A Spatio-Temporal Structured LSTM Model for Short-Term Prediction of Origin-Destination Matrix in Rail Transit With Multisource Data

DEWEI LI\textsuperscript{1}, JINMING CAO\textsuperscript{1,2}, (Student Member, IEEE), RUOYI LI\textsuperscript{3}, (Student Member, IEEE), AND LIFU WU\textsuperscript{1}, (Student Member, IEEE)

\textsuperscript{1}School of Traffic and Transportation, Beijing Jiaotong University, Beijing 100044, China
\textsuperscript{2}Beijing General Municipal Engineering Design and Research Institute Company, Ltd., Beijing 100082, China
\textsuperscript{3}Wuhan Planning and Design Institute, Wuhan 430000, China

Corresponding author: Dewei Li (lidw@bjtu.edu.cn)

This work was supported in part by the National Key Research and Development Plan under Grant 2018YFB1201402, in part by the National Natural Science Foundation of China under Grant 71971019, in part by the Fundamental Research Funds for the Central Universities under Grant 2018JBMO31 and Grant G2019JBMO29, and in part by the Programme of Introducing Talents of Discipline to Universities under Grant B18004.

ABSTRACT

Passenger assignment of rail transit has recently attracted increasing research interest due to its potential applications in large-scale intelligent transportation systems. In the rail transit system, the foundation of passenger assignment is passengers’ origin and destination demand (OD matrix). However, due to the nature of stochastic of the short-term dynamic OD matrix, how to accurately predict the distribution of passenger travel spatio-temporally is still an open challenge. In this paper, combined multisource data with deep learning method is proposed to improve prediction of dynamic OD matrix accuracy. Firstly, multisource data such as smart card data, weather data and mobile phone data are introduced. And after quantitative analysis of the influencing factors, choosing 31 features as model inputs. Secondly, considering the superiority of Long Short-term Memory Network in time series, we improve the structure of LSTM by redesigning the hidden layer and neuron, in view of the spatio-temporal characteristics of spatio-temporal Long Short-term Memory Network (STLSTM) of rail transit passenger flow. Finally, using the Beijing subway network which had 54,056OD for verification. Extensive experiments and evaluations on a large-scale dataset well demonstrate the superiority of STLSTM over commonly used prediction models and standard LSTM model for short-term prediction of dynamic OD matrix. In addition, the application method of multisource data in OD prediction in this paper can deal with more data from other sources to further improve the information exploit effect on passenger flow law.

INDEX TERMS

LSTM, multisource data, rail transit, short-term OD prediction.

I. INTRODUCTION

Rail transit is an important means to alleviate the problems caused by urbanization such as traffic congestion and environmental pollution. Currently, to enhance the operation of rail transit, predictive control strategy is being applied. The basic logic is firstly to predict the passenger distribution in the network (normally every 15 minutes). Dynamically evaluate the distribution of network passenger flow, if there are local congestion problems in the distribution, put some countermeasures such as reschedule [1] and flow control [2] based on the prediction to dynamically improve subway operation management capacity and passenger travel quality. The first step of this strategy is to predict the OD matrix of passenger flow accurately.

However, few methods can significantly improve the accuracy of dynamic OD matrix predictions in rail transit, especially during peak hours [3]. There are two parts of the main reason. One is the restrictions of the data used. Passenger travel is affected by a variety of factors. The data with a single source is difficult to include all passenger travel laws, which makes it difficult to improve the accuracy of prediction. In view of this situation, Ni et al. [4] and Zhang et al. [5] has made effective improvements. The second question is which...
predict method can be used to analyze the information in different dimensions. Valid information for each source needs to be retained to the maximum by the method, and meanwhile random quantities can be eliminated in order to truly improve the accuracy of predictions. This paper first introduces as much as multisource data from different sources, e.g. smart card data, mobile phone data, weather data etc. and then, it exploits the main features behind these data. Then, in view of the spatio-temporal characteristics of rail passenger flow, we improve the structure of LSTM by redesigning the hidden layer and neuron. That is, the ST-LSTM model is designed for the characteristics of spatio-temporal dimension data. By doing so, it shows that the accuracy of prediction can be significantly improved.

A. LITERATURE REVIEW

Accurate passenger demand forecast is an important basis for the release of rail transit information and the implementation of passenger control strategy. Previous studies have mainly focused on forecast of station passenger flow of rail transit system and static OD prediction. Due to the large scale of rail transit network system, the complex of passenger travel behavior and other reasons, short-term dynamic OD passenger flow prediction has been seldom studied. Nevertheless, because the time series data of short-time OD passenger flow are similar to the time series data of station passenger flow and cross section passenger flow, the corresponding research results can also provide valuable insights for short-time OD passenger flow prediction.

The short-term passenger forecasting approaches can be generally divided into two categories: parametric and non-parametric techniques [6]–[8].

In the traditional parametric techniques that can be used to achieve short-term OD forecasting are mainly historical average [9], smoothing techniques [10], and autoregressive integrated moving average (ARIMA) [11]. The operation of the historical average model is the simplest. Although the method is simple to calculate and highly efficient, it is generally less accurate to predict with the uncertainty and nonlinear characteristics of passenger flow [9]. The prediction is not stable enough [10] and it can be greatly affected by sudden passenger flow. The ARIMA model can effectively eliminate the short-term fluctuations in the OD passenger flow time series, so it can better capture the long-term characteristics of OD passenger flow distribution time series data. On the basis of the ARMA model, many researchers have proposed various variant models, some of which are well-known, including the Kohonen-ARIMA model [12], Subset ARMA model [13] Seasonal ARIMA model [11] and so on. When the distribution of OD passenger flow changes greatly, there are some deficiencies in the ARIMA model prediction delay. So, this method is more suitable for relatively smooth traffic flow prediction, and it is difficult to apply to large-scale OD volume forecast.

For the non-parametric techniques, several methods that can be used to achieve short-term OD traffic forecasting are mainly Kalman filtering models [14], non-parametric regression [7], [15], Artificial neural network [16]. Kalman filtering model is an optimal self-regression data processing algorithm based on linear filtering theory, which updates the equation by adding new OD passenger flow data to achieve real-time OD passenger flow prediction. The model performs well in processing both smooth and non-stable data, but the method has the problem of easily insufficient filtering. Okutani and Stephanedes [17] applied the Kalman filtering model to short-term OD passenger prediction problems for the first time. Van Lint [18] had achieved online real-time prediction of passenger flow through Kalman filtering model. Yao [19] established a short-term OD matrix estimation state spatial model based on Kalman filtering method, and realized the dynamic OD matrix estimation of Beijing’s whole network under the time granularity size of 15min, 30min and 60min. In addition, they [19] successfully reduced the forecast error of OD volume at 15min time granularity to 30.8% by minimizing the deviation of the current OD passenger flow at 60min time scale from the historical OD passenger flow in the same period. Non-parametric regression such as the k-Nearest Neighbors algorithm (k -NN) [20] and Support Vector Machine (SVM) is widely used in prediction. Castro-Neto [22] used the Online-SVR (support vector regression) model to achieve short-term traffic flow forecasts in typical and atypical environments. Castro-Neto [23] further improved the online-SVR model into a supervised online weighted SVM model (OLWSVR model) with good prediction results. Jeong et al. [23] used several forecasting models such as SVR to investigate the effect of temporal and spatial features as well as external weather influence on passenger flow forecasting, and the validity of this method is verified by the actual data of Shenzhen metro. Artificial neural network (ANN) trains models by using a large amount of historical data to obtain a more accurate mapping relationship between output and input. ANN has a strong dependence on the data, and needs a lot of sample data training to get a good prediction result. It can be used as a powerful tool to describe OD passenger flow, because it can describe the characteristics of nonlinear system and take consideration into the relationship between OD passenger flow and spatio-temporal dimension. However, in the course of training the traditional neural network model, a great deal of marker data is required. The BP (back propagation) algorithm commonly, used to solve the weight parameters is prone to gradient disappearance and gradient explosion in the case of more model layers, and the result may fall into local optimum. Smith and Demetsky [25] applied the BP neural network model to short-term traffic flow prediction early. Hu et al. [26] pre-processed the historical data firstly, and then put the differential data into the BP neural network to predict the short-term passenger flow. Zhu et al. [27] predicted the passenger flow by using radial base function (RBF), and considered the impact of traffic conditions at intersections. As a major breakthrough in the past ten years, deep learning has achieved good results in many tasks such as...
speech recognition and image processing [28]. As a machine learning method that integrates unsupervised learning, semi-supervised learning and supervised learning, deep learning has a variety of models and algorithms for all kinds of scenarios [29]. As a result, it is ideal for dealing with complex, high-dimensional short-term OD passenger flow forecasting issues. Recurrent Neural Network (RNN) in Deep Learning introduces the concept of timing into existing neural networks [30], this model is ideal for learning long-term dependent timing data such as OD passenger flow data because the output of the precedent moment has a direct impact on the input at the current moment [31]. Long-Short Term Memory (LSTM) is a special RNN network designed to solve the problem of gradient disappearance or gradient explosion. Ma et al. [32] constructed the traffic speed prediction model based on LSTM network and trained the LSTM network with remote microwave sensor data. Tian and Pan [33] discussed the performance of LSTM network in short-term traffic forecasting in detail, and demonstrated the advantages of LSTM by comparing LSTM with several other commonly used models. Rui et al. [34] predicted the short-term traffic flow by LSTM and GRU models, and confirmed the LSTM model is better than the ARIMA model in the short-term traffic flow prediction. Gu et al. [35] combined GRU, ARIMA and RBF to predict short-term traffic volume, and confirmed the combined method is better than other methods in terms of accuracy and stability. Toque et al. [36] used the LSTM model that input the subway and bus passenger flow data to subway predict short-term OD passenger flow. The result showed that LSTM model has obvious advantages in subway short-term OD passenger flow prediction, and after increasing the input of bus data, the prediction accuracy of the model has been further improved. The TABLE 1 is a summary of the recent research.

These methods, which appear in different stages of the development of short-term OD prediction, have good predictive ability, and can reflect the essential characteristics of OD passenger flow time series change to a certain extent. With the sharing of data between different sources, it is difficult to analyze different sources data by existing methods, and it is difficult to exploit the potential temporal and spatial relationships of different data. This paper compares different methods and considers the advantages of LSTM network model in processing time series, selects it as the model basis, and improves the structure of LSTM by redesigning the hidden layer and neuron to fully exploit the spatio-temporal information of data hiding.

B. MAIN CONTRIBUTIONS

This paper studies the short-term OD passenger flow prediction in rail transit, and construct a short-term OD prediction model based on an STLSTM network under multisource conditions through analysis and selection of multisource data, and compare the assessments with other methods.

There are three main contributions. Firstly, introduce multisource data into the problem of short-term OD prediction in the context of rail transit, and to construct a prediction model under multisource data. Compared with the standard LSTM network under the single source, it can be seen that the model can extract more characteristics with an increasing input of multisource data, which improves prediction accuracy of the model significantly under 15min, 30min and 60min time granularity. Secondly, to fully capture the characteristics of each OD in OD passenger flow distribution time series data and the overall characteristics of each period, an STLSTM model is obtained by improving the hidden layer structure and neuron structure of the standard LSTM network. The result shows that after improving the model structure, the model can fully capture the characteristics of each OD pair and overall characteristics of each time series in the OD passenger flow distribution time series, so that the prediction effect of the model under different time granularities can be obviously improved. Finally, large-scale rail transit network and mass data are applied. This paper is verified by using 54,056 OD pairs, 199.6 million smart card data records and 10.5 million mobile phone data in Beijing rail transit network, and the STLSTM can be fully trained based on the massive multidimensional data.

II. PROBLEM DESCRIPTION

Passenger demand in the transportation sector is for specific origins $O$ and destinations $D$, and passenger demand in rail transit systems can be reflected by smart card data in different time periods. The distribution of passenger flow between

| Author | OD prediction | Rail system | Short term prediction | Method | Input data |
|--------|---------------|-------------|----------------------|--------|------------|
| Chen (2011) | ✓ | × | ✓ | LSTM | SC |
| Jeong (2013) | × | × | ✓ | OLWSVR | HS |
| Zhang (2013) | × | × | ✓ | K-NN | HS |
| Ma (2015) | × | × | ✓ | LSTM | GPS |
| Fu (2016) | × | × | ✓ | LSTM/GRU | HS |
| Toque (2016) | ✓ | ✓ | ✓ | LSTM | SC |
| Wang (2016) | ✓ | ✓ | ✓ | RBF/LSL SVM | SC |
| Ma (2017) | × | × | ✓ | Convolutional neural network | GPS |
| Li (2017) | ✓ | ✓ | ✓ | Multiscale RBF | SC |
| Ni (2017) | ✓ | ✓ | ✓ | Linear regression | SC/S |
| Li (2019) | × | × | ✓ | Moop-DBN | HS |
| Tang (2019) | ✓ | ✓ | ✓ | SVR | SC/W |
| This Paper | ✓ | ✓ | ✓ | STLSVM | SC/W/Mb |

Note: SC-smart card; W-Weather; SM-Social media data; HS-Highway surveillance data; *-multi-objective particle swarm algorithm-deep belief network
ODs is influenced by people’s daily activities, which presents inherent law and trend in the time dimension. In addition, there are mutual influence among different ODs. Influenced by the traffic trend, there is a time correlation among ODs, affected by station land properties, distance, etc. and there is also spatial interaction of ODs.

OD pairs are determined by the departure station \( o \) and the destination station \( d \), \((o, d)\) is used to present OD pairs. For the convenience of the description, the following variables are defined as follows: in rail transit networks, the set of network nodes are represented as \( N, N = \{1, 2, \ldots, i, \ldots, j, \ldots, n\} \), the total number of OD pairs is \( M = n \times (n-1) \); time set is \( T, T = \{1, 2, \ldots, k, \ldots, t\} \); \( \tau \) represents prediction time granularity, \( No \) \( k \), time range contained the time range of \( k \tau \), \( k \tau \); \( x_k^{(i,j)} \) is the number of passengers from origin \( i \) to destination \( j \) in time interval \( k \), \( x_k \in \mathbb{R}^M \) is the number of passengers of all OD pairs within the time interval. In practical application, there will be a lag between the inbound time and outbound time collected by smart card system. And there is passenger flow lag effect in OD matrix estimation, that is, the passenger flow information collected in the earlier period has not left the rail transit network, and this kind of passenger flow diffuses results in mismatching of the passenger flow information collected in the same period to each other. Therefore, estimates should not only take into account traffic flow information for the current time period, but also include information from multiple time periods in the previous period, which greatly increases the complexity of the process[19].

The basic relationship between inbound and outbound quantity collection period is shown in Fig. 2.

The goal of this paper is to improve the prediction of short-term OD passenger flow accuracy in rail transit system. That is to predict the quantity of OD \( x_{i+1} \) from the origin to the destination in all the ODs after \( l \) time intervals with the observed values known \( k \) of the time series of OD passenger flow distribution.

III. DATA COLLECTION AND FEATURE SELECTION

In order to improve the accuracy of short-term OD matrix prediction, a large number of multisource data are collected and processed as input during neural network training. Smart card data are a very important data source for short-term OD prediction of rail transit. It is based on the fact that the OD matrixes show significant similarities to historical values, especially during peak hours [19]. Each complete smart card data contains the time of inbound and outbound stations, which are the basis of model input. In addition, weather information [40] will affect passengers’ willingness to travel and the choice of transportation modes. Mobile phone data show the users condition around the station (i.e. crowd gathering hot spots). These three kinds of information can also provide more features to explain irregular passenger flow. Multisource data has a very vital role in rendering the STLSTM construction and the final result for the network, so it should be adapted to the characteristics of the data. Therefore, some statistical methods, such as classification and regression, are used to evaluate the correlation between data as much as possible, in order to choose the suitable data for data mining and avoid repeating calculation or missing important information in this paper. Combined with the above-mentioned data, this paper collects smart card data, mobile phone data and weather data as data sources, and processes and extracts the data as input to the STLSTM.

A. DATA COLLECTION

The data used in this paper consists of three main aspects including smart card data, weather data and mobile phone data [41]. Firstly, smart card data are obtained and filtered by the smart card system, and then process the 199.6 million data records to avoid the impact of holidays on forecasts. The data set contains smart card data and weather data of Beijing from February 15 to April 4, 2014, a total of seven consecutive weeks as a data set. Secondly, weather data are obtained by the weather news Internet (weather data source: http://www.tianqihoubao.com). With the increasing penetration of mobile phones in China, mobile phone users cover most of the population in the city, so it is possible to use mobile phone data to aid OD passenger flow analysis. Thirdly, using mobile phone data to aid short-term OD volume prediction by converting mobile phone data into time-changing user location information table. This paper collects some mobile phone signaling data provided by mobile operators in Beijing on September 2, 2010, with a total of 10.5 million data, which can reflect the attractiveness of the station to some extent. Take the distribution of 8:00-8:15 a.m. on September 2, 2010 for example, the results are shown in Fig.3. Among them, the black dots are the subway station, and the red dots are the target locating points for the mobile phone users. The darker the color is, the denser the points are.

B. FEATURE SELECTION

In order to fully exploit the passenger spatio-temporal information hidden by multisource data, it is necessary to process the characteristics of the data before data input [41]. Therefore, this paper deals with the operating day information, front-order passenger flow information, station attribute
information and other features contained in the data, and explains the reasons in detail for choosing the feature and the quantitative processing methods, which are as follows:

1) INFLUENCE FACTOR OF OPERATION DAY
The distribution of OD passenger flow on the rail transit network will show different characteristics on different operating days. Considering the similarity of the distribution characteristics of the network OD passenger flow, the characteristics of different operating day’s passenger flow can be analyzed and classified by using the collected historical OD passenger flow information. And it can save the analysis of passenger flow characteristics in the same characteristic day for the follow-up study. The sample OD matrix which is made by smart card data clustered by K-Means algorithm. Because the K-Means algorithm is sensitive to the initially selected centroid points, it is necessary to use other methods to determine the number of classifications and the initial centroid points. In order to obtain accurate results, we used within-group Linkage and Ward method [43] (both select square Euclidean distance to calculate the distance between samples) to cluster respectively first. Then it is found that when the number of classifications is 3, the results of the two methods are consistent. Finally, the centers of gravity of these 3 categories are input as the initial centroid points, and the K-Means algorithm is used to obtain the final classification result of the operation day.

According to the results of K-Means method, operating days can be divided into three types of passenger flow characteristic days. It can be found that there is a certain correlation between OD passenger flow distribution and passenger flow characteristic day type. On such normal working days from Monday to Thursday, commuter traffic (mainly work and school) is the main body of rail transit system, so the distribution structure of passenger flow is affected by the attractiveness of places such as work units and schools. The composition of passengers on Friday mornings is basically similar to those made up from Monday to Thursday, but in the afternoon and evening, passengers travel in a manner and purpose that differs from Monday to Thursday. At normal weekends, the main purposes of passenger travel include shopping, sightseeing and visiting relatives and friends, so the distribution structure of passenger flow is affected by attractions such as recreational facilities, tourist attractions and business centers.

In summary, the normal operating day of rail transit (excluding special holidays) can be divided into characteristic day 1 (Monday to Thursday), characteristic day 2 (Friday) and characteristic day 3 (Saturday to Sunday).

2) INFLUENCE FACTORS IN OPERATION PERIOD
Within an operating day, there will be a large difference in passenger flow at different times, for example, during the morning and evening rush hour, due to the large number of commuters, the overall passenger flow of the network will be significantly higher than the off-peak time. (Due to the limited data, this article does not take holiday passenger flow into consideration.)

In the preceding sections, the normal operating day is divided into three types of characteristic days. In theory, different characteristic days will not show exactly the same operating period division results. However, the OD passenger flow data display the characteristics of timing, and the hierarchical clustering method and K-Means method do not apply to the clustering of time series data. The optimal sorting method is used to orderly cluster data set’s operating time in this paper. The clustering process is described below using the characteristic day 1 as an example. Because the number K of classifications is difficult to determine in advance, this paper calculates the classification loss function, plots the trend of the classification loss function as the number of classifications increases, the results of which are shown in Fig. 4.

As can be seen from Fig. 4, when $K \leq 5$, the classification loss function decreases rapidly with the increment of $K$; when $K \geq 10$ during the operating period, the classification loss function decreases slowly with the increment of $K$; When $K = 8$ there is a turning point indicating the slowdown of the changing trend of the classification loss function. Therefore, this paper sets the number of operational period...
classifications to 8, so that for characteristic day 1, at which point the classification loss function value is \( L[p(n, k)] = 2087.637 \).

The same is the clustering of OD passenger flow of characteristic days 2 and 3, and the final result is that the classification number of characteristics day 2 is also set to 8, and the number of characteristic day 3 is set to 6. The divisions of operating hours for different characteristic days are shown in Table 2.

### 3) HISTORICAL PASSENGER FACTORS DURING THE SAME PERIOD

The OD spatial and temporal distribution of passenger flow has a certain stability, that is, the historical passenger flow and the daily passenger flow has a certain correlation. In the actual rail transit operation management, last week’s OD passenger flow is used to assist the current operation management, so the impact of the historical period of passenger flow data on the current time OD passenger flow distribution is analyzed. However, due to passenger travel, different travel needs and times lead to a certain degree of volatility of OD volume. Analysis of the stability of OD passenger flow can judge the correlation between historical same-time passenger flow and current OD passenger flow.

Select the smart card data (2014.3.1-2014.3.14) for two consecutive weeks of Beijing Urban Rail Transit in 2014 to carry out statistical analysis, and establish relevant indicators to analyze the change law of OD passenger flow distribution structure. In this paper, the weighted relative error is selected as a quantitative indicator, which is used to analyze the stability characteristics of the OD passenger flow distribution. The calculation formula is shown in (1)

\[
 RE(t) = \frac{\sum_{i=1}^{N} \sum_{j=1, j\neq i}^{N} \left( \gamma_{ij}(t) \times \frac{q_{ij}(t) - q^{*}_{ij}(t)}{q_{ij}(t)} \right)}{\sum_{i=1}^{N} \sum_{j=1}^{N} q^{*}_{ij}(t)} \times 100\% \quad \forall t \in K
\]

(1)

\( q^{*}_{ij} \) represents historical OD flow; \( q_{ij} \) represents OD flow of the day; \( K \) represents analysis period; \( \gamma_{ij}(t) \) represents the weight coefficient for any OD pair.

The analysis results are shown in Fig. 5 below:

Fig. 5 (a-e) indicated, the trend of the stability of OD passenger flow on weekdays is similar, the weighted relative error of the morning and evening peak is between 15%-30%, the weighted relative error of the off-peak period is between 40%-68%, and the characteristics of OD distribution working days have strong similarity; as can be seen from Fig. 5 (f) to Fig. 5 (g), there is no obvious error in the distribution of OD passenger flow at weekends and the weighted relative error is large throughout the day.

Moreover, the daily statistical results show that with the increase of statistical time granularity, weighted relative error is decreased. If 60min as the statistical time granularity, the weekday OD passenger flow in the morning and evening peak weighted relative error can be reduced to about 18%, and off-peak weighted relative error can be reduced to about 49%.

### 4) PASSENGER FLOW FACTORS FOR A NUMBER OF TIME PERIODS BEFORE THE FORECAST POINT

The OD passenger flow at the current moment is not only affected by the historical OD passenger flow data, but also

| Period (day 1) | Period (day 2) | Period (day 3) |
|---------------|---------------|---------------|
| 0:00-6:00     | 0:00-6:00     | 0:00-6:00     |
| 6:00-7:30     | 6:00-7:30     | 6:00-9:30     |
| 7:30-9:30     | 7:30-9:30     | 9:30-17:00    |
| 9:30-17:00    | 9:30-15:30    | 17:00-19:00   |
| 17:00-19:00   | 15:30-17:00   | 19:00-21:30   |
| 19:00-21:00   | 17:00-20:30   | 21:30-24:00   |
| 21:00-22:30   | 20:30-22:30   | -             |
| 22:30-24:00   | 22:30-24:00   | -             |

**TABLE 2. Operating time division results of different characteristic days.**

![FIGURE 5. Weighted relative error of OD in one week.](image-url)
affected by the OD passenger flow of previous periods. In order to make full use of the historical OD passenger flow distribution data, the OD passenger flow distribution data in the first four time periods of the prediction point were also taken as the influencing factor of OD passenger flow distribution. Taking 15 min time granularity as an example, i.e. 15 min is a time period, studying the effect of OD passenger flow before the forecast point 15 min, 30 min, 45 min, and 60 min on the forecast point OD passenger flow, respectively, the scatter plot is drawn. The horizontal coordinate is the forecast point traffic value, and the ordinate is the traffic value for several periods before the forecast point. Take the all-day OD full-time data of the Beijing metro network as an example, the results are shown in Fig. 6.

In this paper, Pearson correlation coefficient method is used to examine the relationship between actual data and forecast points in the four periods. Pearson correlation coefficient calculation is shown in (3):

$$r = \frac{1}{n-1} \sum_{i=1}^{n} \left( \frac{X_i - \bar{X}}{s_X} \right) \left( \frac{Y_i - \bar{Y}}{s_Y} \right)$$ (3)

In the formula:
- $X_i$ — OD passenger flow;
- $s_X$ — variance of OD passenger flow;
- $\bar{X}$ — the average number of OD passenger flow;
- $Y_i$ — influence factors;
- $s_Y$ — variance of influence factors;
- $\bar{Y}$ — the average number of influence factors.

The value range of $r$ is $[-1, 1]$; if $r < 0$, it indicates a negative correlation between the two variables; if $r > 0$, it indicates a positive correlation between the two variables; the bigger the $|r|$ is, the stronger the correlation between two variables is.

The correlation coefficient between the forecast point and four historical data of OD passenger flow distribution is shown in Table 3. And the values of the Pearson correlation coefficient are within the 99% confidence interval. It can be seen that there is a clear correlation between the forecast point data and the previous period data, and the closer the period of time to the forecast point, the higher the correlation between the actual passenger flow and the prediction.

| Variables                              | Person correlation | Significance (bilateral) |
|----------------------------------------|--------------------|--------------------------|
| OD passenger flow before the forecast point 15 min | 0.953**            | 0.000                    |
| OD passenger flow before the forecast point 30 min | 0.886**            | 0.000                    |
| OD passenger flow before the forecast point 45 min | 0.783**            | 0.000                    |
| OD passenger flow before the forecast point 60 min | 0.672**            | 0.000                    |

Notes: ** Significant correlation on 0.01 level (bilateral); * Significant correlation on 0.05 level (bilateral); Significance (bilateral) is the test value of the differential test.

5) INFLUENCE FACTORS OF LINE PROPERTIES AT THE ORIGIN AND DESTINATION STATIONS

In rail transit system, large passenger flow zones are usually located in urban, while small passenger flow zones are usually located in suburbs. Different types of operating lines have different passenger flow characteristics, so OD passenger flow distribution may be affected by the line properties of the origin and destination stations.

The inbound, transfer and passenger volume of rail transit lines are related to the functional positioning of the line itself. The attributes of the line and the development of land around the line, and the daily average of the station inbound, transfer and passenger volume of each line from February 15 to February 28, 2014 are calculated, as shown in Fig. 7.

According to Fig. 7, line 7, line 8, line 9, line 14, line 15, YIZHUANG line, CHANGPING line, FANGSHAN line daily passenger flow is significantly lower. According to different conditions involves large passenger flow, or with many network points (transfer station), or cross the city center area, the lines conformed one of the above conditions can be considered as urban line otherwise is considered as suburban lines, and the results are shown in Table 4.

Select the average all-day OD passenger flow data from February 15 to February 28, 2014, and statistically analyze the different OD pairs of the origin station in the line type and OD volume. Different types of line between the average OD volume and different types of OD proportion can be shown in Fig. 8.
As can be seen that OD passenger flow of which the origin station is the urban line accounted for 91% of the total volume, and the OD between the urban line and suburban line, and that between the suburban lines, is significantly less.

6) THE TYPE FACTORS OF THE ORIGIN AND DESTINATION STATIONS
In addition to the type of line, the different passenger flow characteristics of rail transit stations and the nature of surrounding land use will also lead to different trends in passenger flow in and out of the station [44]. Therefore, the classification of stations according to passenger flow characteristics and land use is considered and the impact of different origin station types on the OD passenger flow distribution is analyzed.

This paper takes the 15min granularity of passenger in and out of the station on the working day of Beijing from February 15 to March 14, 2014 as the data source, and obtains the average number of in and out passenger flow of each station in 20 working days. With Z-score standardized based on passenger flow characteristics, processed passenger time series are used K-means to cluster. According to the characteristics of the time distribution of passenger flow in and out of the station at different times of the day, 233 stations can be summed up as the following 7 types of stations, namely, residential station, workplace station, residential work mixed station, mixed-bias residential station, mixed-bias work station, shopping and sightseeing station and external hub-type station[45]. The passenger flow time curve for the number of stations in and out of various stations is shown in Fig.9.

The classification results are shown in Fig.10. It can be seen that the station surrounding land types have significant differences, the perimeter is mainly residential-based and the center is mainly employment-oriented.

7) INFLUENCE FACTORS OF THE INBOUND VOLUME OF THE ORIGIN STATION AND OUTBOUND VOLUME OF DESTINATION STATION
The inbound and outbound volume of stations can effectively reflect the passenger flow in the rail transit operation network. When the network passenger flow increases, the OD volume also theoretically increases, and vice versa. The inbound volume at a station at a certain time is all the OD volume.
TABLE 5. Pearson correlation coefficient of OD flow and inbound and outbound volume.

| Variables                        | Pearson correlation | Significance (bilateral) |
|----------------------------------|---------------------|--------------------------|
| The inbound volume of the origin station | 0.332**             | 0.000                    |
| The outbound volume of the destination station | 0.664**             | 0.000                    |

volume sum of this period with this station as the O, so is the outbound volume. Thus, the meaning of a single OD and the inbound and outbound volume is not the same. In order to simplify the calculation and avoid the impact of extreme OD on the correlation, we randomly select 100 ODs from the first 50% of the OD of the passenger flow and use the Z-score standardized method to process the data.

It can be seen that the correlation between OD volume and the outbound volume of the terminal is high, whereas relatively low with the outbound volume at the origin station during the morning peak from Monday to Friday; at Monday to Friday night peak time, the correlation between OD volume and the passenger inbound origin station and terminal outbound volume is high; at weekends, the correlation is relatively low.

The average full-day OD data from February 15 to February 28, 2014, is selected to analyze the correlation between OD volume and origin and destination stations attraction, as shown in Table 5. It is known that the inspection period is satisfied with $\text{sig} < 0.01$. Through the significance test, there is a certain correlation between OD quantity and origin and destination stations attraction, and the attraction of the terminal station has a relatively large influence on the OD volume.

8) INFLUENCE FACTORS OF AVERAGE TRAVEL TIME BETWEEN ODS

In addition, passenger travel can usually be divided into travel with explicit destination (e.g. work, etc.) and travel with vague destinations (e.g. shopping, etc.) [46]. For destination-defined travel, people tend to choose places where they live with relatively short commute times, resulting in more passenger travel between OD pairs with relatively short travel times. Passengers often choose a relatively close shopping center for travel when the destination is unclear. This will result in a relatively short travel time for OD to more passengers. Considering both of these situations, there is a negative correlation between OD volume and travel time.

Passenger travel time between ODs is from swipe card inbound station to swipe outbound station. By counting and analyzing the relevant smart card data, the travel time distribution of any pair of OD pairs of passenger travel can be seen. Using the data, the average travel time between ODs is used to count the travel time of all OD passengers.

The number of OD intervals, transfer time and average travel time data are calculated and collected, with a total of 54,056 OD. Considering that the number of OD intervals has a strong correlation with the average travel time between OD and OD, the average travel time between ODs as the influencing factor is used for analysis. The average travel time of all OD for a week in a row is shown in Fig. 12.

The travel time of passengers taking the Beijing subway is mainly distributed within 20min to 45min, accounting for 41.37% of the total sample. The travel time exceeds 60min is less, only 15.24%, and the travel time of few passengers exceeds 120min. 45min subway system commute time means a one-way commute from home to work is about 60min. After seven years of observation by Huang et al. [47], the 45min subway system’s commute time can be considered the longest commute that Beijing residents can tolerate: if the subway system commute is less than 45min, residents tend to extend the commute to get better employment opportunities or better living conditions; if the subway system commute is more than 45min, residents will prefer to move in order to shorten the commute.

In order to study the relationship between OD and travel time, the average OD of each travel time period at 3min
intervals is calculated. The calculation results for the three characteristic days are shown below:

According to the statistical results of the three characteristic days, there is a clear correlation between OD volume and the average travel time between OD. In most range, the OD volume decreases with the increase in OD average travel time.

9) INFLUENCE FACTORS OF WEATHER INFORMATION
In addition to smart card data, some external factors can have a certain impact on OD passenger flow distribution.

Environmental factors such as weather condition, average temperature and air quality can have a certain impact on people’s travel. Rainy weather, for example, can lead to an increase in the number of passengers taking subway, which can lead to longer morning and evening rush hours when it occurs during morning and evening peak hours. And air quality can also affect some OD travel, for instance, when air quality is good, people are more inclined to outdoor activities [40]. Therefore, this paper considers the relationship between data analysis and OD volume in terms of weather conditions and air quality.

Usually for residents in the daily travel, the air quality data are more sensitive to the quality rating, AQI index and PM2.5 index, and these three sets of data are strongly related.

Considering the convenience of following feature construction, this paper selects the AQI index to depict the air quality data. This paper selects the highest temperature, the lowest temperature, the weather conditions, AQI index and wind for analysis. The time series of weather data is shown in Fig. 14.

It can be seen that the correlation between the highest temperature and the lowest temperature is very strong, so the daily average temperature is depicted; the correlation between AQI index and wind is strong, when the wind is greater than level 3, the AQI index drops, the air quality becomes better, which is also in line with our daily cognition. So, the wind factors must be considered, and AQI index and wind only select AQI index. Therefore, this paper selects the average temperature, weather conditions, AQI index of these three sets of data for the impact analysis of the external environment of the following on OD passenger flow.

The Pearson correlation is calculated by the processed sum of every OD passenger data and the processed weather data from February 15 to April 4, and the list of relevant coefficients is shown in Table 6.

The Pearson relationship value of each influence factor and OD volume is within the 95% confidence interval, indicating that there is a certain correlation between OD quantity and weather factors, but due to the low correlation coefficient, it is needed to divide OD and date to analyze on which characteristic days the weather factors mainly affect which types of OD conditions.

Weather factors such as weather condition, average temperature and air quality can have a certain impact on people’s travel. Generally speaking, weather factors have a greater impact on the travel of leisure groups such as shopping and sightseeing groups, but less on the travel of commuters.

\[ \text{Variables} \quad \text{Person correlation} \quad \text{Significance (bilateral)} \]

| Variables      | Pearson correlation | Significance (bilateral) |
|----------------|---------------------|--------------------------|
| Average temperature | 0.065**            | 0.000                    |
| Weather condition   | 0.041*             | 0.018                    |
| AQI                  | -0.237**           | 0.000                    |

Notes: ** Significant correlation on 0.01 level (bilateral); * Significant correlation on 0.05 level (bilateral); Significance (bilateral) is the test value of the differential test
Therefore, according to the types of travel terminal, the travel purposes of OD are divided. The stations above are divided into commuting, shopping-sightseeing and business as new-type 1, 2 and 3.

All data is standardized for comparison, and weather conditions are numerically represented. The smaller the weather value is, the better weather conditions are. The results of the relationship between OD flow and weather factors under different conditions are analyzed according to the daily passenger flow characteristic days and OD travel purposes, and the results are shown in Fig. 15.

It can be seen that commuting and business groups are basically not affected by external environmental factors during the three characteristic days, because commuter or out-of-office passenger travel is rigid demand, less affected by external factors are. On characteristic day 2, it can be seen that the effect of temperature on shopping-sightseeing group is not obvious, but AQI and weather conditions have some effect on shopping-sightseeing groups, this is because on Fridays, leisure entertainment such as shopping and dinner after work is generally preferred after work on Fridays; external environmental factors can be considered to have a certain impact on OD of shopping-sightseeing, and it can be seen that the influence of AQI is greater among the three attributes, which is consistent with the results of Pearson-related analysis. Therefore, the final choice of AQI is chosen as the characteristics of shopping-sightseeing passenger flow, temperature, and weather are chosen as the characteristics of passenger flow of shopping-sightseeing under characteristic day 2 and 3. Because the interval of the collected weather data is twice a day (8:00-18:00, 18:00-8:00). It is different from the granularity of the forecast, so it is necessary to process the interval of weather data into the granularity of the predicted. The weather data in the interval should be consistent with the information collection time period. Such as 8:00-8:15 and 16:00-16:15 are the values of the original data 8:00-18:00; 18:00-8:15 and 6:00-6:15 are the values of the original data 18:00-8:00.

10) STATION ATTRACTION RATING INFLUENCE FACTORS BASED ON MOBILE PHONE DATA
In addition, in order to avoid queuing outside the station, some of passengers often choose to avoid those stations with ridership and choose other stations with similar distance, and this just explains why the weekend inbound volume and OD volume correlation is not high above mentioned. Therefore, it is necessary to modify the passenger flow caused by this behavior in combination with other source data. Mobile phone data can depict the density of people outside the station and be used to express the attraction of the station. These effect of chooses of OD passenger flow can be characterized by analyzing destination surrounding personnel intensive.

The mobile phone data is divided, calculate the number of users within 500m of each station within each 15min granularity, and get the attraction value, which can reflect the attractiveness of each subway station at different times, as shown in Fig. 16.

According to the calculation, the “Heat” level of the station is divided. Taking 8:00-8:15 as an example, the station attraction level heat based on mobile signaling data is shown in Fig. 17.

Due to data limitations, there is only September 2, 2010 mobile phone data with no matching passenger flow data, so September 2, 2013 OD data and mobile phone data are selected for related analysis. The relationship between the
station’s attractiveness level and OD correlation is shown in Fig. 18.

Due to data limitations, the correlation did not achieve the expectation, however, from the perspective of personnel distribution space, the changes of distribution in urban have been relatively small within three years. [48]. Considering that can be used as a supplement to smart card data, and avoiding neglecting the specific characteristics of the passenger, this paper adds the mobile phone data as a predictor into the forecast variable set.

C. DATA STANDARDIZATION

The data collected in the preceding mentioned on the short-term OD passenger flow prediction data have different volume and volume units, which will affect the prediction effect of the short-term OD prediction model in the later paragraph. In order to avoid the impact of the volume outline between different data, the collected data need to be standardized, to solve the problem of comparability between different volume data. In addition, standardizing the model input data can improve the accuracy of the short-term OD prediction model and improve the convergence speed of the prediction model.

In this paper, the z-score standardized method is applied to process the data. The calculation formula is:

\[ x' = \frac{x - \mu}{\sigma} \]  

where \( x' \) represents standardized processed data, \( x \) represents raw data, \( \mu \) represents the mean of all sample data, and \( \sigma \) represents the standard deviation of all sample data.

D. INPUT FEATURE SORTING

The values of influencing factors are divided into numeric type, category type, two-value type and other types, different types of variables have different processing methods. This paper standardizes all numerical variables with Z-score, and code all classification characteristics with One-hot. Then All data is organized in 15min, 30min, and 60min time granularity and matched with the passenger data. Variable selection is performed using recursive feature elimination, and logistic regression is used as the base model. At first, different variable combinations are generated and evaluated. Then multiple rounds of training are conducted by using the base model. After each round of training, the feature of multiple weight coefficients is eliminated, and the next round of training is conducted based on the new variable subset until the final result is selected. The final selection results of the predictive features in this paper are shown in Table 7, and 31 features are finally obtained.

The spatial information contains the origin station, passing station and destination station between different ODs. Each OD is different. In other words, different information should contain spatial information unique to each OD. So, features with different values between different OD are classified as space-dependent features, and those with the same value are classified as time-dependent features.

IV. MODEL

The OD passenger flow distribution of rail transit system has timing smoothness, periodicity and trend, and then the short-term OD passenger flow forecast needs to make effective use of the OD passenger flow distribution information in the pre-sequence period. LSTM is considered to perform well with time series prediction problems such as OD passenger flow prediction. However, in view of the spatio-temporal attributes of input features, it is necessary to adjust the structure of LSTM network to exploit hidden spatio-temporal information. Therefore, based on the standard LSTM network, this paper construct STLSTM model by improving the hidden layer structure and neuron structure of the standard LSTM network.

A. RECURRENT NEURAL NETWORK

LSTM is a kind of time recurrent neural network (RNN), which is specially designed to solve the long-term

| Category        | Feature                                      | Description |
|-----------------|----------------------------------------------|-------------|
| Time-dependent  | Characteristic Day 1,2,3                     | [0, 1]      |
| feature         | Operating period 1,2,3,4,5                   | [0, 1]      |
|                 | Average temperature of the day              | Value       |
|                 | Weather 1,2,3,4                            | [0, 1]      |
|                 | AOI                                          | Value       |
| Space-dependent | OD flux before the forecast point           | Value       |
| feature         | 15min,30min,45min,60min                    | Value       |
|                 | OD volume in the same-time of previous week | Value       |
|                 | first two weeks                             | Value       |
|                 | Line properties at the origin and destination station | [0, 1] |
|                 | The type of origin and destination station  | Value       |
|                 | The outbound volume of destination station  | Value       |
|                 | Average travel time between OD             | Value       |
|                 | Station attraction level 1,2,3,4,5,6,7     | [0, 1]      |

FIGURE 18. Relationship between attraction level of destination station and OD flow.
dependence problem of general RNN. In other words, the LSTM network is essentially an RNN network using LSTM units.

A basic RNN model is shown in Fig. 19(a). It can be seen that compared with other neural network models, the nodes of the RNN hidden layer are connected to each other, and the input of the hidden layer includes not only the output of the input layer but also the output of the previous time hidden layer. And these self-connected hidden layers that span the time points enable the RNN to make explicit modeling of time. As can be seen from Fig. 19(b), the characteristics of the chain reveal that RNN is essentially related to the sequence, and its output value is affected by the previous input values. So RNN is the most natural neural network architecture for OD passenger flow distribution time series data.

However, because the chain-based method in the process of RNN network piloting forms the form of multiplication, the error accumulates in the update which leads to gradient explosion and gradient disappearance easily [49].

B. NEURON STRUCTURE IMPROVEMENT

To overcome the aforementioned disadvantages of traditional RNNs, and fully exploit the spatio-temporal attributes of multisource data, STLSTM is proposed in this paper to predict OD passenger flow prediction. These features are especially desirable for passenger prediction in the transportation domain. A STLSTM is composed of input layer, hidden layer and output layer. The improvement of the RNN by using STLSTM is by adding time state $C_{time}$ and space state $C_{space}$ to save the long-term state. Three control switches are used to control $C_{time}$ and $C_{space}$. Two channels are constructed for independent extraction of spatio-temporal characteristics of multisource data. Among them, the control methods of $C_{time}$ and $C_{space}$ are consistent, so this paper takes $C_{time}$ as an example to explain.

The switch of the STLSTM network is implemented through the gate. The gate is actually a fully connected layer, and its input is a vector, and the output is a real vector between 0 and 1. The use of gate is to multiply the output vector by the element multiplied by the vector that needs to be controlled. Two gates are used in STLSTM to control $C_{time}$. One is the forgotten gate which determines how much of the unit state of the previous moment $C_{time,t-1}$ is retained to the current moment $C_{time,t}$; the other is input gate, which determines how much input $X_{time,t}$ from the network is saved to $C_{time,t}$ at the current time. The output gate is used to control how much $C_{time,t}$ is output to the current output value $H_{time,t}$.

1) FORGET GATE $F_{time}$

Identify information that should be forgotten by neurons. The $F_{time,t}$ reads $H_{time,t-1}$ and $X_{time,t}$, and finally assigns the value of the output to $C_{time,t-1}$. The output value is between [0, 1]. 0 means ‘completely discarded’ and 1 means ‘completely reserved’. Its calculation is shown in (5).

$$F_{time,t} = \text{sigmoid} \left( W_f \cdot \left[ X_{time,t}, H_{time,t-1} \right] + b_f \right) \quad (5)$$

From (5) to (10), $W_f, W_i, W_o$ represents the corresponding weight, $b_f, b_i, b_o$ represents the corresponding paranoia, $\text{sigmoid}$ and ReLU represents activation functions.

2) INPUT GATE $I_{time}$

$I_{time}$ determines which new information is saved into the $C_{time}$. The $I_{time}$ consists of two modules, one using the sigmoid layer to determine the value to be entered, and the other using the ReLU layer to determine a new candidate vector, which is then added to the new $C_{time,t}$. The calculation is as follows:

$$I_{time,t} = \text{sigmoid} \left( W_i \cdot \left[ X_{time,t}, H_{time,t-1} \right] + b_i \right) \quad (6)$$

The new contribution to neuronal state is:

$$\tilde{C}_{time,t} = \text{relu} \left( W_c \cdot \left[ X_{time,t}, H_{time,t-1} \right] + b_c \right) \quad (7)$$

The two parts of the $I_{time}$ will update $C_{time,t}$, $C_{time,t-1}$ will be updated to $C_{time,t}$. $F_{time,t}, \tilde{C}_{time,t}$ indicates the resulting information that needs to be discarded, and then adds it to $I_{time,t} \cdot \tilde{C}_{time,t}$. That is:

$$C_{time,t} = I_{time,t} \cdot \tilde{C}_{time,t} + F_{time,t} \circ C_{time,t-1} \quad (8)$$

3) OUTPUT GATE $O_{time}$

$O_{time}$ determines the value to be output. This output is based on $C_{time,t}$ and is filtered through a filter. 1) The sigmoid function is used to determine the part of $C_{time,t}$ that needs to be output; 2) $C_{time,t}$ is processed by using the ReLU layer (the purpose is to keep the output value between the values of the -1 and 1); 3) multiply the output of the sigmoid gate to determine the final output. The specific calculations are:

$$O_{time,t} = \text{sigmoid} \left( W_o \cdot \left[ X_{time,t}, H_{time,t-1} \right] + b_o \right) \quad (9)$$

$$H_{time,t} = O_{time,t} \cdot \text{relu} \left( C_{time,t} \right) \quad (10)$$
The improved neuronal structure is shown in Fig. 21. The improved neurons always maintain two channels to model the OD passenger flow distribution time series, the time-dependent feature channels in the sequence (blue line channels in Fig. 20) and the spatial dependent feature channels (green line channels in Fig. 20).

Similarly, the control of $C_{\text{space}}$ is as follows:

\[
\begin{align*}
X'_{\text{space,} t} & = X_{\text{space,} t} \cdot S \\
I_{\text{space,} t} & = \text{sigmoid} \left( W'_{i} \cdot \left[ X'_{\text{space,} t}, H_{\text{space,} t-1} \right] + b'_{i} \right) \\
F_{\text{space,} t} & = \text{sigmoid} \left( W'_{f} \cdot \left[ X'_{\text{space,} t}, H_{\text{space,} t-1} \right] + b'_{f} \right) \\
O_{\text{space,} t} & = \text{sigmoid} \left( W'_{o} \cdot \left[ X'_{\text{space,} t}, H_{\text{space,} t-1} \right] + b'_{o} \right) \\
\tilde{C}_{\text{space,} t} & = \text{relu} \left( W'_{c} \cdot \left[ X'_{\text{space,} t}, H_{\text{space,} t-1} \right] + b'_{c} \right) \\
C_{\text{space,} t} & = I_{\text{space,} t} \cdot \tilde{C}_{\text{space,} t} + F_{\text{space,} t} \cdot C_{\text{space,} t-1} \\
H_{\text{space,} t} & = O_{\text{space,} t} \cdot \text{relu} \left( C_{\text{space,} t} \right) \\
H_{t} & = [H_{\text{time,} t}, H_{\text{space,} t}] \\
C_{t} & = [C_{\text{time,} t}, C_{\text{space,} t}] 
\end{align*}
\]

\[W'_{i}, b'_{i} \text{ are learning parameters, } l \text{ represents the number of layers, when } X_{\text{time,} t}^{l-1} = H_{\text{time,} t}^{l-1} \in R^{d \times N}, X_{\text{space,} t}^{l} = H_{\text{space,} t}^{l} \in R^{l \times N}, l = 1, S \text{ is entered into the model as a priori knowledge, and its initialization can be based on the distance between ODs, whether each OD pair contains the same area and other information. The input gate, the output gate and the forgotten gate still use sigmoid as the activation function.}\]

**C. FUSION SPACE DEPENDENCY AND TIME DEPENDENCY**

The fusion module will combine the two-channel results of the output of the improved LSTM neurons in space-time, with a structure shown in Fig. 21.

As shown in the (12), the parameters and the characteristics of the time dimension and space dimension are firstly weighted and summed. Then the fusion features are flattened. Finally, the space-time features are further abstracted and extracted with a fully connected network as shown in the (13).

\[X_{\text{fusion,} t} = \Theta_{1} \circ H_{\text{time,} t} + \Theta_{2} \circ H_{\text{space,} t} \]

\[x_{\text{seq,} t} = \text{relu} \left( W_{\text{seq}} \cdot X_{\text{fusion,} t} + b_{\text{seq}} \right) \]

where $x'_{\text{seq,} t} \in R^{d' \times N}$ is flattened by $X_{\text{seq,} t} \in R^{d' \times N}$, $\Theta_{1}, \Theta_{2} \in R^{d' \times 1}, W_{\text{seq}, t} \in R^{d' \times N_{d}}, b_{\text{seq}, t} \in R^{d}$ represents learning parameters. The module also uses ReLU as an activation function, with an output of $x_{\text{seq,} t}$.

In each time slice, $x_{\text{seq,} t}$, which combines time and spatial features, is extracted again through a fully connected layer. Finally, through the tanh function activation, the output value is mapped $[-1, 1]$ to get the final prediction $\hat{X}_{t+k}$. The calculation formula is:

\[\hat{X}_{t+k} = \text{tanh} \left( W_{\text{seq}, t} + b \right) \]
D. Li et al.: Spatio-Temporal Structured LSTM Model for Short-Term Prediction of Origin-Destination Matrix

FIGURE 22. Framework of OD flow prediction model based on STLSTM under multisource data.

\[ W \in R^{N \times a_{seq}}, b \in R^{N}, b_{seq,t} \in R^{u} \] represents learning parameters.

D. MODEL FRAMEWORK BASED ON IMPROVING SPACE-TIME LSTM NETWORK UNDER MULTISOURCE DATA

Combined with the neuronal structure and the spatial-time dependence fusion method, this paper proposes STLSTM model, explicitly model the spatial-time correlation of the demand sequence, so as to predict the OD passenger flow distribution more accurately. The STLSTM model is shown in Fig. 22. Compared to the traditional model, the new LSTM neuron structure redesign the hidden layer structure, and the integration layer and the full connection layer, after adding two channels of STLSTM neuron layer.

V. CASE STUDY

A. CASE DESCRIPTIONS

In this paper, the Beijing rail transit network is selected as the case to validate the proposed model. The network contains 233 stations and 54,056 ODs. Based on the data demand of the short-term OD passenger flow forecast research on the rail transit network, this paper collects the original records of smart card of the whole network for 7 consecutive weeks through the Beijing Rail Transit system as the basis of this study. Considering that February 14, 2014 is the Lantern Festival, April 5, 2014 is Qingming Festival, the study period was further identified as February 15, 2014 to April 4, 2014, thus avoiding the impact of holidays on OD passenger flow in the rail network. At the same time, the corresponding weather data and mobile signaling data were collected to match the OD demand data. All the data are settled in 15min, 30min and 60min time granularity. Among which, the data for six consecutive weeks from February 15, 2014 to March 28, 2014 are used as training data set for the analysis of the characteristics of the OD passenger flow distribution of the network and the construction of the short-term OD passenger flow prediction model. The data for the week from March 29, 2014 to April 4, 2014 are used as test data set to evaluate the predictive effect of the short-term OD passenger flow prediction model at different times. The experiment is implemented on a personal computer with Intel i7 2.5GHZ CPU and 8 GB memory.

B. MODEL PARAMETER SETTING

The parameters of the network connection, the number of layers, the number of neurons per layer, etc. in the prediction model based on LSTM network are not generated by machine learning, but need to be set in advance by human beings. These parameters are called Hyper-Parameters. Next, some of the hyper-parameters of the model are tested in this paper, and the optimal parameter results are selected.

In order to analyze the effects of the number of neurons and epoch in the prediction model on the accuracy of the prediction model, this paper sets the value range of the number of neurons is \( N = \{50, 100, 150\} \), and epoch takes 30, 50, 70, 90, and the loss function takes MAE or MAPE. Combine each neuron number, epoch, and loss function, and compare their prediction result to select the optimal parameters. In the combined prediction results of each neuron number, epoch, and loss function, the minimum values of MAE and MAPE are respectively 1.31161 and 0.42287 when the number of neurons is 100, and epoch is 70, achieving a better prediction effect. By comparison with loss functions, the prediction effect is better when the loss function takes MAPE. In other experimental tests in this paper, the number of neurons is set to 100 and the loss function as MAPE. In order to avoid overfitting, it is necessary to set dropout [50]. To test the effect of dropout parameter settings on experimental prediction errors, this paper builds a prediction model with dropout values of 0, 0.1, 0.3, and 0.5, respectively, and enters epoch ranging from 30-90.

From Fig. 23 to 24, it can be concluded that using dropout prediction model is less perused in the MAE and MAPE values under different epoch conditions than without dropout prediction model. This is due to the fact that in the training phase, the hidden layer neurons appear randomly at a set probability at each iteration when the weight matrix adjustment is made using the data of the input model. This way...
of updating the weight matrix no longer depends on the joint action of the neurons with the hidden layer of complex relationship, enhances the learning ability of the short-term OD prediction model, avoids the problem that some OD passenger flow distribution time series data features only have effect under certain circumstances. Thus, improving the generalization ability of the short-term OD prediction model. Dropout value are set as 0.1, 0.3 and 0.5, the OD prediction model error differences are relatively small, but overall epoch as 70, dropout value as 0.5, the prediction model perform slightly better. Therefore, the value of dropout is chosen to be 0.5 in the model.

C. RESULTS

1) ANALYSIS OF THE OVERALL PREDICTION PERFORMANCE OF DIFFERENT CHARACTERISTIC DAYS

The three predictions at 15min, 30min and 60min time granularity are obtained by constructing a short-term passenger flow forecast model based on STLSTM under the multisource data to predict the full week of OD passenger flow distribution from March 29, 2014 to April 4, 2014. Take the period from 8:00 a.m. to 8:15 for example, the forecast results are illustrated, and the OD passenger flow distribution results of different characteristic days are shown in Fig. 25 (b-d) below. In the figure, the black dot represents the subway station, the different colors of inter-lines between OD represents the OD passenger flow of different sizes, the greater the passenger flow is, the closer the color is to the red, otherwise, the closer the color is to the blue, the less the OD with passenger flow of 0 does not show.

It can be seen that passenger flow distributions in the morning 8:00-8:15 of characteristic day 1 and 2 are very similar, but they have large differences with day 3. Passenger flow of day 3 is significantly smaller than the other two days. Based on the forecast results, the predicted accuracy results for calculating the days of different passenger flow characteristics with different time granularity are shown in Fig. 25 (a):

Analysis of the model’s overall prediction accuracy results found that the MAPE of 15min interval are all below 45%, MAPE of 30min time interval are below 40%, MAPE of 60min time interval are below 35%, and in the ordinary working day (characteristic day 1), the MAPE of the 60min interval reached 22.20%. In addition, the predicted MAE value under the granularity size of each time was basically below 1.5, of which the prediction effect is relatively better to the Beijing subway network passenger flow level. It can be seen that the short-term passenger flow prediction model based on STLSTM under multisource data has good overall prediction effect and can be used as a reference when the operator of rail transit system formulates the transportation organization strategy.

2) ANALYSIS OF PREDICTION EFFECTS IN DIFFERENT TIME PERIODS

Through the above analysis, it is found that the overall prediction accuracy of the STLSTM under multisource data is high. A specific analysis of the precision of the time-sharing prediction which comparing the prediction results of different characteristic day in different time periods is shown in Fig 26.

As can be seen from Fig. 26:(1) On the whole, the error of characteristic 1 day is basically less than the other two characteristic days, indicating that the characteristic daily 1 passenger flow regularity is relatively strong. (2) Characteristic day 2 is similar to the prediction effect of characteristic day 1 in each time of the morning. However, from about 12:00 noon, the OD passenger flow prediction error of the characteristic day 2 is significantly larger than that of characteristic day1, because it is affected by weekend travel.(3) The MAE value and MAPE value of characteristic day 3 are also relatively small, probably because the passenger flow distribution stability of characteristic day 3 is not high, while the passenger flow is also small. Therefore, the MAE value is lower than that of the passenger flow characteristic day 2, and the MAPE value is also smaller because there are great number of OD pairs with zero passenger flow.(4) In the forecast results of each period, the forecast effect of
As can be seen from Fig. 27: (1) The MAE of the peak time is basically below 20, the MAPE is basically below 25%. However, because the MAPE value is higher at the point of small passenger flow, the overall MAPE value is larger; (2) The median MAE value is basically below 10, but the MAPE is basically below 70%. This is because when the OD volume is small, a large MAPE value is possibly obtained even though the MAE value is small.

3) ANALYSIS OF THE PREDICTION EFFECT OF DIFFERENT PASSENGER DEMAND TYPES

In order to further illustrate the short-term prediction effect of the model with different pick-up point station type of OD, the OD pair (XI’ERQI-SHAHE) with relatively large OD passenger flow and the OD pair (ZHICHUNLU –YONG’ANLI) with relatively small OD passenger flow are selected from the forecast results, and the prediction results are analyzed. In order to further illustrate the short-term prediction effect of the model for different passenger ODs, the OD pair (XI’ERQI-SHAHE) with relatively large OD volume (top 10% of passenger volume) and OD pair (ZHICHUNLU-YONG’ANLI) with relatively small OD volume (50% after passenger flow) are randomly selected to analyze the forecast results. Fig. 28 is the results of OD passenger flow forecasts for OD pairs (XI’ERQI-SHAHE) and OD pairs (ZHICHUNLU –YONG’ANLI) at 15min forecast intervals, respectively.

Through Fig. 28, the prediction effect of the OD pair (XI’ERQI-SHAHE) with the larger passenger demand and the OD pair (ZHICHUNLU-YONG’ANLI) with the smaller passenger demand for passenger flow can be seen. The MAE of OD pair (XI’ERQI-SHAHE) is 6.902 and the
MAPE is 23.37%; and the MAE of OD pair (ZHICHUNLU-YONG’ANLI) is 0.875 and the MAPE is 31.43%. The results show that the prediction effect of the model is related to the type of OD passenger flow, and when the volume of flow is large and relatively stable, the model can learn the data characteristics better, and the prediction effect is often higher. In contrast, when OD flow is small and relatively unstable, the model learning effect is relatively poor, the prediction effect is general. The results show that the prediction effect of the model is related to OD flow. When there is more OD flow with less stochastic effect, the model performs better in learning and predicts more accurately. This further verifies the model’s extraction of regular information in the data.

4) COMPARATIVE ANALYSIS OF MULTIPLE MODELS AT DIFFERENT TIME GRANULARITY

In order to study whether the short-term OD passenger flow prediction model improves the prediction accuracy, it is necessary to compare with other methods indicating historical average method, such as ARIMA, standard LSTM under single-source data (only enter smart card data), standard LSTM under multisource data, and compare the prediction effect of each model at different time granularity.

The prediction results of 4 method are calculated. The prediction results including the OD passenger flow distribution of a full week from March 29, 2014 to April 4, 2014 under the time granularity of 30min and 60min time. The statistical model has the overall prediction error at different time granularity listed. And the results are compared with the results of the STLSTM. The results are shown in Fig.29.

These summaries can be seen from the results:

(1) In comparison of the prediction results of each model, it can be seen that the STLSTM obtains a better prediction effect. MAE value of STLSTM network is about 3%-13% lower than other methods’, and MAPE value is about 5%-13% lower than other methods’, which shows that the STLSTM is better than the other four methods under multisource data.

(2) The standard LSTM network under single-source data is compared with the historical average method and ARIMA model under single-source data, it is difficult for the model to extract enough features to draw the peak of passenger flow when no other features are collected. Therefore, some OD passenger flow cannot be accurately predicted, so the prediction effect at peak time is poor. Because only smart card data is used, other information that affects passenger travel, especially during peak hours, is poorly predicted for peak passenger volume.

(3) In comparison of the standard LSTM network under the single-source data and the multisource data, it can be seen that by adding multisource data, the model can extract more features, so that the prediction effect of the model under different time granularities is significantly improved. MAE is reduced by about 0.2 and MAPE by about 4%, indicating that the introduction of multisource data is very effective for the improvement of the model.

(4) In comparison of the standard LSTM network and the STLSTM, it can be seen that after improving the model structure, the STLSTM can effective capture the characteristics of each OD in the OD passenger flow distribution time series. That makes the prediction effect of the model significantly improve under different time granularities, and the MAPE reduce by about 6%. It shows that the improvement of the model structure effectively promotes the prediction accuracy.

VI. CONCLUSION

Short term OD matrix prediction of rail transit is a well-known difficult problem due to stochasticity. In order to improve the accuracy of the prediction, STLSTM is proposed to utilize multisource spatiotemporal data to improve the short-term predictability of the dynamic passengers’ origin and destination demand (OD matrix) in a rail transit system. The proposed network extends the standard LSTM by incorporating spatial information available from multisource data. The approach is validated on the dataset obtained from the Beijing subway network. The results show an improved

![Figure 29. Comparison of different models’ prediction results.](image-url)
prediction of the OD matrix over existing predictive models and the standard LSTM.

(1) It is very effective to introduce multisource data as input to improve prediction accuracy. Comparing the prediction effect of the standard LSTM network under the single source data and the multisource data, the prediction effect can be extracted to more characteristics by increasing the multisource data, so that the prediction effect under the granularity of time can be significantly improved.

(2) The results proved that STLSTM model can effectively exploit the potential information of multisource data and improve the prediction effect. Comparing the standard LSTM and STLSTM under multisource data, the prediction effect of STLSTM at each time granularity is significantly improved, and the MAPE is reduced by about 6%, which explained that the improvement of model structure effectively promotes the prediction accuracy.

(3) STLSTM has a good prediction effect in the case of large-scale complex road network. The results of case verification conducted through the whole network of Beijing, which contains 233 stations and 54,056 ODs, can be obtained. In comparison of commonly used prediction methods, the STLSTM network under multisource data has obtained a good prediction effect. The partially stochastic passengers’ activity increases the difficulty of the prediction. However, such ‘stochasticity’ is possible to be capture from the social networking data such as Facebook, WeChat etc. In the future, based on the STLSTM, the social networking site data will be further imputed into model to check whether it can improve the accuracy. At the same time, it is expected to obtain complete data with the same time interval and matching time dimension to improve prediction accuracy. Finally, to reduce calculation time, the more easily calculated GRU would be used to reduce the calculation number of large-scale networks in the future. Moreover, the application ability of the model for some abnormal cases such as sudden large passenger flow burst scenarios will be studied.

REFERENCES

[1] Y. Wang, T. Tang, B. Ning, T. J. J. van den Boom, and B. De Schutter, “Passenger-demands-oriented train scheduling for an urban rail transit network,” Transp. Res. C, Emerg. Technol., vol. 60, pp. 1–23, Nov. 2015.
[2] H. Li, Y. Wang, X. Xu, L. Qin, and H. Zhang, “Short-term passenger flow prediction under passenger flow control using a dynamic radial basis function network,” Appl. Soft Comput., vol. 83, Oct. 2019, Art. no. 105620.
[3] X. Yao, B. Hao, D. Li, and H. Ren, “Simulation-based dynamic passenger flow assignment modelling for a schedule-based transit network,” Discrete Dyn. Nature Soc., vol. 2017, pp. 1–15, Feb. 2017.
[4] M. Ni, Q. He, and J. Gao, “Forecasting the subway passenger flow under event occurrences with social media,” IEEE Trans. Intell. Transp. Syst., vol. 18, no. 6, pp. 1623–1632, Jun. 2017.
[5] J. Zhang, D. Shen, L. Tu, F. Zhang, C. Xu, Y. Wang, C. Tian, X. Li, B. Huang, and Z. Li, “A real-time passenger flow estimation and prediction method for urban bus transit systems,” IEEE Trans. Intell. Transp. Syst., vol. 18, no. 11, pp. 3168–3178, Nov. 2017.
[6] E. I. Vlahogianni, J. C. Golas, and M. G. Karlaftis, “Short-term traffic forecasting: Overview of objectives and methods,” Transp. Rev., vol. 24, no. 5, pp. 533–557, Sep. 2004.
[7] B. L. Smith, B. M. Williams, and R. K. Oswald, “Comparison of parametric and nonparametric models for traffic flow forecasting,” Transp. Res. C, Emerg. Technol., vol. 10, no. 4, pp. 303–321, Aug. 2002.
[8] N. Zhang, F.-Y. Wang, F. Zhu, D. Zhao, and S. Tang, “DynaCAS: Computational experiments and decision support for ITS,” IEEE Intell. Syst., vol. 23, no. 6, pp. 19–23, Nov. 2008.
[9] B. L. Smith and M. J. Demetsky, “Traffic flow forecasting: Comparison of modeling approaches,” J. Transp. Eng., vol. 123, no. 4, pp. 261–266, Jun. 1997.
[10] B. M. Williams, P. K. Durvasula, and D. E. Brown, “Urban freeway traffic flow prediction: Application of seasonal autoregressive integrated moving average and exponential smoothing models,” Transp. Res. Rec., J. Transp. Res. Board, vol. 1644, no. 1, pp. 132–141, Jan. 1998.
[11] B. M. Williams and L. A. Hoel, “Modeling and forecasting vehicular traffic flow as a seasonal ARIMA process: Theoretical basis and empirical results,” J. Transp. Eng., vol. 129, no. 6, pp. 664–672, Nov. 2003.
[12] M. Van Der Vooit, M. Dougherty, and S. Watson, “Combining Kohonen maps with ARIMA time series models to forecast traffic flow,” Transp. Res. C, Emerg. Technol., vol. 4, no. 5, pp. 307-318, 1996.
[13] S. Lee and D. B. Fambro, “Application of subset autoregressive integrated moving average model for short-term freeway traffic volume forecasting,” Transp. Res. Rec., J. Transp. Res. Board, vol. 1678, no. 1, pp. 179–188, Jan. 1999.
[14] X. Chen, S. Guo, L. Yu, and B. Hellinga, “Short-term forecasting of transit route OD matrix with historical data,” in Proc. 14th Int. Conf. IEEE Conf. Intell. Transp. Syst. (ITSC), Oct. 2011, pp. 1513–1518.
[15] S. Clark, “Traffic prediction using multivariate nonparametric regression,” J. Transp. Eng., vol. 129, no. 2, pp. 161–168, Mar. 2003.
[16] M. Dougherty, “A review of neural networks applied to transport,” Transp. Res. C, Emerg. Technol., vol. 3, no. 4, pp. 247–260, Aug. 1995.
[17] I. Okutani and Y. J. Stepahenedes, “Dynamic prediction of traffic volume through Kalman filtering theory,” Transp. Res. B, Methodol., vol. 18, no. 1, pp. 1–11, Feb. 1984.
[18] J. C. van Lint, “Online learning solutions for freeway travel time prediction,” IEEE Trans. Intell. Transp. Syst., vol. 9, no. 1, pp. 38–47, Mar. 2008.
[19] X. Yao, “Theory and method of dynamic passenger flow distribution and coordinated inflow control for urban rail transit network,” Ph.D. dissertation, Dept. Eng., Beijing Jiaotong Univ., Beijing, China, 2014.
[20] X. Yao, P. Zhao, and D. Yu, “Dynamic origin-destination matrix estimation for urban rail transit based on averaging strategy,” J. Jilin Univ., vol. 46, no. 1, pp. 92–99, 2016.
[21] B. L. Smith and M. J. Demetsky, “Multiple-interval freeway traffic flow forecasting,” Transp. Res. Rec., vol. 1554, no. 1554, pp. 136–141, 1996.
[22] M. Castro-Neto, Y.-S. Jeong, M.-K. Jeong, and L. D. Han, “Online-SVR for short-term traffic flow prediction under typical and atypical traffic conditions,” Expert Syst. Appl., vol. 36, no. 3, pp. 6164–6173, Apr. 2009.
[23] Y.-S. Jeong, Y.-J. Byon, M. M. Castro-Neto, and S. M. Easa, “Supervised weighting-online learning algorithm for short-term traffic flow prediction,” IEEE Trans. Intell. Transp. Syst., vol. 14, no. 4, pp. 1700–1707, Dec. 2013.
[24] L. Tang, Y. Zhao, and J. Cabrera, “Forecasting short-term passenger flow: An empirical study on shenzhen metro,” IEEE Trans. Intell. Transp. Syst., vol. 20, no. 10, pp. 3613–3622, Oct. 2019.
[25] B. L. Smith, “Short-term traffic flow prediction: Neural network approach,” Transp. Res. Rec., no. 1453, pp. 98–104, 1994.
[26] W. Hu, Y. Liu, L. Li, and S. Xin, “The short-term traffic flow prediction based on neural network,” Transp. Res. Rec., vol. 1, pp. 293, May 1994.
[27] J. Z. Zhu, J. X. Cao, and Y. Zhu, “Traffic volume forecasting based on radial basis function neural network with the consideration of traffic flows at the adjacent intersections,” Transp. Res. C, Emerg. Technol., vol. 47, pp. 139–154, Oct. 2014.
[28] T. Mikolov, A. Deoras, D. Povey, L. Burget, and J. Cernocky, “Strategies for training large scale neural network language models,” in Proc. IEEE Workshop Autom. Speech Recognit. Understand., Dec. 2011, pp. 196–201.
[29] R. Collobert and J. Weston, “A unified architecture for natural language processing: Deep neural networks with multitask learning,” in Proc. 25th Int. Conf. Mach. Learn. (ICML), Aug. 2011, pp. 160–167.
[30] J. F. Kolen and S. C. Kremer, “Gradient Flow in Recurrent Nets: The Difficulty of Learning Long-Term Dependencies,” Hoboken, NJ, USA: Wiley, 2009.
[31] A. Graves, A.-R. Mohamed, and G. Hinton, “Speech recognition with deep recurrent neural networks,” in Proc. IEEE Int. Conf. Acoust., Speech Signal Process., May 2013, pp. 6645–6649.
[32] 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.

[33] Y. Tian and L. Pan, “Predicting short-term traffic flow by long short-term memory recurrent neural network,” in Proc. IEEE Int. Conf. Smart City/ SocialCom/SustainCom (SmartCity), Dec. 2015, pp. 153–158.

[34] P. Wang, C. Wu, and X. Gao, “Research on passenger flow combination prediction model based on RBF neural networks and LSSVM,” in Proc. Chin. Control Decis. Conf. (CCDC), May 2016, pp. 6064–6068.

[35] L. Li, Y. Wang, G. Zhong, J. Zhang, and B. Ran, “Short-to-medium term passenger flow forecasting for metro stations using a hybrid model,” KSCE J. Civ. Eng., vol. 22, no. 5, pp. 1937–1945, May 2018.

[36] Z. He, B. Wang, J. Huang, and Y. Du, “Station passenger flow forecasting for urban rail transit based on station attributes,” in Proc. IEEE Int. Conf. Cloud Comput. Intell. Syst., Nov. 2014, pp. 410–414.

[37] Z. He, B. Wang, J. Huang, and Y. Du, “Station passenger flow forecast for urban rail transit based on station attributes,” in Proc. IEEE Int. Conf. Cloud Comput. Intell. Syst., Nov. 2014, pp. 410–414.

[38] L. Huang, C. Tao, and W. Yan, “Forecasting daily pedestrian flows in the Tiananmen square based on historical data and weather conditions,” ISCRAM, vol. 18, pp. 1315–1320, May 2015.

[39] Q. Xu and M. Wang, “Dynamic origin-destination extraction and spatial-temporal feature analysis based on cell phone data,” in Proc. Annu. Conf ITS China, 2018, p. 18.

[40] K. Chen, Y. Zhou, and F. Dai, “A LSTM-based method for stock returns prediction: A case study of China stock market,” in Proc. IEEE Int. Conf. Big Data (Big Data), Oct. 2015, pp. 2823–2824.

[41] M. Fionn and L. Pierre, “Ward’s hierarchical clustering method: Clustering criterion and agglomerative algorithm,” J. Classification, vol. 31, no. 3, pp. 274–295, 2014.

[42] Z. He, B. Wang, J. Huang, and Y. Du, “Station passenger flow forecasting for urban rail transit based on station attributes,” in Proc. IEEE Int. Conf. Cloud Comput. Intell. Syst., Nov. 2014, pp. 410–414.

[43] L. Huang, C. Tao, and W. Yan, “Forecasting daily pedestrian flows in the Tiananmen square based on historical data and weather conditions,” ISCRAM, vol. 18, pp. 1315–1320, May 2015.

[44] Q. Xu and M. Wang, “Dynamic origin-destination extraction and spatial-temporal feature analysis based on cell phone data,” in Proc. Annu. Conf ITS China, 2018, p. 18.

[45] K. Chen, Y. Zhou, and F. Dai, “A LSTM-based method for stock returns prediction: A case study of China stock market,” in Proc. IEEE Int. Conf. Big Data (Big Data), Oct. 2015, pp. 2823–2824.

[46] M. Fionn and L. Pierre, “Ward’s hierarchical clustering method: Clustering criterion and agglomerative algorithm,” J. Classification, vol. 31, no. 3, pp. 274–295, 2014.

[47] Z. He, B. Wang, J. Huang, and Y. Du, “Station passenger flow forecasting for urban rail transit based on station attributes,” in Proc. IEEE Int. Conf. Cloud Comput. Intell. Syst., Nov. 2014, pp. 410–414.

[48] L. Huang, C. Tao, and W. Yan, “Forecasting daily pedestrian flows in the Tiananmen square based on historical data and weather conditions,” ISCRAM, vol. 18, pp. 1315–1320, May 2015.

[49] Q. Xu and M. Wang, “Dynamic origin-destination extraction and spatial-temporal feature analysis based on cell phone data,” in Proc. Annu. Conf ITS China, 2018, p. 18.

[50] K. Chen, Y. Zhou, and F. Dai, “A LSTM-based method for stock returns prediction: A case study of China stock market,” in Proc. IEEE Int. Conf. Big Data (Big Data), Oct. 2015, pp. 2823–2824.

[51] M. Fionn and L. Pierre, “Ward’s hierarchical clustering method: Clustering criterion and agglomerative algorithm,” J. Classification, vol. 31, no. 3, pp. 274–295, 2014.

[52] Z. He, B. Wang, J. Huang, and Y. Du, “Station passenger flow forecasting for urban rail transit based on station attributes,” in Proc. IEEE Int. Conf. Cloud Comput. Intell. Syst., Nov. 2014, pp. 410–414.

[53] L. Huang, C. Tao, and W. Yan, “Forecasting daily pedestrian flows in the Tiananmen square based on historical data and weather conditions,” ISCRAM, vol. 18, pp. 1315–1320, May 2015.