606	Chaoyangmen	Beijingzhan
3	42656***	9708	210	20160309115600	20160309124428	Tongzhoubeiyuan	Beijingzhan

TABLE 2. Example of processed AFC data for March 9, 2016.

Index	Card number	Entering station number	Exiting station number	Entering time	Exiting time
1	74873***	19	37	849	875
2	19727***	44	43	452	466
3	42656***	29	43	356	464

proven to be feasible for network traffic matrix prediction. First, we introduce an OD attraction degree (ODAD) indicator and ODAD levels to capture the importance of OD pairs in the entire URT network. Based on ODAD, some OD pairs that have lower OD flows (lower ODAD) are deleted to simplify the URT network, decrease model complexity and computation, and increase forecasting precision. Next, the OD matrix preprocessing method is presented in detail to prepare the data for the LSTM model. Then, the LSTM model is applied to verify the effect of ODAD when conducting OD forecasting. The essential ability of the LSTM model to operate on input sequences of variable lengths by leveraging the masking layer is creatively utilised. Finally, several cases under different time granularities (TGs) and ODAD levels are studied to evaluate the model performance and to explore the optimal combination of TG and ODAD levels.

The remaining sections of this study are organized as follows. In Section 2, we provide the OD matrix data extraction and preprocessing, as well as the introduction of ODAD. The LSTM model development is briefly described in Section 3. The case study and result discussion are presented in Section 4, and the conclusions are summarized in Section 5.

II. DATA PREPROCESSING AND INTRODUCTION OF ODAD

In this section, we first present automatic fare collection (AFC) data. The definition of ODAD is then described in detail. Finally, the OD matrix extraction procedure is introduced. Based on ODAD, some OD pairs are deleted, and the processed OD matrix is used as input to the LSTM model.

A. AFC DATA DESCRIPTION

The AFC data used in this study is taken from February 29, 2016 to April 3, 2016 in a Beijing subway. There were a total of 17 lines and 276 stations in operation in March 2016. Therefore, there are 76,176 OD pairs, which constitutes a very large network structure. To extract the OD matrix, every station is given only an index, and the entering and exiting times are transformed into minutes, from 0 to 1080, which represents 05:00 to 23:00 (18 h); this is done to simply divide the time period. Examples of raw and processed AFC data are shown in Table 1 and Table 2, respectively.

B. ODAD AND ODAD LEVEL

Existing studies have proven that OD pairs with low flow volumes can have a significant impact on the OD forecasting accuracy [30]. Within the URT, there are many OD pairs that attract only a few passengers; among these, most of the trips are randomly generated. To exclude randomness in the prediction process and thus improve accuracy, ODAD is introduced. ODAD is used to describe the attraction degree of OD pairs, namely the average OD flow volume in different time periods over a long duration. It can be calculated according to (1).

$$
\begin{aligned}
&ODAD_{ij}^{t} \\
&= \frac{1}{n}\sum\nolimits_{k=1}^{n} q_{ij}^{k,t} \\
&= \begin{pmatrix}
ODAD_{11}^{t} & ODAD_{12}^{t} & ODAD_{13}^{t} & \cdots & ODAD_{1j}^{t} \\
ODAD_{21}^{t} & ODAD_{22}^{t} & ODAD_{23}^{t} & \cdots & ODAD_{2j}^{t} \\
ODAD_{31}^{t} & ODAD_{32}^{t} & ODAD_{33}^{t} & \cdots & ODAD_{3j}^{t} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
ODAD_{i1}^{t} & ODAD_{i2}^{t} & ODAD_{i3}^{t} & \cdots & ODAD_{ij}^{t}
\end{pmatrix}
\end{aligned}
\tag{1}
$$

where i is the entering station number, and j is the exiting station number. The variable n is the number of days considered in a continuous time period, k is the particular day within this time period, t is the time period number during the day, and q is the flow volume. From this formula, we can infer that ODAD is a dynamic indicator. For the same OD pairs, the ODAD might be high during peak rush-hour periods and low in the middle of the night. Moreover, the ODAD is an average OD flow volume, thus excluding randomness.

In this study, i and j vary from 1 to 276, representing the 276 stations. n is 25, representing the 25 weekdays from

TABLE 3. ODAD levels adopted in this study (15 min).

ODAD value	Level
ODAD = 0	Lowest
0 < ODAD ≤ 1	Low
1 < ODAD ≤ 2	Middle
2 < ODAD ≤ 3	High
ODAD > 3	Highest

February 29, 2016 to April 3, 2016, and k thus varies from 1 to 25. If the TG used to aggregate the AFC data is 15 min, then t varies from 1 to 72 (18 h). If the TG is 30 min or 60 min, then t varies from 1 to 36 or from 1 to 18, respectively.

The ODAD level when TG is 15 min is given in Table 3, which can be used as a reference to pre-process the OD matrix. When TG is 30 min or 60 min, corresponding values are doubled or quadrupled, respectively.

C. OD MATRIX EXTRACTION AND PREPROCESSING

Based on the processed AFC data and ODAD levels, the historical OD matrix can be extracted in the form of Equation (2). The extracted OD matrix and ODAD matrix (Equation (1)) exhibit one-to-one correspondence. Given a certain ODAD level (e.g., ODAD level = "Low"), some OD pairs, namely those whose ODAD value is less than or equal to 1, will be deleted according to ODAD matrix. Equation (2) is then transformed into the form of Equation (3). Finally, because deleting OD pair results in variable-length sequences, OD pairs are synchronously moved to the left, and the vacant position was replaced by 0 to obtain the final OD matrix as shown in Equation (4), which is the input of the LSTM model.

$$OD_{ij}^{k,t} = \begin{pmatrix} q_{11}^{k,t} & q_{12}^{k,t} & q_{13}^{k,t} & \cdots & q_{1j}^{k,t} \\ q_{21}^{k,t} & q_{22}^{k,t} & q_{23}^{k,t} & \cdots & q_{2j}^{k,t} \\ q_{31}^{k,t} & q_{32}^{k,t} & q_{33}^{k,t} & \cdots & q_{3j}^{k,t} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ q_{i1}^{k,t} & q_{i2}^{k,t} & q_{i3}^{k,t} & \cdots & q_{ij}^{k,t} \end{pmatrix} = \begin{pmatrix} \boldsymbol{Q}_1^{k,t} \\ \boldsymbol{Q}_2^{k,t} \\ \boldsymbol{Q}_3^{k,t} \\ \vdots \\ \boldsymbol{Q}_i^{k,t} \end{pmatrix} \tag{2}$$

$$OD_{ij}^{\prime k,t} = \begin{pmatrix} q_{11}^{k,t} & \times & \times & \cdots & q_{1j}^{k,t} \\ q_{21}^{k,t} & q_{22}^{k,t} & q_{23}^{k,t} & \cdots & q_{2j}^{k,t} \\ q_{31}^{k,t} & \times & q_{33}^{k,t} & \cdots & q_{3j}^{k,t} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \times & q_{i2}^{k,t} & q_{i3}^{k,t} & \cdots & q_{ij}^{k,t} \end{pmatrix} \tag{3}$$

$$OD_{ij}^{\prime\prime k,t} = \begin{pmatrix} q_{11}^{k,t} & \cdots & q_{1j}^{k,t} & 0 & 0 \\ q_{21}^{k,t} & q_{22}^{k,t} & q_{23}^{k,t} & \cdots & q_{2j}^{k,t} \\ q_{31}^{k,t} & q_{33}^{k,t} & \cdots & q_{3j}^{k,t} & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ q_{i2}^{k,t} & q_{i3}^{k,t} & \cdots & q_{ij}^{k,t} & 0 \end{pmatrix} = \begin{pmatrix} \boldsymbol{Q}_1^{\prime k,t} \\ \boldsymbol{Q}_2^{\prime k,t} \\ \boldsymbol{Q}_3^{\prime k,t} \\ \vdots \\ \boldsymbol{Q}_i^{\prime k,t} \end{pmatrix} \tag{4}$$

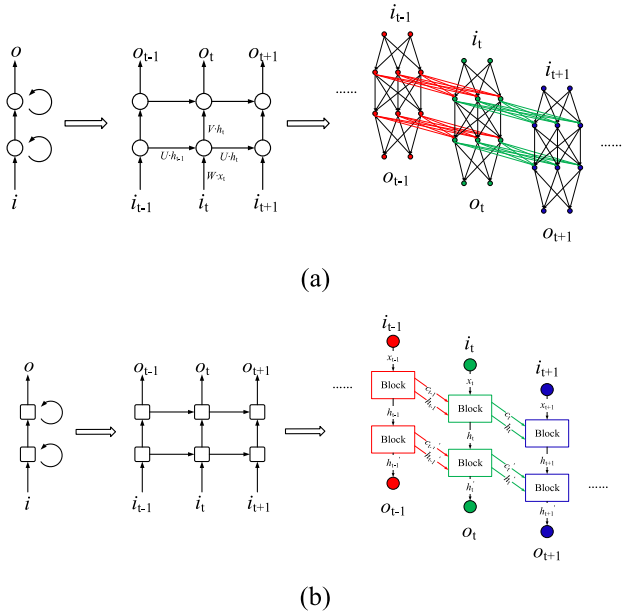

FIGURE 1. Diagrams of (a) conventional RNN and (b) LSTM model architecture.

where $OD_{ij}^{k,t}$ represents the OD matrix in the time period of t during the day k, and $q_{ij}^{k,t}$ is the OD flow entering station i in the time period of t and leaving for station j. $\boldsymbol{Q}_{ij}^{k,t}$ is a row vector.

The LSTM model has the advantage of automatically handling variable-length sequences by leveraging the masking layer. In other words, it can automatically filter "0" used as position completion during model training. This feature is also one of the important features to verify whether the introduction of ODAD can improve prediction accuracy. The form of $OD''^{k,t}_{ij}$ in Equation (4) is absolutely the input of LSTM model. It should be noted that the location of OD pairs in the output is the same as the input. Therefore, after determining the output, the location of OD pairs will be relocated according to their original place.

III. MODEL DESCRIPTION

LSTM model has been proved to be effective, mature and stable in traffic forecast tasks. It can avoid vanishing or exploding gradient problems, thus can capture long-term temporal dependencies of time series and achieve satisfactory results. Unlike conventional neurons in an RNN (Figure 1 (a)), the memory block (Figure 2) in an LSTM network (Figure 1 (b)) is composed of four parts: input gate, output gate, forget gate, and memory cell [30]. The three gates can determine what can be input, output, and forgotten during training process. The memory cell is closely related to the three gates and can record and pass historical useful information into the present task. The data flow can be calculated according to Equations (5) to (12) [31]–[33].

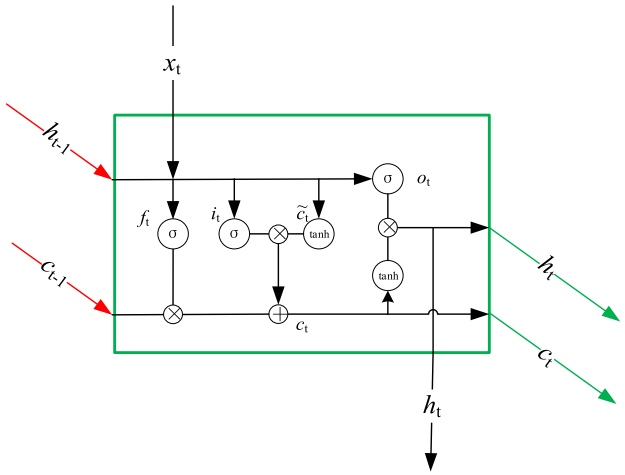

FIGURE 2. Diagram of a LSTM model memory block.

$$f_t = \sigma\left(W_f x_t + U_f h_{t-1} + V_f c_{t-1} + b_f\right) \quad (5)$$

$$i_t = \sigma\left(W_i x_t + U_i h_{t-1} + V_i c_{t-1} + b_i\right) \quad (6)$$

$$\tilde{c}_t = tanh\left(W_c x_t + U_c h_{t-1} + b_c\right) \quad (7)$$

$$c_t = f_t \odot c_{t-1} + i_t \odot \tilde{c}_t \quad (8)$$

$$o_t = \sigma\left(W_o x_t + U_o h_{t-1} + V_o c_t + b_o\right) \quad (9)$$

$$h_t = o_t \odot tanh\left(c_t\right) \quad (10)$$

$$\sigma\left(x\right) = \frac{1}{1+e^{-x}} \quad (11)$$

$$tanh\left(x\right) = \frac{e^{x}-e^{-x}}{e^{x}+e^{-x}} \quad (12)$$

where x_t, i_t, o_t, f_t, c_t, and h_t represent input data, input gate, output gate, forget gate, cell state, and final output, respectively. The variables W, U, and V are weight matrixes, and b is the bias vector. The back propagation through time (BPTT) algorithm and "rmsprop" optimizer are utilized to optimize these parameters. The t is the time step, and $\odot$ stands for Hadamard product. The initial values of c_t and h_t are all set as 0.

To estimate the OD matrix in the time period t of day k, we use the OD matrix in time periods t and $t-1$ of day $k-1$, combined with the entering and exiting passenger volume as the input of the LSTM model, as shown in Equations (13) and (14). All these input data are real rather than forecasted to avoid accumulated errors. Unlike Zhao *et al.* [16] and Azzouni and Pujolle [17], we forecast the OD matrix using a row of the matrix, namely a subway station, rather than making the matrix into a vector; this method is more refined and focused. After obtaining forecasting results of all stations in the specific time period t, the forecasting results will be integrated into an OD matrix, which is the final result of the LSTM model. Because there are 276 stations, parallel computing is used to train the LSTM network for different stations.

$$\boldsymbol{Q}_i'^{k,t} = f(\boldsymbol{Q}_i'^{k-1,t}, \boldsymbol{Q}_i'^{k-1,t-1}, I_i^{k,t}, O_i^{k,t}) \quad (13)$$

$$\boldsymbol{Q}_i'^{k,t} = (\, q_{i1}'^{k,t} \quad q_{i2}'^{k,t} \quad q_{i3}'^{k,t} \quad \cdots \quad q'^{k,t}{}_{ij} \quad (14)$$

where $\boldsymbol{Q}_i'^{k,t}$, $\boldsymbol{Q}_i'^{k-1,t}$, and $\boldsymbol{Q}_i'^{k-1,t-1}$ are row vectors (as shown in Equation (5)) representing OD flows entering station i in the specific time period t of day k, t of day $k-1$, and $t-1$ of day $k-1$, respectively. The $q_{i1}'^{k,t}$ is the OD flow after OD relocating and padded with zeros. The $I_i^{k,t}$ and $O_i^{k,t}$ are the entering and exiting passenger volumes, respectively, in the time period t of day k.

IV. CASE STUDY AND RESULT DISCUSSION

In this section, the detailed input data and output data are first described. Then, several evaluation metrics are chosen to evaluate model performances. Finally, cases for different TGs are studied and the OD estimate result is presented.

A. MODEL CONFIGURATION

The training dataset comprises the former 22 workdays, and the testing dataset is the 23rd workday (Wednesday). In actual subway operations, the prediction time (for which TG is adopted) should be seriously considered. A smaller TG implies more detailed passenger flow information will be captured, but the predictability will be worse; a larger TG is the opposite. In this study, estimation of the OD matrix with three TG settings will be discussed (15 min, 30 min, and 60 min) to find the appropriate TG used for actual operation. For each TG, three ODAD levels ("Lowest", "Low", and "Middle") are applied to evaluate model performance, thus giving a total of 9 studied cases.

We then use Keras to build the LSTM model. By trial and error, the architecture includes a masking layer, two LSTM layers, a dense layer with linear activation, as shown in Figure 3. Because the four types of input data have different sequence lengths, they will be padded into the same length of 276 with zeros. Therefore, in every time step, there are 276 features. Fortunately, the masking layer before the LSTM layer has the advantage of masking the padded zeroes, thus allowing the variable-length sequences to be input into the LSTM model, as well as decreasing model complexity and computation. Moreover, it is noteworthy that the epoch in every model with a TG of 15 min, 30 min, and 60 min is set as 50, 100, and 200, respectively. The batch size is set as 72.

B. EVALUATION METRICS

To evaluate forecasting performance, the root mean squared error (RMSE), mean absolute error (MAE), and weighted mean absolute percentage error (WMAPE) in different time

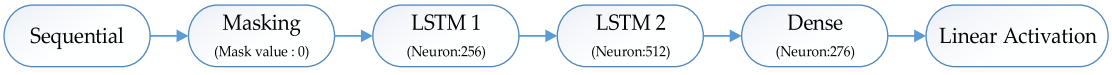

FIGURE 3. Architecture of LSTM mode.

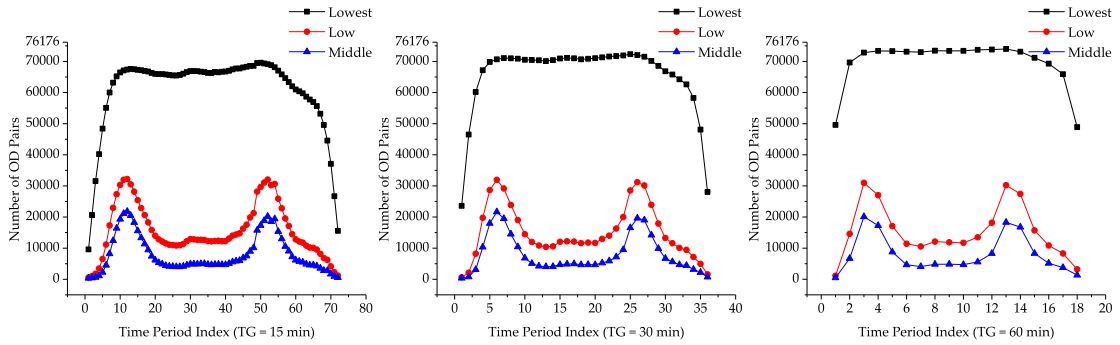

FIGURE 4. Variation of OD pair numbers for each TG case for one day.

periods, as calculated according to Equations (15) to (17), are applied in this study. Corresponding average indicators for a whole day, weighted by total passenger volume in different time periods, are calculated according to Equations (18) to (20).

$$RMSE^t = \sqrt{\frac{1}{n*n}\sum_{i=1}^{n}\sum_{j=1}^{n}\left(q_{ij}^{k,t} - \hat{q}_{ij}^{k,t}\right)^2} \quad (15)$$

$$MAE^t = \frac{1}{n*n}\sum_{i=1}^{n}\sum_{j=1}^{n}\left|q_{ij}^{k,t} - \hat{q}_{ij}^{k,t}\right| \quad (16)$$

$$WMAPE^t = \sum_{i=1}^{n}\sum_{j=1}^{n}\frac{q_{ij}^{k,t}}{\sum_{i=1}^{n}\sum_{j=1}^{n}q_{ij}^{k,t}}\left|\frac{q_{ij}^{k,t} - \hat{q}_{ij}^{k,t}}{\hat{q}_{ij}^{k,t}}\right| \quad (17)$$

$$RMSE = \sum_{t=1}^{m}\frac{\sum_{i=1}^{n}\sum_{j=1}^{n}q_{ij}^{k,t}}{\sum_{t=1}^{m}\sum_{i=1}^{n}\sum_{j=1}^{n}q_{ij}^{k,t}}RMSE^t \quad (18)$$

$$MAE = \sum_{t=1}^{m}\frac{\sum_{i=1}^{n}\sum_{j=1}^{n}q_{ij}^{k,t}}{\sum_{t=1}^{m}\sum_{i=1}^{n}\sum_{j=1}^{n}q_{ij}^{k,t}}MAE^t \quad (19)$$

$$WMAPE = \sum_{t=1}^{m}\frac{\sum_{i=1}^{n}\sum_{j=1}^{n}q_{ij}^{k,t}}{\sum_{t=1}^{m}\sum_{i=1}^{n}\sum_{j=1}^{n}q_{ij}^{k,t}}WMAPE^t \quad (20)$$

where $q_{ij}^{k,t}$ is the prediction value, $\hat{q}_{ij}^{k,t}$ is the actual value, t is the time period during a day k, n is the number of stations, and m is total number of time periods in a day.

C. DISCUSSION OF RESULTS

After the application of ODAD, the number of OD pairs is significantly decreased because of the deletion of OD pairs that actually have little impact on the system, as shown in Figure 4. It can be seen that regardless of which TG is adopted, the number of OD pairs has a similar variation trend.

Applying the "Lowest" ODAD level, the number of OD pairs remains steady from morning peak hours to evening peak hours; nearly 10,000 OD pairs with no passengers. Early in the morning and late at night, these values decline sharply, indicating that there are more OD pairs (from about 10,000 to 60,000) with no passengers at all during in these periods. Therefore, if these OD pairs are taken into consideration in the prediction process, significant computing resources will be wasted, and model performance will also be negatively affected.

Going from "Lowest" to "Low" level, the biggest decrement occurs, signifying that more than half of the OD pairs have very few passengers, i.e., less than or equal to 1, 2, and 4 passengers in 15 min, 30 min, and 60 min, respectively. It is specifically these OD pairs that have a significant impact on the prediction performance. Additionally, the number variations are related to passenger volume, showing obvious double peaks. Even in peak periods, only approximately half of the OD pairs have a higher ODAD level than "Low".

For the "Middle" ODAD level, the number of OD pairs show a similar trend as that of the "Low" level. However, more OD pairs are deleted when the "Middle" level is applied. These observations indicate that the choice of which ODAD level to apply should also be seriously considered to avoid ignoring some important OD pairs.

The model performance under different TG and ODAD levels is highlighted in Figure 5. As clearly shown, nearly all indicators under the three TG settings decrease as ODAD level increases, which strongly indicates that the introduction of ODAD level can significantly improve forecasting precision. In addition, all three indicators are closely related to passenger volume, with the same or the opposite variation trend. Moreover, the decrement from "Lowest" level to "Low" level is larger than that from "Low" to "Middle"; this shows similar regularity with that of the OD pair numbers in Figure 5. Therefore, we recommend the "Low" ODAD level be adopted for subway operation. This will not only ensure forecasting performance to some extent, but also allow as many OD pairs as possible to be considered.

The average model performance for a whole day is presented in Figure 6 and Table 4 to Table 6. The average RMSE with the different TGs of 15 min, 30 min, and 60 min is approximately 1.3, 2.3, and 4.4, respectively. Although the RMSE for the same TG is approximately the same, there also exists a slight decline. By contrast, the average MAE and WMAPE vary greatly with the same TG. The smallest MAE (0.36921) occurs when a TG of 15 min and ODAD level of "Middle" are adopted.

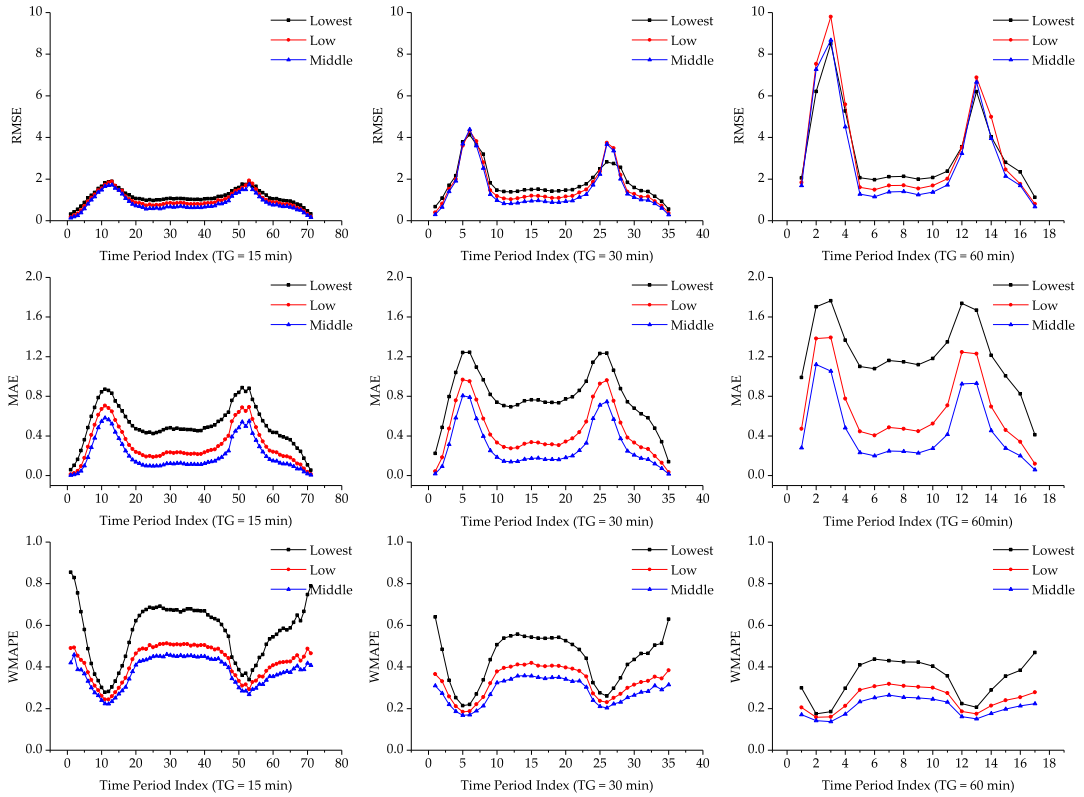

FIGURE 5. Model performance in different time periods.

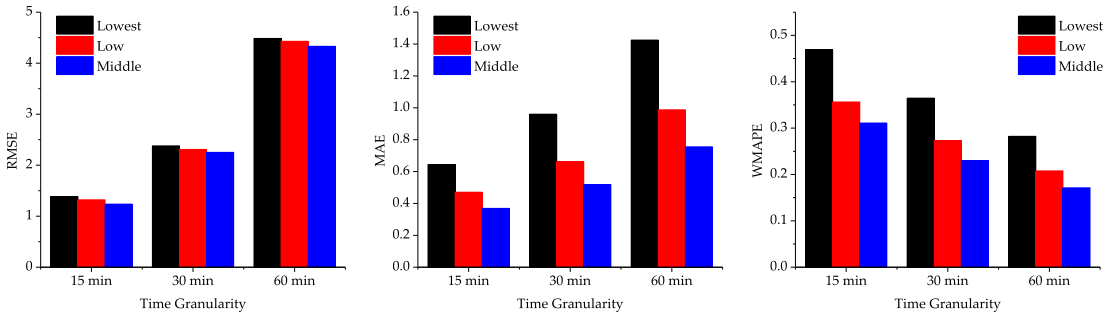

FIGURE 6. Average model performance for a whole day.

TABLE 4. Average RMSE for a whole day.

ODAD\TG	15 min	30 min	60 min
Lowest	1.38486	2.37934	4.48547
Low	1.32064	2.30958	4.42929
Middle	1.23677	2.25166	4.33042

TABLE 5. Average MAE for a whole day.

ODAD\TG	15 min	30 min	60 min
Lowest	0.6443	0.96018	1.42481
Low	0.47066	0.66186	0.98759
Middle	0.36921	0.51825	0.75546

WMAPE increases to 46.9% when a TG of 15 min and ODAD level of "Lowest" is adopted, signifying that the model performance is slightly worse in this setting. Even when the ODAD level of "Middle" is adopted, WMAPE exceeds 30%. There are several reasons leading to this situation. One is that the trip regularity is lower from a statistical perspective when passenger flow is aggregated in a smaller TG. Another is the uncertainty of travel time. For example, the starting travel times of passengers are often not limited to a specific 15-min interval (e.g., 7:00 to 7:15). This may change with a slightly longer interval (e.g., 30 min). When a TG of 30 min and an ODAD level of "Low" is adopted, WMAPE (27.28%) begins to be less than 30%. When the TG is 60 min, WMAPE is always less than 30%. However, a bigger TG is not beneficial for real time control of subway systems, and more detailed passenger flow

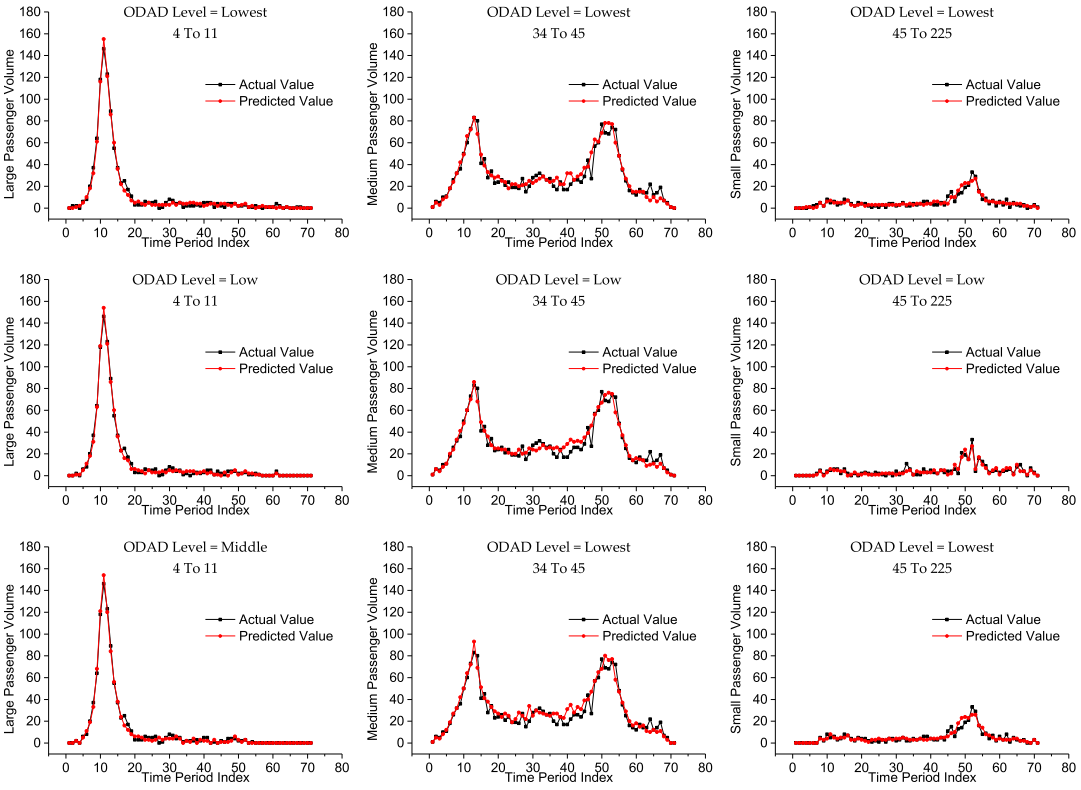

FIGURE 7. Comparison of actual values and predicted values when TG equals 15 min.

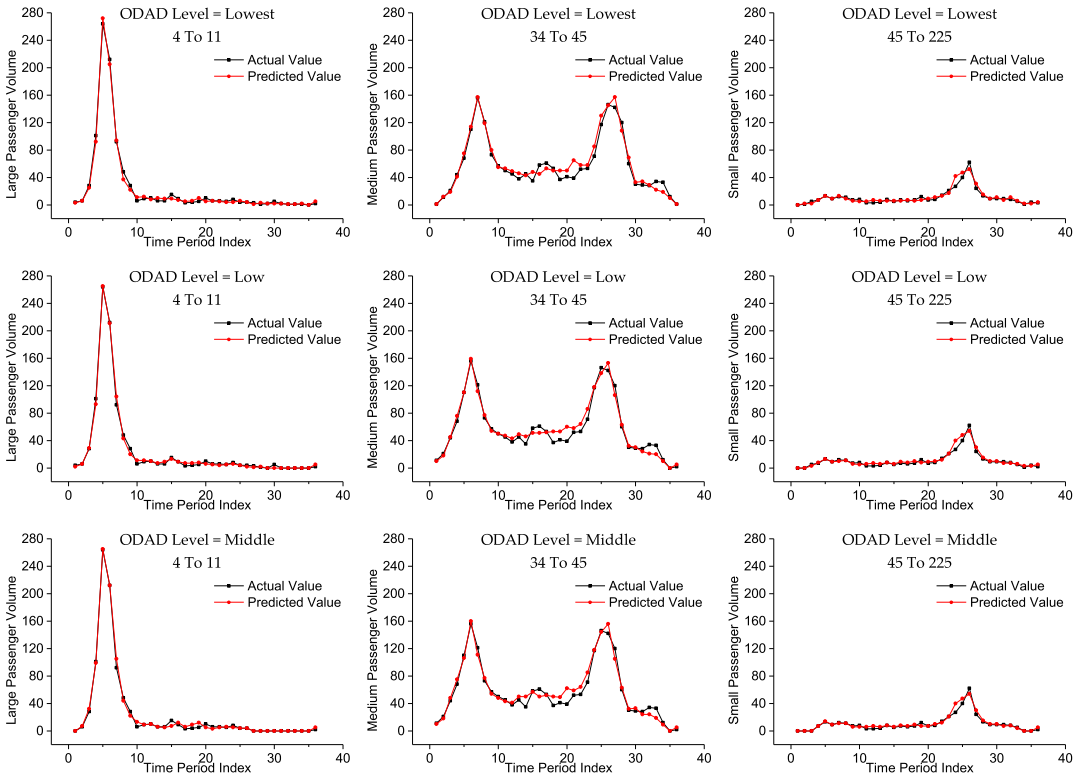

FIGURE 8. Comparison of actual values and predicted values when TG equals 30 min.

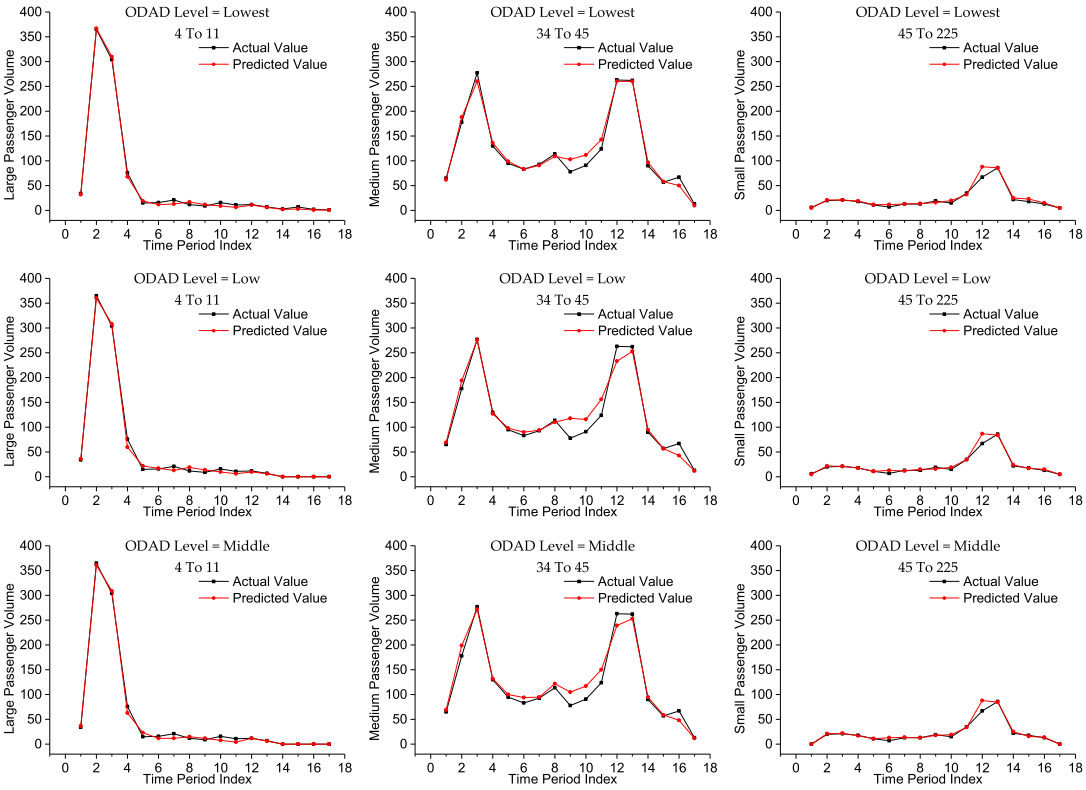

FIGURE 9. Comparison of actual values and predicted values when TG equals 60 min.

TABLE 6. Average WMAPE for a whole day.

ODAD\TG	15 min	30 min	60 min
Lowest	46.91%	36.45%	28.22%
Low	35.63%	27.28%	20.77%
Middle	31.10%	22.97%	17.12%

TABLE 7. Description of OD pairs.

OD pairs	4 To 11	34 To 45	45 To 225
Name	Yuquanlu To Fuxingmen	Xizhimen To Dongzhimen	Dongzhimen, To Wangjingxi
Passenger Volume	Large	Medium	Small
passenger flow characteristics	Morning Peak	Double Peak	Evening Peak

information will be ignored. Therefore, a TG of 30 min and an ODAD level of "Low" is recommended for actual subway operation.

In addition, we have chosen three OD pairs, namely from 4 to 11, 34 to 45, and 45 to 225, to show the actual values and predicted values in different time periods. The station numbers of 4, 11, 34, 45, and 225 represent Yuquanlu, Fuxingmen, Xizhimen, Dongzhimen, and Wangjingxi stations, respectively. Moreover, the three OD pairs represent stations with large, medium, and small passenger volumes as well as different passenger flow characteristics as shown in Table 7; this facilitates a more general representation, as shown in Figure 7 to Figure 9. It can be clearly seen that whichever the case, the LSTM model is able to capture the passenger flow variation trend; this demonstrates the feasibility of the LSTM model.

V. CONCLUSION

This study innovatively proposed an ODAD indicator to conduct short-term OD forecasting in URT. The mature and stable LSTM model is selected to verify the effect of ODAD. First, we introduced an ODAD indicator to capture the attraction degree of OD pairs. The ODAD level with different TG values was also given. Based on the ODAD, a detailed OD matrix preprocessing method was presented to prepare data for the LSTM model. Second, the theory of the LSTM model was clearly described to conduct short-term OD forecasting. Finally, nine cases were studied, covering three TGs and three ODAD levels, to verify the effect of introducing the ODAD indicator; the best optimal combination of TG and ODAD levels was also explored. Several critical findings are summarized as follows:

1. The introduction of the ODAD can effectively improve forecasting model performance. Moreover, the ideal adopted ODAD level depends on actual operational requirements.
2. It is feasible that the LSTM model can be applied to conduct short-term OD forecasting in URT with a large-scale network. In addition, LSTM can skilfully operate on input sequences of variable lengths by leveraging the masking layer, which facilitates the

application of the ODAD. Furthermore, the LSTM model can clearly capture passenger flow variations throughout a day.

3. A smaller TG can decrease OD matrix predictability. A larger ODAD level will cause many OD pairs to be ignored. Using case study results, we recommend a TG of 30 min and an ODAD level of ''Low'', which can ensure satisfactory forecasting results. In this case, the RMSE, MAE, and WMAPE are 2.31%, 0.66%, and 27.28%, respectively, in a network of nearly 300 subway stations.

Overall, these findings can provide insights for subway operators to conduct short-term OD forecasting in URT using LSTM or other deep learning methods. However, there are also some limitations. For example, only OD flows on weekdays were studied, and weekends were not included. In addition, other factors that influence OD flows, such as the weather and precipitation, were not considered. Moreover, we did not use more complicated deep learning models. We only chose a relatively simple, mature, and stable LSTM model to verify the effect of introducing ODAD, which is the main contribution of this study. Therefore, we did not compare LSTM with other models. Future work can conduct additional research in these areas.

CONFLICTS OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGMENT

The authors are grateful to the Fundamental Research Funds for the Central Universities under Grant 2018YJS128 and the China Scholarships Council under Grant 71871027 for their financial support. This work was facilitated through the use of advanced computational, storage, and networking infrastructure provided by the Hyak supercomputer system at the University of Washington.

REFERENCES

[1] X. Wang *et al.*, ''Short-term forecasting of urban rail transit ridership based on ARIMA and wavelet decomposition,'' in *Proc. AIP Conf.*, vol. 1, no. 1. New York, NY, USA: AIP, 2018, Art. no. 040025.

[2] J. Roos, G. Gavin, and S. Bonnevay, ''A dynamic Bayesian network approach to forecast short-term urban rail passenger flows with incomplete data,'' *Transp. Res. Procedia*, vol. 26, pp. 53–61, Jan. 2017.

[3] 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.

[4] X. Feng, H. Zhang, T. Gan, Q. Sun, F. Ma, and X. Sun, ''Random coefficient modeling research on short-term forecast of passenger flow into an urban rail transit station,'' *Transport*, vol. 31, no. 1, pp. 94–99, 2016.

[5] J. W. Li, ''Short-time passenger volume forecasting of urban rail transit based on multiple fusion,'' *Appl. Mech. Mater.*, vols. 641–642, pp. 773–776, Sep. 2014.

[6] 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.

[7] E. I. Vlahogianni, M. G. Karlaftis, and J. C. Golias, ''Short-term traffic forecasting: Where we are and where we're going,'' *Transp. Res. C, Emerg. Technol.*, vol. 43, pp. 3–19, Jun. 2014.

[8] L. Nie, D. Jiang, L. Guo, S. Yu, and H. Song, ''Traffic matrix prediction and estimation based on deep learning for data center networks,'' in *Proc. IEEE Globecom Workshops (GC Wkshps)*, Dec. 2016, pp. 1–6.

[9] P.-W. Lin and G.-L. Chang, ''A generalized model and solution algorithm for estimation of the dynamic freeway origin–destination matrix,'' *Transp. Res. B, Methodol.*, vol. 41, no. 5, pp. 554–572, Jun. 2007.

[10] Y. Vardi, ''Network tomography: Estimating source-destination traffic intensities from link data,'' *J. Amer. Statist. Assoc.*, vol. 91, pp. 365–377, 1996.

[11] Y. Zhang, M. Roughan, N. Duffield, and A. Greenberg, ''Fast accurate computation of large-scale IP traffic matrices from link loads,'' in *Proc. ACM SIGMETRICS Perform. Eval. Rev.*, Jun. 2003, pp. 206–217.

[12] A. Lakhina, K. Papagiannaki, M. Crovella, C. Diot, E. D. Kolaczyk and N. Taft, ''Structural analysis of network traffic flows,'' in *Proc. ACM SIGMETRICS Perform. Eval. Rev.*, Jun. 2004, pp. 61–72.

[13] L. Tan and X. Wang, ''A novel method to estimate ip traffic matrix,'' *IEEE Commun. Lett.*, vol. 11, no. 11, pp. 907–909, Nov. 2007.

[14] D. Jiang, X. Wang, L. Guo, H. Ni, and Z. Chen, ''Accurate estimation of large-scale IP traffic matrix,'' *AEU Int. J. Electron. Commun.*, vol. 65, no. 1, pp. 75–86, Jan. 2011.

[15] H. Zhou, L. Tan, Q. Zeng, and C. Wu, ''Traffic matrix estimation: A neural network approach with extended input and expectation maximization iteration,'' *J. Netw. Comput. Appl.*, vol. 60, no. 6, pp. 220–232, Jan. 2016.

[16] J. Zhao, H. Qu, J. Zhao, and D. Jiang, ''Towards traffic matrix prediction with LSTM recurrent neural networks,'' *Electron. Lett.*, vol. 54, no. 9, pp. 566–568, May 2018.

[17] A. Azzouni and G. Pujolle, ''A long short-term memory recurrent neural network framework for network traffic matrix prediction,'' 2017, *arXiv:1705.05690*. [Online]. Available: https://arxiv.org/abs/1705.05690

[18] W. E. Leland, M. S. Taqqu, W. Willinger, and D. V. Wilson, ''On the self-similar nature of Ethernet traffic,'' in *Proc. ACM SIGCOMM Comput. Commun. Rev.*, Oct. 1993, pp. 183–193.

[19] M. G. H. Bell, ''The estimation of origin-destination matrices by constrained generalised least squares,'' *Transp. Res. B, Methodol.*, vol. 25, no. 1, pp. 13–22, Feb. 1991.

[20] M. G. H. Bell, ''The real time estimation of origin-destination flows in the presence of platoon dispersion,'' *Transp. Res. B, Methodol.*, vol. 25, nos. 2–3, pp. 115–125, 1991.

[21] H. Yang, ''Heuristic algorithms for the bilevel origin-destination matrix estimation problem,'' *Transp. Res. B, Methodol.*, vol. 29, no. 4, pp. 231–242, 1995.

[22] H. Yang, T. Sasaki, Y. Iida, and Y. Asakura, ''Estimation of origin-destination matrices from link traffic counts on congested networks,'' *Transp. Res. B, Methodol.*, vol. 26, no. 6, pp. 417–434, 1992.

[23] H. Yang, Y. Iida, and T. Sasaki, ''An analysis of the reliability of an origin-destination trip matrix estimated from traffic counts,'' *Transp. Res. B, Methodol.*, vol. 25, pp. 351–363, Oct. 1991.

[24] J. Ou, J. Lu, J. Xia, C. An, and Z. Lu, ''Learn, assign, and search: Real-time estimation of dynamic origin-destination flows using machine learning algorithms,'' *IEEE Access*, vol. 7, pp. 26967–26983, 2019.

[25] L. Liu, Z. Qiu, G. Li, Q. Wang, W. Ouyang, and L. Lin, ''Contextualized spatial-temporal network for taxi origin-destination demand prediction,'' *IEEE Trans. Intell. Transp. Syst.*, to be published.

[26] J. Zhao, A. Rahbee, and N. H. Wilson, ''Estimating a rail passenger trip origin-destination matrix using automatic data collection systems,'' *Comput. Aided Civil Infrastruct. Eng.*, vol. 22, no. 5, pp. 376–387, Jul. 2007.

[27] X. M. Yao, P. Zhao, and D.-D. Yu, ''Real-time origin-destination matrices estimation for urban rail transit network based on structural state-space model,'' *J. Central South Univ.*, vol. 22, no. 11, pp. 4498–4506, Nov. 2015.

[28] X. M. Yao, P. Zhao, and D. D. Yu, ''Dynamic origin-destination matrix estimation for urban rail transit based on averaging strategy,'' *J. Jilin Univ. (Eng. Technol. Ed.)*, vol. 46, no. 1, pp. 92–99, 2016.

[29] Z.-J. Chen, B.-H. Mao, Y. Bai, Q. Xu, and T. Zhang, ''Short-term Origin-destination Estimation for Urban Rail Transit Based on Multiple Temporal Scales,'' *Transp. Syst. Eng. Inf. Technol.*, vol. 15, no. 5, pp. 166–172, 2017.

[30] A. Soule, A. Lakhina, N. Taft, K. Papagiannaki, K, Salamatian, A. Nucci, M. Crovella, and C. Diot, ''Traffic matrices: Balancing measurements, inference and modeling,'' in *Proc. ACM SIGMETRICS Perform. Eval. Rev.*, Jun. 2005, pp. 362–373.

[31] C. Olah. (2015). *Understanding LSTM Networks*. [Online]. Available: https://colah.github.io/posts/2015-08-Understanding-LSTMs/

[32] F. A. Gers, J. Schmidhuber, and F. Cummins, ''Learning to forget: Continual prediction with LSTM,'' in *Proc. 9th Int. Conf. Artif. Neural Netw. (ICANN)*, 1999, pp. 850–855.

[33] S. Hochreiter and J. Schmidhuber, ''Long short-term memory,'' *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.

JINLEI ZHANG was born in Hebei, China. He is currently pursuing the Ph.D. degree with Beijing Jiaotong University.

He is currently a Researcher with the University of Washington, as a Visiting Ph.D. Student. His research interests include traffic data mining and applications, and dynamic traffic modeling and management.

FENG CHEN received the B.S., M.A., and Ph.D., degrees in railway engineering from Beijing Jiaotong University, Beijing, China, in 1983, 1990, and 2007, respectively.

He is currently a Professor with Beijing Jiaotong University and the President of Chang'an University. His research interests include passenger flow management, traffic data mining and application for urban rail transit.

Prof. Chen received the First Prize of the National Science and Technology Progress, as the second accomplisher, in 2017.

ZIJIA WANG was born in Henan, China. He received the Ph.D. degree from the School of Civil Engineering, Beijing Jiaotong University, where he is currently an Associate Professor.

His research interests include transportation planning, including pedestrian flow modeling, pedestrian simulation research, and transportation demand analysis and forecasting.

HANXIAO LIU was born in Hebei, China. She received the M.E. degree from Beijing Jiaotong University, in 2019.

She is currently with Beijing General Municipal Engineering Design and Research Institute Co., Ltd. Her current research interests include traffic planning, rail transit operation, and traffic data mining.