

Research Article

Spatiotemporal Virtual Graph Convolution Network for Key Origin-Destination Flow Prediction in Metro System

Jun Yang ^{1,2}, Xiao Han ¹, Tan Ye ¹, Yinghao Tang ¹, Weidong Feng ³, Aili Wang ³,
Huijun Zuo ⁴ and Qiang Zhang ⁴

¹Big Data and Internet of Things Research Center, China University of Mining and Technology, Beijing, China

²Key Laboratory of Intelligent Mining and Robotics, Ministry of Emergency Management, Beijing, China

³SinoRail (Beijing) Network Technology Research Institute Co. Ltd, Beijing, China

⁴Unit 61741 of PLA, Beijing 100094, China

Correspondence should be addressed to Xiao Han; 874326845@qq.com

Received 5 July 2022; Accepted 22 August 2022; Published 19 September 2022

Academic Editor: Li Zhu

Copyright © 2022 Jun Yang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Short-term Origin-Destination (OD) flow prediction plays a major part in the realization of Smart Metro. It can help traffic managers implement dynamic control strategies to improve operation safety. Also, it can assist passengers in making reasonable travel plans to improve the passenger experience. However, there are problems that the dimension of OD short-term traffic prediction is much higher than the base number of metro stations and the OD matrix is sparse. To resolve the above two problems, a threshold-based method is proposed to extract key OD pairs first. OD passenger flow contains the attribute information of the Origin-Destination station and exhibits similar time evolution characteristics, so the spatial and temporal correlation needs to be considered in the prediction. Pearson correlation matrix is used to build a virtual graph and model the virtual connection between OD pairs. A spatiotemporal virtual graph convolutional network (ST-VGCN), which combines the advantages of a graph neural network and gated recurrent neural network, is proposed to identify spatial associations and temporal patterns simultaneously. The proposed method is evaluated on 39 days of real-world data from Shenzhen Metro, which outperforms other benchmarks. The research in this work can contribute to the development of short-term OD flow forecasts and help to provide ideas for the research on real-time operation and management of rail transit. Furthermore, it can help to establish passenger flow prediction and early warning mechanisms to quickly evacuate a large number of passengers in case of emergency.

1. Introduction

In the wake of developments in intelligent cities, the analysis and mining of big traffic data have attracted widespread attention in the field of intelligent transportation. As an important means of transportation to alleviate road traffic congestion in big cities, urban rail transit has gradually entered the era of network operation. At the same time, a large amount of passenger flow data is generated, which makes the research on passenger flow prediction more and more important [1]. The most basic thing is modern technology has greatly guaranteed the safety of train operation [2], ensuring the growing demand for safe travel of residents. Shenzhen Metro, for example, the average travel scale of

which in a single day is up to 5 million person-times, carries more than one-third of the traffic flow of the entire city and occupies an important position in the urban public transportation system. However, the ever-increasing travel demand has brought greater and greater pressure to the operation of urban rail transit. Especially during the morning and evening rush hours, the demand for commuting is very large, so the operating efficiency of the metro directly affects the overall commuting efficiency of the city. In addition to the research on the operation safety of train control systems [3], more and more researchers have paid attention to the passenger flow data generated by metro operations and conducted multilevel analysis [4] on it to ensure the operation safety. Because of the enormous scale of

the current metro network system, passenger travel shows great complexity. Most of the existing research pays more attention to the prediction of inbound and outbound passenger flow, daily ridership, hourly ridership, and sectional passenger flow. OD passenger flow reflects the direction of flow in and out of the station, which is a visual display of passenger travel needs and contains precious information.

Because of the good connectivity between stations in the metro network, the travel demand of passengers is usually expressed by an abstract OD matrix. The elements in the matrix represent the OD passenger flow. As important carriers of the metro passenger flow, they represent the travel times between station O and station D within a specific time interval. OD matrix can well describe the source and destination of the ridership between any two stations, which is the foundation for travel behavior characteristic analysis. The existing information platform establishes a channel of communication between the metro system and passengers to provide passengers with metro network status prediction information and help passengers plan their trips reasonably. At the same time, OD passenger flow prediction combined with real-time detection of passengers in densely crowded areas of metro stations [5] can provide better services for ensuring operational safety. However, the prediction results may differ from the actual situation, especially during peak periods. This inaccuracy disrupts travel plans and creates a sense of mistrust for the information while affecting metro managers to make judgments. Therefore, how to accurately predict the OD passenger flow of the whole network is an important issue for the refined and comprehensive prediction of the rail transit network. By predicting the fine-grained passenger flow and its dynamic changes in the network state, metro managers can timely adjust the operation plan and carry out reasonable passenger flow organization and control [6], which is of great practical significance to ensure traffic safety. Among them, passenger flow organization and control includes, but is not limited to, adjusting schedules, adding services, disseminating information to passengers in response to the condition of demand surge, and incentivizing passengers to delay or change travel time.

The research of metro OD passenger flow based on AFC data mainly includes three aspects, OD matrix, the main passenger flow direction of the origin station, and key OD pairs. Since the OD matrix is sparse and highly skewed, combined with the analysis of the main direction of the origin station, it can be known that the key OD pairs reflect the travel routes of large demand, which largely determines the metro operation status of the network. Meanwhile, they are the main object of metro vehicle allocation which means higher requirements for ensuring efficient and orderly trains [7]. At the same time, the OD pair implicitly includes the information of station O and station D and is regular in time variation like time series. For instance, most of the flow in the morning rush is commuters whose commuting routes are relatively fixed. OD pairs with O as residential areas and D as business areas show a similar regularity. That is, OD pairs show some dependence on the consecutive time intervals and the similar spatial properties of the Origin-

Destination station. Therefore, how to determine the key OD pairs and capture the spatiotemporal dependence between them at the level of metro network becomes a key problem.

In this work, we propose a method to extract and predict flow of key OD pairs in metro network. According to the historical data of the AFC system, we extract the OD matrix and then obtain the key OD pairs through filtering. We introduce a spatiotemporal virtual graph convolutional network (ST-VGCN) to model the spatiotemporal dependencies between key OD pairs. The virtual topology is used to establish a similarity map. The graph convolution combined with GRU is used to extract spatiotemporal features and predict the key OD passenger flow. We carry out an experiment on the real-world dataset from Shenzhen Metro to prove the advantage of ST-VGCN. The main contributions of this study are as follows:

- (1) We obtain OD matrix by processing AFC historical data and extract the key OD pairs that have a large demand for metro travel by setting reasonable thresholds
- (2) We propose a spatiotemporal virtual graph convolutional network (ST-VGCN) model, in which we establish a similarity graph based on the Pearson coefficient as a virtual topology, to capture the spatial and temporal dependencies between key OD pairs
- (3) We have verified our method through a case study of Shenzhen Metro, and the validity of the model was confirmed by comparative tests

The rest of this paper is organized as follows: In Section 2, we review related work on OD traffic forecast. In Section 3, we describe our main work on the key OD flow prediction. In Section 4, we evaluate our method based on real data and present our results and analysis. In Section 5, we summarize our paper and discuss several possible directions for future work.

2. Related Work

Most of the researches concentrate on the prediction of road traffic flow and traffic speed [8]. More and more researchers focus on urban rail transit forecasting, among which the research on OD prediction is relatively less. Considering the similarity between OD sequence data and inbound/outbound data of urban rail transit stations, we will introduce relevant studies on traffic prediction in related work and further summarize the research on OD passenger flow prediction.

Early researchers used methods based on statistical learning, such as Historical Average [1] (HA), Autoregressive Integrated Moving Average [9] (ARIMA), Vector Autoregressive model [10] (VAR), etc. Later, traditional machine learning and neural network methods were applied to traffic flow prediction. Wu et al. [11] studied the travel time prediction problem using support vector regression(SVR). Zhu et al. [8] came up with a linear conditional Gaussian Bayesian network model for short-term traffic flow forecast, which took into account spatiotemporal

characteristics and velocity information. Yang and Hou [12] used a hybrid model based on wavelet analysis and least squares support vector machine to complete short-term rail transit passenger flow prediction. Jiao et al. [13] proposed an improved Kalman filter model based on Bayesian combination and nonparametric regression, in which real-time passenger traffic is deviated from historical data to mitigate the volatility of original data.

Because of the good performance of deep learning models in other fields, many scholars try to use them in traffic prediction. Zhang et al. [14] presented deep-ST, which is the first model that uses convolutional neural network to mine spatial dependence between grids. On this basis, ST-ReSNet [15] adopted the framework of residual convolution network and considered the time series of three different trends to mine the spatiotemporal relationship; then, external factors such as weekdays and weather were integrated to improve the accuracy of traffic flow prediction. Liu et al. [16] used Long Short-Term Memory (LSTM) and full connection layer(FC) to predict passenger flow inbound and outbound of metro stations. Ma et al. [17] converted the data of metro passengers (i.e., Destination station) in Beijing into images by using Convolutional Neural Networks(CNN) and a bi-directional LSTM layer. Yang et al. [18] proposed a novel attention mechanism-based end-to-end neural network to predict the inbound and outbound passenger flow, which improved the prediction effectiveness.

All the above models can effectively capture complex time features, and the proposed Graph Convolutional Neural Network (GCN) solves the problem of Network spatial correlation effectively. Traditional neural networks can only effectively extract some features from standard grid data and cannot deal with the complex and nonlinear traffic data well. In this case, we need to consider using graph convolutional neural network. Chen et al. [19] constructed topology, similarity (based on dynamic time-warping distance), and correlation graph to represent the dependency between passengers at different transfer stations, and then used a variant of graph neural network to conduct demand forecasting. Defferrard et al. [20] applied a three-dimensional convolution operation to seamlessly capture irregular spatiotemporal dependence on metro network. Guo et al. [21] proposed the ASTGCN model, which added an attention mechanism to take into account the dynamic influence of different time periods and different places on adjacent time periods and adjacent places. Yan et al. [22] proposed a spatiotemporal graph convolution model (STGCN) combining graph convolution and CNN based on spatial domain. Zhao et al. [23] put forward a special passenger flow forecasting prediction model based on temporal graph convolutional network (T-GCN), which combines a graph convolution network (GCN) and gated recursive network (GRU). In this model, GCN learns complex nonlinear structures to capture spatial topology, while the gated loop unit understands dynamic changes of traffic and captures time-dependent data.

There are many studies on the estimation and prediction of short-term traffic origin and destination, especially in the area of taxi and car-hailing travel [24]. The research on

passenger flow OD prediction starts from residents travel, road traffic, and urban public transportation network, which is called travel distribution prediction or traffic distribution prediction.

One of the important differences between OD prediction of traffic and public transport is the high dimension of data. A network with N sites consists of $N \times N$ OD pairs, and it is not extensible to exploit a model with a traditional method for each OD pair. Matrix/tensor decomposition is an effective method to solve the high-dimensional problem of OD matrix prediction, and many researches [25] have explored this based on it. Deep learning is also one of the mainstream OD prediction methods. Because the data is high dimensional, Toqué et al. [26] applied LSTM networks to selected high-traffic OD pairs with heavy traffic. Wang et al. [27] improved OD flow prediction network of GCN + LSTM by multi-task learning. Shen et al. [28] mixed CNN with gravity model to predict the OD matrix of the metro system. The effect of deep learning model is usually affected by noise in sparse Metro OD matrix. To reduce the influence of noise, Zhang et al. [29] developed an index called OD attraction degree (ODAD) to cover up non-important OD pairs, indicating that shielding OD pairs close to zero can improve the prediction of LSTM. Meanwhile, a Channel-wise Attentive Split-CNN (CAS-CNN) model [30] is developed for metro OD matrix prediction. Gong et al. [31] proposed a real-time delayed data collection problem and discussed how to address it. Peyman et al. [32] considered the issue of delayed data availability, which is a challenge in the prediction of dynamic Origin-Destination (OD) demand.

From the traditional method to the current artificial intelligence method, it is all to better capture the regularity of the prediction content. Most of the predictions for inflow and outflow are at the station level, while the prediction of OD flow often involves stations in the entire network. The traditional method mainly makes predictions based on statistical laws, but it cannot show the influence of various factors, and the prediction accuracy is low. The GRU method is often used for simple time series prediction, which is very effective, but it lacks the capture of spatial information in the problem of rail transit OD passenger flow prediction. Using the knowledge of graph theory and convolution, GCN can well capture the physical space information brought by the metro network and the virtual space information brought by the OD passenger flow. Masking unimportant OD pairs was proposed in previous methods, which has a strong inspiration for us. Therefore, in this work, we study how to extract important OD pairs for research. At the same time, according to the analysis of the properties of key OD passenger flow, GRU combined with GCN is selected to capture spatiotemporal information to achieve good results.

3. Methodology

3.1. Key OD Pairs Prediction Problem. The Metro AFC system records the original travel data of passengers, including card number, entry station number and time, transaction type (inbound or outbound), and other data. In

this way, we know the entry and exit stations and the corresponding time. We summarize them at a fixed time interval and calculate the traffic demand from O to D in the statistical period based on O , regardless of whether the journey is completed or not. We finally form an $N \times N$ OD matrix, where N represents the total number of metro stations in the dataset. We extract the key OD pair set as

$$K_{OD} = \{k_1, k_2, \dots, k_n\}. \quad (1)$$

In our work, a virtual graph between key OD pairs denoted as $G = (V, E)$ is established, which is built to represent the connection relationship between OD pairs. V and E denote the set of nodes and the set of edges, respectively. In our work, V is equivalent to the set of key OD pairs used to encode the characteristics of nodes, which refers to the traffic time series of each OD pair. The adjacency matrix A that only contains elements of 0 and 1 represents the virtual connection relationship between OD pairs. The corresponding element is 1 if there is an association relationship between nodes, otherwise, it is 0. The data collected by the key OD pair within t consecutive moments is expressed as $X = [x^1, \dots, x^t]$, where $x^t = [x(k_1), x(k_2), \dots, x(k_n)]^T$ represents the data collected by each key OD pair at the time t . The feature matrix $X^{n \times p}$ represents the attribute characteristics of the node, n represents the number of key OD pairs, and p represents the historical time sequence length. The purpose of our paper is to predict the passenger flow of the future m time intervals according to the historical passenger flow data of key OD pairs in previous n time intervals, which can be expressed as the following learning function:

$$[X^{T-n}, \dots, X^{T-1}, X^T; G] \xrightarrow{f} [X^{T+1}, \dots, X^{T+m}], \quad (2)$$

where X^T represents the values of all OD pairs at time interval T and f is a mapping function.

3.2. Key OD Pairs Extraction. The OD matrix describes the number of travels between each OD pair in the system during the time interval. The amount of stations in the metro system is N , and the size of the generated OD pairs is $N \times N$, which means a huge number. The heat map formed by the OD matrix at the 10-minute granularity is shown in Figure 1(a). As shown by the color bar on the right side of Figure 1(a), the brighter the color indicates the greater the amount of OD between the two stations. It can be seen that the graph has a large black part, that is, most of the elements in the OD matrix are 0, which means that there is no passenger travel demand between most of the stations. Therefore, not all OD pairs are the ones we need to pay attention to. The passenger flow of some noncritical OD pairs is scarce, and the travel demand is very random. The contribution of this kind of OD passenger flow is very small, implying that relatively few key OD pairs account for the vast majority of the overall OD passenger flows. So we use the historical passenger flow data set $H^s = \{h^1, h^2, \dots, h^s\}$ of each OD pair to filter OD pairs, where s is the length of the selected historical data, h^i represents the OD value of the period. In order to achieve the extraction of key OD pairs, we set three thresholds, including

the proportion z of nonzero value, the randomness judgment value r and the proportion p of value that is greater than the randomness judgment value. The key OD pairs meet the following conditions:

$$\frac{|H^{s'}|}{|H^s|} * 100\% \leq z \text{ and } \frac{|H^{s''}|}{|H^s|} * 100\% > p, \quad (3)$$

where $H^{s'} = \{h^i \mid h^i \in H^s \text{ and } h^i = 0, i = 1, 2, \dots, s\}$ means a subset of key OD pairs set in which elements value is zero and $H^{s''} = \{h^i \mid h^i \in H^s \text{ and } h^i > r, i = 1, 2, \dots, s\}$ a subset in which elements value is greater than randomness judgment value r .

We selected 14 OD pairs and showed them in Figure 1(b). The horizontal axis represents time. From this picture, we can see that the critical and noncritical OD pairs exhibit large differences. And different OD pairs show different time-varying laws. The color bar of OD14 is all black, indicating that there is no OD demand between the two stations. OD11–OD13 have some faint colors, indicating that the passenger travel demand between such noncritical OD pair is very random. The remaining OD pairs show different color distributions, representing different passenger flow characteristics.

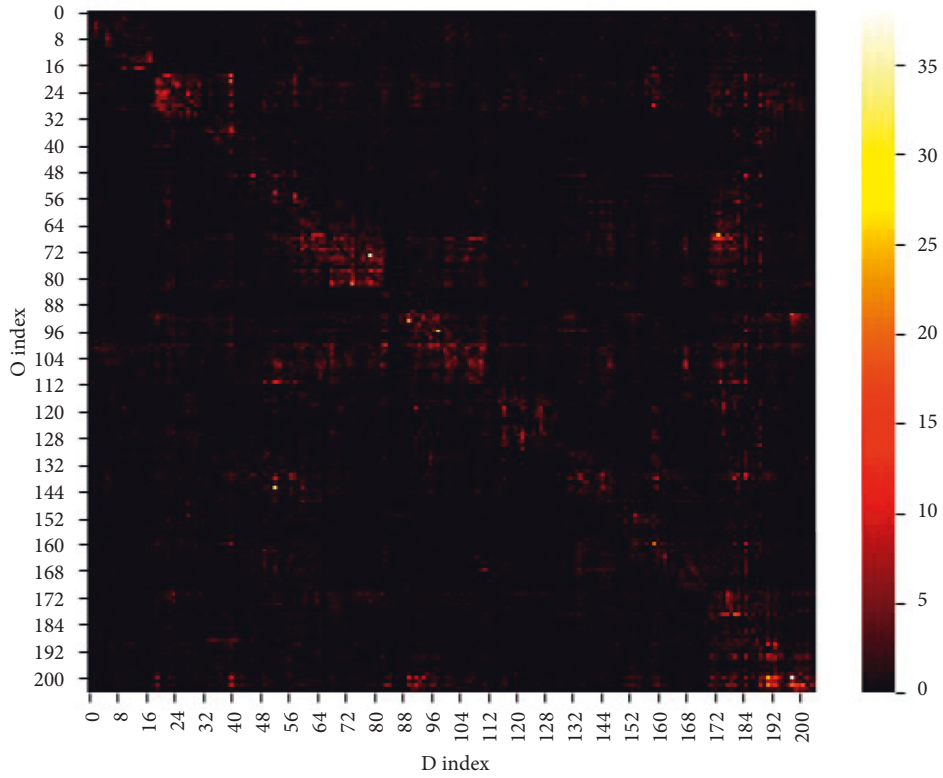
3.3. Virtual Graph Construction. Graph neural network performs graph convolution based on the relationships between nodes. We conduct research and prediction based on key OD pairs, so it loses the dependency information that comes with the physical topology of the real stations. However, the OD pairs combine the spatial properties of the origin-destination station by themselves. For example, if station O is a residential area and station D is an office area, this kind of OD pair has similar properties. The OD pairs show correlation, that is, OD pairs may have similar traffic distribution characteristics because of similar functionality. So virtual connection edges can be established to generate an adjacency matrix.

To measure the degree of correlation between two variables, researchers usually use the Pearson correlation coefficient. In the quantification of the correlation between time series, it can be used to measure how two continuous signals change together with time, and the correlation coefficient shows their relationship. The Pearson correlation coefficient can be calculated by the following formula:

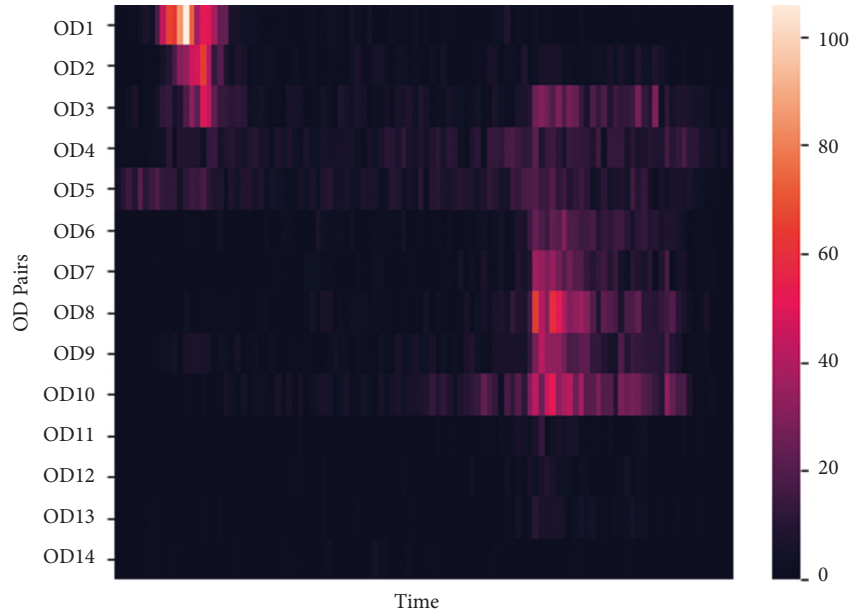
$$\rho_{X,Y} = \frac{\sum (X - \bar{X})(Y - \bar{Y})}{\sqrt{\sum (X - \bar{X})^2 \sum (Y - \bar{Y})^2}}, \quad (4)$$

where X and Y represent two variables and \bar{X} and \bar{Y} represent the mean of the variables.

According to the above calculation formula, we obtain the correlation matrix R that describes the relation between key OD pairs. R_{ij} represents the Pearson correlation coefficient between observed historical passenger flow data series of the i_{th} and j_{th} OD pair. Classical GCNs encode adjacency between nodes to represent arbitrarily structured graphs. A binary-encoded adjacency matrix A is usually used to represent the connectivity between nodes. $A_{ij} = 1$ if nodes i and j are directly connected in the graph, otherwise $A_{ij} = 0$. According to the meaning of the correlation coefficient, if



(a)



(b)

FIGURE 1: Visual display of raw data. (a) Heat map formed by an OD matrix. (b) Heat map of some OD pairs over time.

the absolute value of the relative coefficient is larger, the correlation between variables is stronger. We set a threshold c to determine whether to establish a virtual connection between OD pairs. The formula is as follows:

$$A_{ij} = \begin{cases} 1, & R_{ij} \geq c, \\ 0, & R_{ij} < c. \end{cases} \quad (5)$$

3.4. Spatiotemporal Virtual-Graph Convolution Network. After completing the virtual graph construction, we use the combined model of GCN and GRU to model the spatiotemporal dependencies of key OD pairs. The prediction framework of this paper is shown in Figure 2. The main structure of the model is divided into three parts, which are the input layer, feature extraction layer, and output layer.

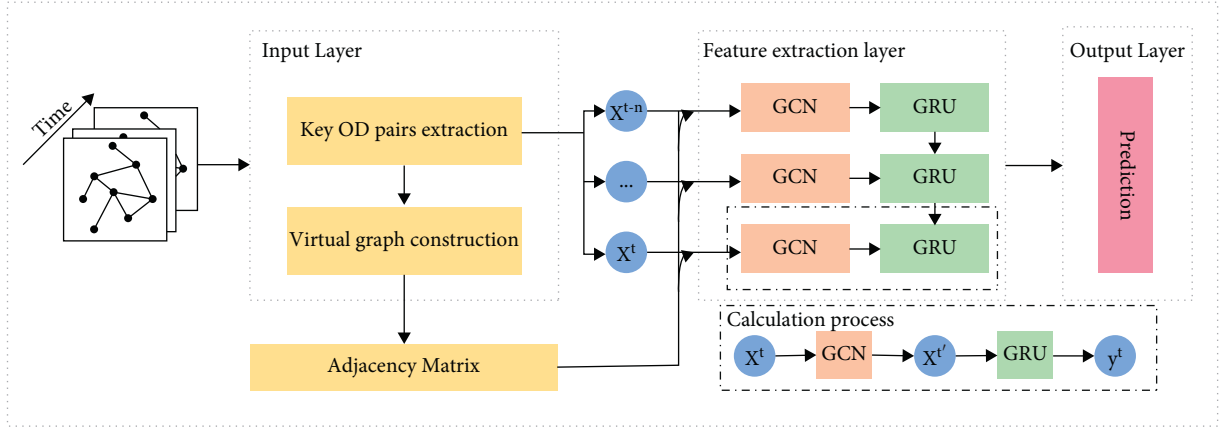


FIGURE 2: The structure of the ST-VGCN model.

The input layer receives raw historical OD matrix data. After Key OD pairs extraction, feature matrix X that represents multiple time steps historical flow of all key OD pairs is generated. And through virtual graph construction, adjacency matrix A is generated. The feature extraction layer is composed of a graph convolution module and a GRU module, and the inputs of it are X and A . Firstly, the graph convolution module receives the data from the input layer and learns both node features and structural information end-to-end through graph convolution operation to obtain rich node information and aggregate spatial features. Then the result with spatial features of each node is sent to the GRU module to capture the time series features. The final output layer gets the prediction result.

There are two main categories of GCN methods: spectral-based and spatial-based. In our work, we use a graph convolutional network based on the spectral method. Given a feature matrix X representing node features and an adjacency matrix A representing structural features, the graph convolution operation computes the information of nodes using the information of related nodes. The core calculation formula is as follows:

$$f(X, A) = \sigma(\hat{D}^{-1/2} \hat{A} \hat{D}^{-1/2} X W). \quad (6)$$

Here, $\hat{A} = A + I_N$ represents the sum of the adjacency matrix A and the identity matrix I_N ; \hat{D} represents the degree matrix of \hat{A} ; $\hat{D}^{-1/2} \hat{A} \hat{D}^{-1/2}$ represents the \hat{A} is normalized; W represents the weight matrix; and σ denotes the activation function.

GRU is one of the most widely used recurrent neural networks for processing series data. It can be regarded as a combination of reset gate and update gate. It is used to model the sequence information that has undergone graph convolution operations to capture its temporal features and complete the prediction task. After the original sequence $X = [x^1, \dots, x^t]$ passing through the graph convolution layer, new sequence data containing spatial information is obtained as $X' = [x^{1'}, \dots, x^{t'}]$. We input the new sequence data into the GRU network. The feature extraction layer improves the basic GRU structure in combination with graph convolution operation. The result is shown in Figure 3.

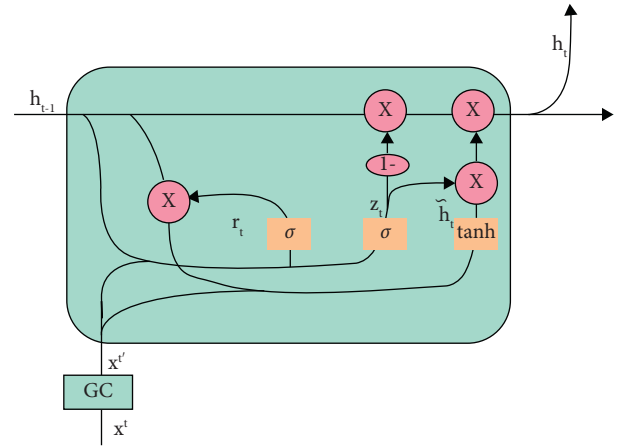


FIGURE 3: Basic GRU structure in combination with graph convolution operation.

This process can be described by the following equations:

$$\begin{aligned} z_t &= \sigma(W_z [f(X^t, A), h_{t-1}] + b_z), \\ r_t &= \sigma(W_r [f(X^t, A), h_{t-1}] + b_r), \\ \tilde{h}_t &= \tanh(W_h [f(X^t, A), (r_t * h_{t-1})] + b_h), \\ h_t &= z_t * h_{t-1} + (1 - z_t) * \tilde{h}_t. \end{aligned} \quad (7)$$

In the formula, h_{t-1} is the hidden state at time $t - 1$; X^t is the flow information of all key OD pairs at time t ; r_t is the reset gate in the GRU model, which is about how the new input information is integrated with the previous memory; the z_t update gate indicates the amount of previous memory saved to the current time step; \tilde{h}_t is the memory content stored at time t ; h_t is the output state at time t . GRU takes the hidden state at time $t - 1$ and the current key OD pair flow information that has undergone graph convolution operations as input, and obtains the flow state at time t . The key OD flow is predicted in order to make the forecast result as close as possible to the actual traffic demand. Therefore, we need to select the loss function to estimate the degree of inconsistency between the predicted value and the real value of the model. Our ultimate goal is to minimize losses during

training process. The loss function chosen in this paper is shown as follows:

$$\text{loss} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2, \quad (8)$$

where y_t and \hat{y}_t are the true OD flow and predicted values and n represents the length of observation window.

4. Experiments

4.1. Data Description. Because of few public benchmarks for metro passenger forecasts, we construct the dataset MetroSZ2020 with 39 consecutive days of metro smart card data which records card number, origin number, destination number, and the entry and exit time of each metro trip from Shenzhen, China. Because some data in the Shenzhen metro network cannot be obtained, after data cleaning, the data of 205 stations are available. As shown in Table 1, MetroSZ2020 covers 205 metro stations in Shenzhen from August 23rd to September 30th, 2020. We select 6:00–24:00 as the metro operating period. We count the OD matrix every 10 minutes, which means 108 matrixes per day and contains a total of 42025 OD pairs per matrix.

Three thresholds for z , r , and p are set as 80%, 10 and 10% respectively. We test different threshold combinations in experiments and obtain different numbers of key OD pairs. However, considering the hardware resource problem and the model training situation, the above threshold combination is finally selected. By analyzing the OD matrix heat map similar to Figure 1(a), it can be seen that it is meaningful to set the nonzero value ratio z to at least 80%. The randomness judgment value r needs to be changed according to the change of time granularity. The time granularity selected in this paper is 10 min. For the OD matrix of 205 stations, setting r to 10 can already indicate that the OD passenger flow has a certain regularity. For a station with regular passenger flow, it takes a sustained period of time to express the regularity, so setting p to 10% is a reasonable choice. The key OD pairs that meet less than 80% of the historical time series data is zero and more than 10% of the data value is greater than 10 are extracted. A total of 490 key OD pairs of data are used as input and we scale the data to (0, 1], divide by the maximum value into the data. When proceeding with virtual graph construction, we set the threshold c as 80%. The data is divided into training data and validating data according to the ratio, which is 0.8 in our research.

4.2. Model Configurations. Our experiment is completed in the Pytorch environment on a workstation equipped with an Intel(R) Core(TM) i7-6800k processor whose cache is 15M and working frequency is up to 3.40 GHz, 16 GB memory space, and NVIDIA GeForce GTX 1080 Ti graphics card. We train the model using the Adam optimizer. To obtain the best experimental results, we manually adjust the determined parameters including the number of hidden units and the number of training epoch which may greatly affect the prediction precision. As shown in Figure 4, the horizontal axis indicates the different parameter choices and the vertical axis

TABLE 1: The description of the dataset.

Description	MetroSZ2020
Date	2020/8/23–2020/9/30
Time	6:00–24:00
Station number	205
Matrix dimension	205*205
Key OD pairs number	490
Time interval	10 min

indicates the variation of the different metrics. The red dots indicate the performance of each metric under the selected parameters. First, we test the training epoch in the set [500, 1,000, 1,500, 2,000, 3,000, 3,500, 4,000, 4,500, 5,000] and analyze the variation of the model performance. Figure 4(a) shows the results of metrics for different training epochs, and Figure 4(b) shows the variation of metrics for different hidden units. With the increase of training epoch value, the variation of evaluation metrics stabilizes, and there is a turning point at 3,000. So we fix the training epoch at 3,000 and select the number of hidden cells from the set [32, 64, 100, 128, 256]. As shown in Figure 4(b), the model becomes stable when hidden units reach 128. Therefore, the training epoch is affirmed as 3000 and the number of hidden cells as 128. Besides, the time step is affirmed as 12, the batch size as 64, and the learning rate as 0.001.

4.3. Evaluation Metrics. In this work, we choose root mean square error (RMSE), mean absolute error (MAE), and linear regression coefficient of determination (R^2) as evaluation metrics. In real experiments, each evaluation metric represents a different meaning. RMSE and MAE represent the error between the predicted value and the actual value. R^2 reflects the fitting effect of the model, which measures the ability of prediction results to represent actual data by calculating correlation coefficients. As we all know, the smaller the error, the closer the predicted value is to the real value, and the higher the fitting degree, the better the prediction effect of the model. The specific calculation formulae are as follows:

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (Y_t - \hat{Y}_t)^2},$$

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |Y_t - \hat{Y}_t|, \quad (9)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (Y_t - \hat{Y}_t)^2}{\sum_{i=1}^n (Y_t - \bar{Y}_t)^2},$$

where Y_t represents the real OD information, \hat{Y}_t represents the predicted OD information, \bar{Y}_t represents the mean of the predicted OD information, and n is the number of nodes.

4.4. Results and Discussion. In this section, we select the following five baseline methods for comparison with the model in this paper, including three traditional time series

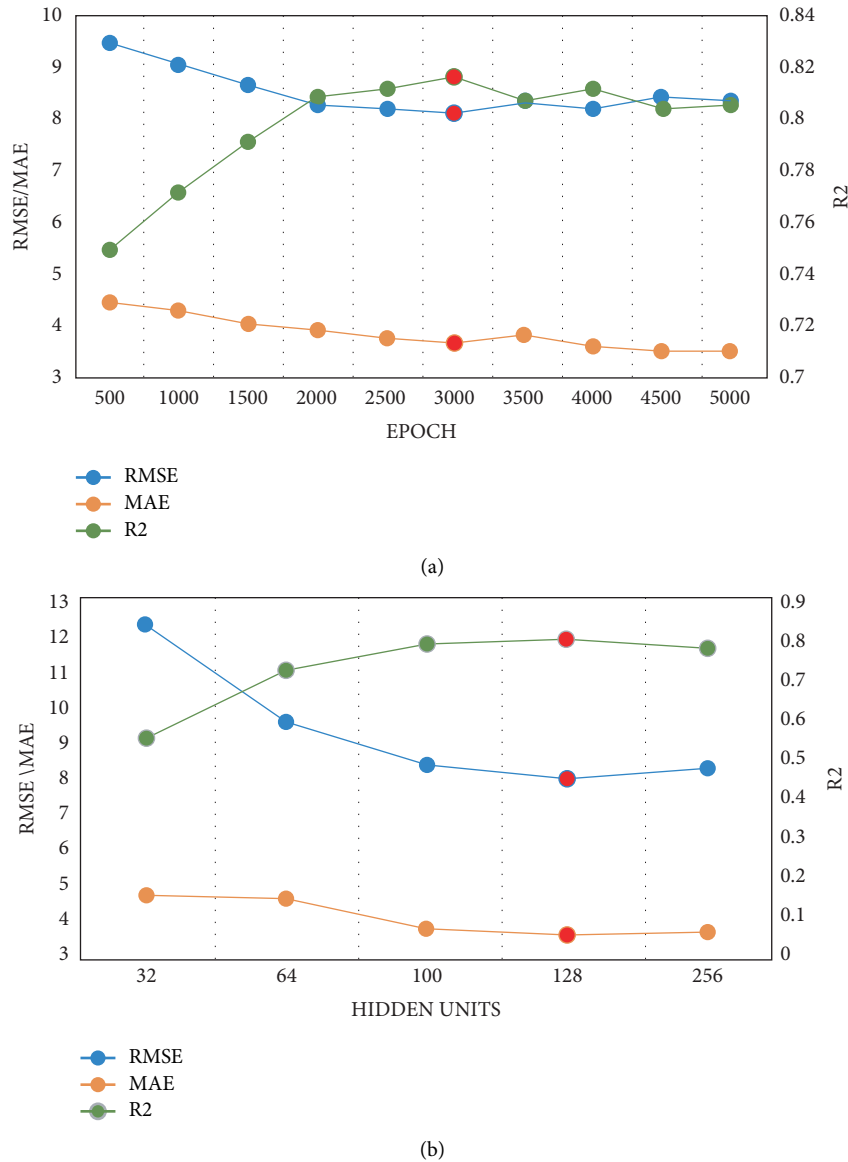


FIGURE 4: Parameter tuning results. (a) Variation of predicted performance under different epochs. (b) Variation of predicted performance under different hidden units.

models as well as two generalized deep learning models, namely (1) Historical Averaging model (HA), (2) Autoregressive Integrated Moving Average model (ARIMA), (3) Support Vector Regression model (SVR), (4) Graph Convolutional Network model (GCN), and (5) Gated Recurrent Unit model (GRU). To verify the effectiveness of the model in our work for the key OD flow prediction problem, we compare the evaluation metrics with some comparative experiments, as shown in Table 2.

We explore the performance of ST-VGCN for key OD pair flow prediction at 10-minute time granularity. We can find that ST-VGCN performs much better than HA model on all performance metrics, and its RMSE is reduced by approximately 26.15% and 0.6% compared to ARIMA and SVR models, respectively. GCN and GRU are both deep learning methods, but GCN only focuses on spatial

TABLE 2: Comparison of performances of different models.

Models	RMSE	MAE	R^2
HA	15.134	5.787	0.360
ARIMA	11.321	6.065	0.105
SVR	8.416	3.748	0.802
GCN	13.619	5.191	0.482
GRU	12.071	5.015	0.593
ST-VGCN	8.360	3.492	0.805

relations while GRU only focuses on temporal relations. ST-VGCN model takes both into account and its RMSE is reduced by about 38.62% and 30.74% compared to GCN and GRU models, respectively. Other indicators also perform significant improvements. Through the above analysis and comparison, we verify the improvement of the

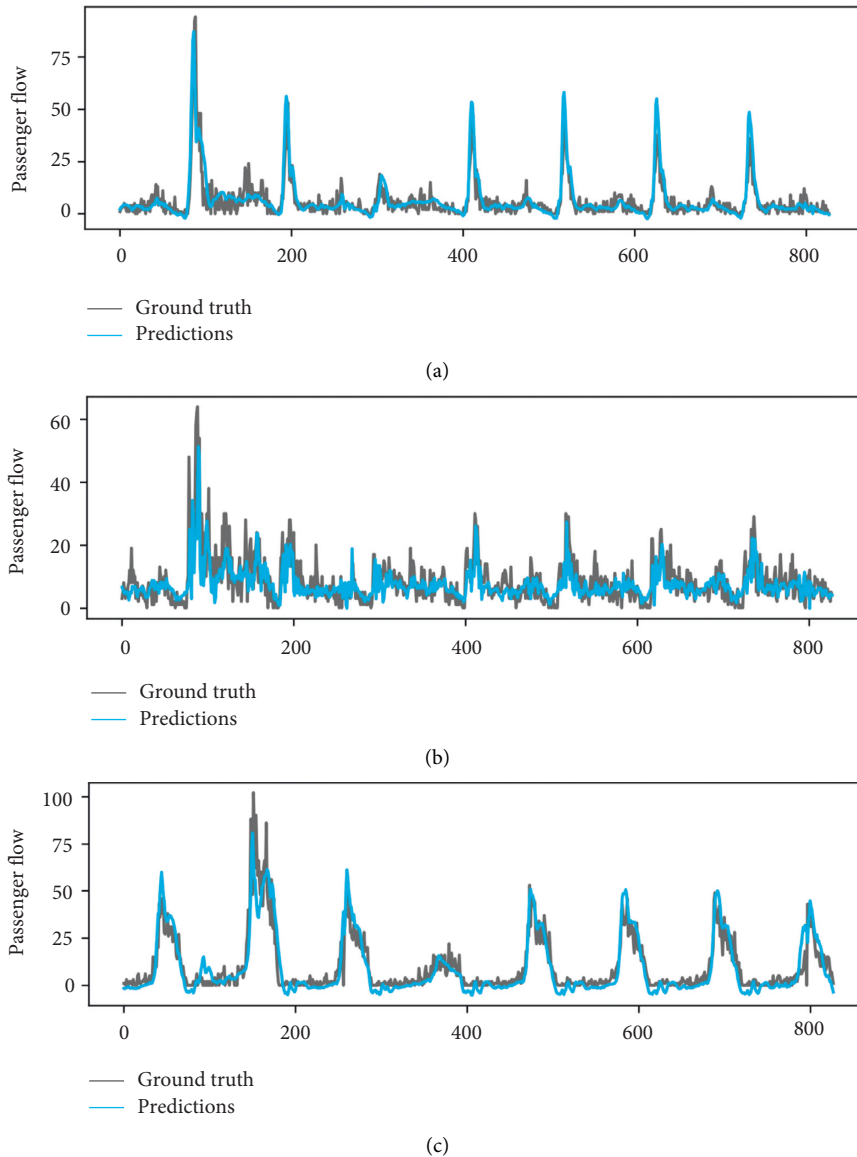


FIGURE 5: The real and predicted time series of three selected OD pairs.

proposed ST-VGCN model in three indicators. And we believe that the model in our work is effective. The comparison results validate the superiority of the proposed ST-VGCN model. For the 10-minute granularity of key OD pairs flow, the prediction result is visualized. As shown in Figure 5, three different passenger travel patterns can be seen. Figures 4(b) and 5(c) have explicit spike moments, and the spike moment in 5(c) is later than that in 5(a). After analysis, it is concluded that the nature of the OD pair of stations represented in Figure 5(a) is that station O is a residential area and station D is an office area. Figure 5(c) represents that station O is an office area and station D is a residential area, which is in line with the residents' commuting pattern. Figure 5(b) maintains a relatively flat trend, where station D is an airport station, so it maintains a certain level of passenger travel demand throughout the day.

As can be seen, our model can well model different passenger flow demands. Accurate predictions can provide effective reference information for passenger travel. At the same time, the modeling and analysis of the key OD pairs flow can assist the emergency response [33] when an unexpected event occurs, and reduce the continuous impact caused by the occurrence of emergencies.

5. Conclusions

Our work mainly defines and studies the key OD pair flow prediction problem and proposes a complete set of procedures for accomplishing the prediction. The key steps are to obtain key OD pairs, model the correlation between them, and use the ST-VGCN model to complete the prediction. The ST-VGCN model achieved the best prediction results on a real-world dataset when compared with five preexisting

models. Through experiments and analysis, we have obtained the following conclusions:

- (1) OD flow has the problem that the data is too sparse and its dimension is too high, yet most OD pairs have a small contribution to the overall passenger flow. So it is necessary to extract key OD pairs for research by setting a threshold filter.
- (2) The OD pairs exhibit certain correlations with each other due to the same nature of the origin-destination station and time variation pattern. It makes sense to consider the spatiotemporal characteristics of OD passenger flow when predicting.
- (3) The proposed ST-VGCN model can combine temporal and spatial information to improve the ability of prediction. By establishing a virtual graph, GCN is used to capture spatial properties. GRU is used to capture time series information.

Overall, research on key OD pair flow prediction can provide important insights for metro operation and management. In the future, the impact of multiple sources of data needs to be considered which can further improve the prediction ability of the model. It is also an open question worthy to be explored how to determine the threshold value for filtering key OD pairs more reasonably. Also, the problem of not being able to obtain real-time OD information due to trip duration is not explored in this study, and we will consider the problem of how to obtain real-time information in the next work.

Data Availability

Due to the nature of this research, participants of this study did not agree for their data to be shared publicly, so supporting data are not available.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was funded by the Beijing Municipal Natural Science Foundation (Grant No. L201015); the National Key R&D Program of China (Grant No. 2020YFC0833104); and the Green, Intelligent and Safe Mining of Coal Resources (Grant No. 52121003).

References

- [1] B. L. Smith and M. J. Demetsky, "Traffic flow forecasting: comparison of modeling approaches," *Journal of Transportation Engineering*, vol. 123, no. 4, pp. 261–266.
- [2] Li. Zhu, Y. Li, F. R. Yu, B. Ning, T. Tang, and X. Wang, "Cross-layer defense methods for jamming-resistant cbtc systems," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 11, pp. 7266–7278, 2021.
- [3] Y. Li, Li. Zhu, H. Wang, F. R. Yu, and S. Liu, "A cross-layer defense scheme for edge intelligence-enabled cbtc systems against mitm attacks," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 4, pp. 2286–2298, 2021.
- [4] K. Kumar, M. Parida, and V. K. Katiyar, "Short term traffic flow prediction for a non urban highway using artificial neural network," *Procedia - Social and Behavioral Sciences*, vol. 104, pp. 755–764, 2013.
- [5] J. Yang, Y. Zheng, K. P. Yan et al., "SPDNet: a real-time passenger detection method based on attention mechanism in subway station scenes," *Wireless Communications and Mobile Computing*, vol. 2021, Article ID 7978644, 13 pages, 2021.
- [6] G. Filmon, "Habtemichael and Mecit Cetin. Short-term traffic flow rate forecasting based on identifying similar traffic patterns," *Transportation Research Part C: Emerging Technologies*, vol. 66, pp. 61–78, 2016.
- [7] Li. Zhu, H. Liang, H. Wang, B. Ning, and T. Tang, "Joint security and train control design in blockchain-empowered cbtc system," *IEEE Internet of Things Journal*, vol. 9, no. 11, pp. 8119–8129, 2022.
- [8] Z. Zhu, Bo. Peng, C. Xiong, and L. Zhang, "Short-term traffic flow prediction with linear conditional Gaussian bayesian network," *Journal of Advanced Transportation*, vol. 50, no. 6, pp. 1111–1123, 2016.
- [9] B. M. Williams and L. A. Hoel, "Modeling and forecasting vehicular traffic flow as a seasonal arima process: theoretical basis and empirical results," *Journal of Transportation Engineering*, vol. 129, no. 6, pp. 664–672, 2003.
- [10] E. Zivot and J. Wang, "Vector Autoregressive Models for Multivariate Time Series," *Modeling Financial Time Series with S-Plus®*, Springer, New York, NY, USA, pp. 385–429, 2003.
- [11] C. H. Wu, J. M. Ho, and D. T. Lee, "Travel-time prediction with support vector regression," *IEEE Transactions on Intelligent Transportation Systems*, vol. 5, no. 4, pp. 276–281, 2004.
- [12] J. Yang and Z. Hou, "A wavelet analysis based ls-svm rail transit passenger flow prediction method," *China Railway Science*, vol. 34, no. 3, pp. 122–127, 2013.
- [13] P. Jiao, R. Li, T. Sun, Z. Hou, and A. Ibrahim, "Three revised kalman filtering models for short-term rail transit passenger flow prediction," *Mathematical Problems in Engineering*, vol. 2016, Article ID 9717582, 10 pages, 2016.
- [14] J. Zhang, Y. Zheng, D. Qi, R. Li, and X. Yi, "Dnn-based prediction model for spatio-temporal data," in *Proceedings of the 24th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, SIGSPACIAL '16, New York, NY, USA, 2016.
- [15] J. Zhang, Y. Zheng, and D. Qi, "Deep spatio-temporal residual networks for citywide crowd flows prediction," in *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence*, AAAI Press, San Francisco, CA, USA, February 2017.
- [16] Y. Liu, Z. Liu, and R. Jia, "Deeppf: a deep learning based architecture for metro passenger flow prediction," *Transportation Research Part C: Emerging Technologies*, vol. 101, pp. 18–34, 2019.
- [17] X. Ma, J. Zhang, B. Du, C. Ding, and L. Sun, "Parallel architecture of convolutional bi-directional lstm neural networks for network-wide metro ridership prediction," *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, no. 6, pp. 2278–2288, 2019.
- [18] J. Yang, X. Dong, and S. Jin, "Metro passenger flow prediction model using attention-based neural network," *IEEE Access*, vol. 8, Article ID 30953, 2020.

- [19] J. Chen, L. Liu, H. Wu, J. Zhen, G. Li, and L. Lin, "Physical-virtual collaboration graph network for station-level metro ridership prediction," 2020, <https://arxiv.org/abs/2001.04889>.
- [20] M. Defferrard, X. Bresson, and P. Vandergheynst, "Convolutional Neural Networks on Graphs with Fast Localized Spectral Filtering," 2016, <https://arxiv.org/abs/1606.09375>.
- [21] S. Guo, Y. Lin, F. Ning, C. Song, and H. Wan, "Attention based spatial-temporal graph convolutional networks for traffic flow forecasting," in *Proceedings of the Thirty-Third AAAI Conference on Artificial Intelligence and Thirty-First Innovative Applications of Artificial*, AAAI Press, Honolulu, HI, USA, February 2019.
- [22] S. Yan, Y. Xiong, and D. Lin, "Spatial temporal graph convolutional networks for skeleton-based action recognition," 2018, <https://arxiv.org/abs/1801.07455>.
- [23] L. Zhao, Y. Song, M. Deng, and H. Li, "Temporal graph convolutional network for urban traffic flow prediction method," 2018, <https://arxiv.org/abs/1811.05320>.
- [24] L. Liu, Z. Qiu, G. Li, Q. Wang, W. Ouyang, and L. Lin, "Contextualized spatial-temporal network for taxi origin-destination demand prediction," *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, no. 10, pp. 3875–3887, 2019.
- [25] J. Liu, F. Zheng, H. J. van Zuylen, and J. Li, "A dynamic od prediction approach for urban networks based on automatic number plate recognition data," *Transportation Research Procedia*, vol. 47, pp. 601–608, 2020.
- [26] F. Toqué, E. Côme, M. Khalil El Mahrsi, and L. Oukhellou, "Forecasting dynamic public transport origin-destination matrices with long-short term memory recurrent neural networks," in *Proceedings of the 2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC)*, pp. 1071–1076, IEEE, Rio de Janeiro, Brazil, November 2016.
- [27] Y. Wang, H. Yin, H. Chen, T. Wo, J. Xu, and K. Zheng, "Origin-destination matrix prediction via graph convolution: a new perspective of passenger demand modeling," in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp. 1227–1235, Anchorage AK USA, August 2019.
- [28] L. Shen, Z. Shao, Y. Yu, and X. M. Chen, "Hybrid approach combining modified gravity model and deep learning for short-term forecasting of metro transit passenger flows," *Transportation Research Record*, vol. 2675, no. 1, pp. 25–38, 2021.
- [29] J. Zhang, F. Chen, Z. Wang, and H. Liu, "Short-term origin-destination forecasting in urban rail transit based on attraction degree," *IEEE Access*, vol. 7, Article ID 133452, 2019.
- [30] J. Zhang, H. Che, F. Chen, W. Mae, and Z. He, "Short-term Prediction of Urban Rail Transit Origin-Destination Flow: A Channel-wise Attentive Split-Convolutional Neural Network Method," *Transportation Research Part C: Emerging Technologies*, vol. 124, 2020.
- [31] Y. Gong, Z. Li, J. Zhang, W. Liu, and Yu. Zheng, "Online spatio-temporal crowd flow distribution prediction for complex metro system," *IEEE Transactions on Knowledge and Data Engineering*, vol. 34, no. 2, pp. 865–880, 2022.
- [32] N. Peyman, H. N. Koutsopoulos, and J. Zhao, "Dynamic origin-destination prediction in urban rail systems: a multi-resolution spatio-temporal deep learning approach," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 6, pp. 5106–5115, 2022.
- [33] J. Yang and Z. Hou, "A grey Markov based on large passenger flow real-time prediction model," *Journal of Beijing Jiaotong University*, vol. 37, no. 02, pp. 119–123, 2013.