





## Research Article

# Short-Term Forecasting of Dockless Bike-Sharing Demand with the Built Environment and Weather

Yang Yang <sup>1</sup>, Xin Shao <sup>1</sup>, Yuting Zhu <sup>2</sup>, Enjian Yao <sup>1</sup>, Dongmei Liu,<sup>3</sup> and Feng Zhao<sup>4</sup>

<sup>1</sup>Key Laboratory of Transport Industry of Big Data Application Technologies for Comprehensive Transport, Beijing Jiaotong University, No. 3 Shangyuancun, Haidian District, Beijing 100044, China

<sup>2</sup>School of E-Business and Logistics, Beijing Technology and Business University, Higher Education Garden, Liangxiang, Beijing 102488, China

<sup>3</sup>Research and Development Center of Transport Industry of Big Data Processing Technologies and Application for RIOH High Science and Technology Group, Research Institute of Highway Ministry of Transport, No. 8 Xitucheng Road, Haidian, Beijing 100088, China

<sup>4</sup>Tianjin Intelligent Traffic Operation Monitoring Center, No. 169 Weiguo Road, Hedong, Tianjin 300250, China

Correspondence should be addressed to Yuting Zhu; zhuyuting@btbu.edu.cn

Received 9 September 2022; Revised 6 December 2022; Accepted 25 December 2022; Published 3 February 2023

Academic Editor: Yajie Zou

Copyright © 2023 Yang 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.

To help related operators to allocate and dispatch the number of bike-sharing and provide good guidance for setting up electronic fences, this paper proposes a spatiotemporal graph convolution network prediction model (SGCNPM) with multiple factors to enhance the accuracy of predicting the demand for bike-sharing. First, we consider time, built environment, and weather. We use a multigraph convolution network (GCN) to model the built environment, utilize a long short-term memory (LSTM) network to extract temporal features, and utilize a fully connected network (FCN) to model weather influence. We construct SGCNPM which can effectively fuse GCN, LSTM, and FCN, thus creating a prediction method considering the influence of multiple factors. The results of the real case in Tianjin, China, show that the proposed model can perform well in improving prediction accuracy. Also, we analyze the influence of factors on model prediction results in different periods.

## 1. Introduction

Over the past few decades, with increasing concerns about global warming and rapid urbanization, many efforts have been put into cities to promote bike-sharing as a viable green mobility solution. The successful implementation of bike-sharing systems could provide alternative solutions to many urban problems such as traffic congestion, air pollution, energy shortages, and deterioration in human health. Taking a shared bike as an example, it is estimated that it will generate about 76 kg of carbon dioxide emissions in the whole life cycle of production, transportation, and scrapping. However, during its service life, the average cycling distance is more than 4,000 km, which is estimated to reduce carbon emissions by about 105 kg, enough to achieve “zero

carbon” [1]. Bike-sharing promotes comprehensive resource conservation and recycling, meets people’s green travel needs with lower resource input and higher operation efficiency, and contributes to the realization of carbon peaking and carbon neutrality goals.

Dockless bike-sharing has been introduced in China since 2014 and has grown rapidly in the past few years, spreading to overseas countries, including the US and the UK. In 2022, the number of bike-sharing in China reached 23 million, far surpassing docked bike-sharing and becoming one of the most important modes of shared transportation. The bike-sharing service platforms, e.g., Hellobike and Mobike, are emerging technologies with the mobile Internet boom. The platforms dispatch and transfer bike-sharing to match supply with demand. At present, bike-

sharing has problems such as low utilization rate, low turnover rate, and limited distribution. According to the report by the Beijing Municipal Commission of Transportation, in Beijing, the average daily turnover rate of bike-sharing is only 1.1, the average daily active bike-sharing accounts for only 16% of the total reported bike-sharing, and the average weekly active bike-sharing accounts for only 30% [2]. Understanding the short-term passenger demand in different spatial regions can help the platform and the operator solve this problem effectively. Then, bike-sharing can be dispatched to regions with more potential passenger demand to enhance the utilization and turnover rates.

The selection of factors is one of the most important steps in demand forecasting. In addition to time, weather and built environment are also important factors. Weather factors include temperature and weather state. Regarding the temperature, Heinen et al. investigated that travel demand was positively correlated when the temperature was between 0 and 20°C; demand reached its highest level when the temperature was between 20 and 30°C [3]. Regarding the weather state, Hyland and Hong found that rainfall and snowfall were the most unfavorable weather conditions, and there was a negative correlation between them and bike-sharing demand [4]. Built environment refers to the artificial environment provided for human activities, including a large urban environment. It is generally characterized by diversity (or land use type), design (particularly design for public transportation), and so on [5, 6]. Regarding the land use type, Eren and Uz concluded that bike-sharing demand was 15 times higher in business districts than in residential areas and 3–5 times higher in parks than in schools and subways [6]. Regarding the design for public transportation, Kabak et al. investigated that bike-sharing can effectively solve the last-kilometer connection problem, and bike-sharing demand was high near transportation hubs such as trains, buses, and subways [7].

Although many scholars have carried out relevant research on short-term forecasting of dockless bike-sharing demand based on considering multiple factors, few of them considered the impact of both weather and the built environment. Sun et al. demonstrated positive associations between taxi demand and built environment variables as well as weather conditions [8].

This paper intends to propose a short-term forecasting method with multiple factors to enhance the accuracy of predicting the demand for bike-sharing. The contributions are summarized as follows. First, this paper considers time, built environment, and weather. This paper uses a multi-graph convolution network (GCN) to model the built environment, utilizes a long short-term memory (LSTM) network to extract temporal features, and utilizes a fully connected network (FCN) to model weather influence. This paper constructs a spatiotemporal graph convolution network prediction model (SGCNPM) which can effectively fuse GCN, LSTM, and FCN, thus creating a prediction method considering the influence of multiple factors. The results of the real case in Tianjin, China, show that the proposed model can perform well in improving prediction

accuracy. Also, this paper analyzes the influence of factors on model prediction results in different periods.

The rest of this paper is organized as follows. Section 2 first introduces the explanatory variables. Section 3 proposes the modeling framework of this research. Section 4 gives the research of enterprise bike-sharing data and compares the predictive performance between the SGCNPM and other models. Section 5 summarizes the research results and proposes the future research direction.

## 2. Literature Review

There is much research on demand forecasting of bike-sharing. According to the research objects, they can be divided into station-based and dockless.

Focusing on station-based bike-sharing, Sathishkumar et al. discussed five models for hourly rental demand prediction, including linear regression, gradient boosting machine, support vector machine, boosted trees, and extreme gradient boosting trees [9]. Li et al. proposed a spatial-temporal memory network to predict short-term bike-sharing usage [10]. Sohrabi et al. proposed generalized extreme value count models which can predict hourly arrivals and departures at each station [11]. Reynaud et al. proposed a panel mixed generalized ordered logit model to estimate hourly bicycle availability; the model accommodated exogenous variables and a station-level model [12]. Hu et al. proposed a set of generalized additive models to delineate temporal interactions between station-level daily bike-sharing usage and independent variables, including land use, station characteristics, and COVID-19 [13]. Collini et al. used bidirectional long short-term memory (Bi-LSTM) networks to predict the number of available bikes and free bike slots in bike-sharing stations [14]. Mehdizadeh Dastjerdi and Morency first used the Louvain algorithm to identify six communities in the bike-sharing network and utilized CNN-LSTM to predict pickup demand in each community [15]. Unlike station-based bike-sharing, dockless sharing bikes can be parked by users in any appropriate place. This characteristic improves the availability of bicycles and service coverage but increases prediction difficulty. The forecast method for station-based bicycles cannot be used directly for dockless bike-sharing systems.

For the dockless bike-sharing system, Chang et al. utilized deep learning algorithms to predict the number and location of shared bikes, which adopted the encoder-decoder architecture embedded with the attention mechanism to enhance prediction ability further [16]. Shang et al. used big data to analyze the impacts of COVID-19 on the user behaviors and environmental benefits of bike-sharing [17]. They employed the topological indices arising from complex network theory to analyze the transformation of user behavior patterns. Yang et al. analyzed dockless bike-sharing in Nanchang, China, over a period when a new metro line came into operation, which utilized spatial statistics and graph-based approaches to quantify changes in travel behaviors [18]. Ai et al. proposed a convolutional long short-term memory (Conv-LSTM) network to forecast short-term distribution [19], which solved the problem of spatial

dependence and time dependence of dockless bike-sharing. Li and Shuai proposed a deep learning model called CLTFP to predict the travel distance and OD distribution of bike-sharing under different conditions of time and space [20]. Previous studies have mainly considered the time correlation when simulating the correlation, and this paper finds that the built environment is also important for the demand forecast for bike-sharing. For example, area A and area B are far apart. However, the land use types of the two areas are similar, or the two areas are connected by urban rail, so the two areas will influence each other, and the demand for bike-sharing may be similar.

Considering the built environment, Xu et al. used a four-month GPS dataset to reconstruct the temporal usage patterns of shared bikes at different places and applied an eigendecomposition approach to uncover their hidden structures [21]. Li et al. applied ordinary least squares (OLS) regression and geographically weighted regression (GWR) models to explore how the built environment and social-demographic characteristics influence bike-sharing utilization [22]. Dong et al. proposed DestiFlow based on points of interest (POIs) clustering to predict the demand for dockless bike-sharing [23]. Yan et al. investigated the travel distance distributions of dockless bike-sharing near metro stations to provide the basis for the service area of dockless bike-sharing [24]. Li et al. utilized the data of the dockless bike-sharing service Mobike to quantify short-trip transportation patterns and analyze the comprehensive view of mobility patterns [25]. Li et al. proposed a framework based on the gravity model and Bayesian rules to infer the purpose of dockless bike-sharing trips at the individual level [26]. In addition to bike-sharing, many scholars have also made short-term forecasts for other fields. Zeng et al. proposed a DWT-Bi-LSTM model to predict parking space availability based on historical parking data [27]. Ma et al. proposed a short-term traffic flow prediction model to improve the accuracy of short-term traffic flow prediction [28]. Ziheng et al. proposed a deep learning model MOS-BiAtten to predict ride-hailing demand during COVID-19 in Beijing [29]. Zhu et al. proposed a deep learning model to achieve accurate and stable taxi demand in dynamic areas [30].

Although there were many studies that analyzed demand, they also demonstrated the importance of weather and built environment for demand forecasting. However, few researchers considered both the built environment and the weather environment as variables. This paper sorts out the variables involved in the main literature, as shown in Table 1. The purpose of this paper is to explore bike-sharing demand forecasting with both the built environment and the weather environment.

**2.1. Preliminaries.** The short-term bike-sharing demand forecasting is essentially a time-series forecast problem. The nearest historical demand can provide valuable information for forecasting future demand. This paper also observes that the built environment influences the short-term bike-sharing demand. In this paper, the built environment is

characterized by two factors, land use type and accessibility of public transportation.

When forecasting a region's demand, other regions with similar functions can be intuitively referred to. If both region A and region B are residential areas, the spatiotemporal characteristics of bike-sharing demand in the two regions are comparable, and mutual reference can be made in predicting demand. Public transportation accessibility is also an essential factor in spatiotemporal prediction. Objectively speaking, geographically remote but accessible areas can be correlated. This connection is caused by public transportation such as buses and subways. Furthermore, the attributes of time-of-day, day-of-week, and weather conditions also impact the short-term bike-sharing demand. In this paper, the research area is divided into several grids consistently. Each grid refers to a zone and is represented by  $(m, n)$ . Also, a day is divided into different periods according to the same time interval. Then, the related variables are defined as follows.

**2.1.1. Demand Intensity.** The demand at the  $t$ -th time slot (hour) lying in the grid  $(m, n)$  is defined as the number of demands during this time slot within the grid, which is denoted by  $d_t^{m,n}$ . The bike-sharing demand in all  $M \times N$  grids at the  $t$ -th time slot is defined as the matrix  $D_t \in R^{M \times N}$  ( $R$  is the real set), in which the  $(m, n)$ -th element is  $(D_t)_{m,n} = d_t^{m,n}$ .

**2.1.2. Land Use Type.** The number of each land use type is used to measure this factor. Let  $p_i^{m,n}$  denote the number of land use types  $i$  in the grid  $(m, n)$ . The land use type  $i$  is divided into nine categories: subway stations, parks, shopping centres, training institutions, office buildings, schools, well-known enterprises, housing estates, and comprehensive restaurants. The number of land use type  $i$  in all  $M \times N$  grids is defined as the matrix  $P_i \in R^{M \times N}$  ( $R$  is the real set), in which the  $(m, n)$ -th element is  $(P_i)_{m,n} = p_i^{m,n}$ .

**2.1.3. Accessibility of Public Transportation.** The public transport system is the foundation of public transport accessibility. The spatial layout of the public transport network, the actual operation scheme, and the cooperation between the modes of public transport and rail transit within the system will all affect the choice of travellers' travel paths and modes in the public transport system, thus affecting the public transport accessibility. Let  $con^{m,n}$  indicate the accessibility of public transport in the grid  $(m, n)$ , representing the number of public transportation modes from area  $m$  to area  $n$ . The accessibility of public transportation in all  $M \times N$  grids is defined as the matrix  $CON \in R^{M \times N}$  ( $R$  is the real set), in which the  $(m, n)$ -th element is  $(CON)_{m,n} = con^{m,n}$ .

**2.1.4. Day.** By empirically examining the distribution of demand intensity concerning time in the training set, day-of-month, day-of-week, and time-of-day are selected. Let  $x_{da} \in [1, 31]$  denote the day-of-month,  $x_{we} \in [1, 7]$  denote day-of-week, and  $x_{tp} \in [1, 24]$  denote time-of-day. In

TABLE 1: Variables involved in the literature about dockless bike-sharing.

Literature	Involved time variables	Involved weather variables	Involved built environment variables
Chang et al. 2021	The number and location of bike-sharing	×	×
Ai et al. 2019	Number of bicycles, distribution uniformity, usage distribution	×	×
Kim D. 2018	Weekend, public holiday	Temperature, humidity, precipitation, wind speed	×
Hyland et al. 2018	Weekdays, weekends	Snowfall, temperature	×
Dong et al. 2019	×	×	POI density
Yan et al. 2020	×	×	Land use type, population density
Li et al. 2020	Spatiotemporal distribution of bike-sharing usage	×	Road characteristics, urban land use
This paper	The number of bike-sharing, working day, rest day	Temperature, weather state, wind speed, air quality index	Land use type, public transport accessibility

addition, a dummy variable  $x_{ho}$  is introduced to catch up on the distinguished properties between rest days (including holidays and weekends) and working days (i.e., weekdays). It is given by

$$x_{ho} = \begin{cases} 0, & \text{rest days,} \\ 1, & \text{working days.} \end{cases} \quad (1)$$

**2.1.5. Weather.** This paper considers five categories of weather variables: weather state, maximum temperature, minimum temperature, wind level, and air quality index. Let  $x_{wt}$  denote the weather state. It is given by

$$x_{wt} = \begin{cases} 0, & \text{cloudy,} \\ 1, & \text{fine,} \\ 2, & \text{haze,} \\ 3, & \text{snow,} \\ 4, & \text{rain.} \end{cases} \quad (2)$$

Let  $x_{ma} \in [-50, 50]$ ,  $x_{mi} \in [-50, 50]$ ,  $x_{wl} \in [0, 5]$ , and  $x_{aq} \in [0, \infty]$  denote maximum temperature, minimum temperature, wind level, and air quality index, respectively. The weather state, the maximum temperature, and the minimum temperature take a value for a time interval, while wind level and air quality index are taken on average in each time interval.

### 3. Methodology

**3.1. Framework.** SGCNPM consists of two submodels: (1) the spatiotemporal and built environment variable model based on multigraph convolution network and LSTM and (2) the weather variable model based on fully connected network. The output results of the two submodels are fused with different weights to obtain the bike-sharing demand of each region, as shown in Figure 1.

Firstly, we encode relationships between regions through multiple graphs, such as distance, POI similarity, and public transport accessibility. The non-Euclidean relationship is captured by using multigraph convolution network, and temporal features are extracted by LSTM, and the prediction result of the submodel is output; the historical demand and the weather labels are input by fully connected network, and the prediction result of the submodel is output. Lastly, the two results are weighted and fused, and the forecast value is output.

At the spatial level, multigraph convolution network is used to model the non-Euclidean relationship [32].

At the temporal level, considering the historical travel characteristics of bike-sharing, the LSTM is used to model the travel characteristics of bike-sharing, which obtains the demand rule of bike-sharing.

#### 3.2. Spatial Modeling

**3.2.1. Distance Map.** Each grid is regarded as a vertex of the graph, the centre of the grid is the vertex's location, and the distance between vertices is the distance between the centres of the grids. If the distance between two grids is close, the demand for bike-sharing is similar. The range set of distances can be expressed as

$$\alpha_i = \{a_i \mid \text{dis}(a_m, a_n)\}, \quad (3)$$

where  $a_m$  and  $a_n$  represent the centre of region  $m$  and region  $n$ , respectively, and  $\text{dis}(a_m, a_n)$  denotes the Euclidean distance between region  $m$  and region  $n$ .

**3.2.2. POI Map.** When forecasting a region's demand, this paper can refer to other regions with similar POI. The function of a region can be represented by each category of POI in the region, and the following formula can represent the similarity of POI between the two regions:

$$\gamma_i = \text{sim}(b_m, b_n) \in [0, 1], \quad (4)$$

where  $b_m$  and  $b_n$ , respectively, represent the POI vectors of region  $m$  and region  $n$ , the vector dimension is the number of POI categories, and the input value is the total number of all kinds of POI.

**3.2.3. Interconnection Map.** The transportation system is an important factor that affects spatiotemporal prediction. Although some areas are geographically far away, they can be connected by public transportation. The following formula can express public transport interoperability:

$$\delta_i = \frac{\text{con}(v_m, v_n)}{\max[\text{con}(v_m, v_n)]} \in [0, 1], \quad (5)$$

where  $\text{con}(v_m, v_n)$  indicates the accessibility of public transport in region  $m$  and region  $n$ . If there are  $n$  kinds of public transport connections between the two areas, the value equals  $n$ .

**3.2.4. Modeling Spatiotemporal Dependency by Multigraph Convolution.** This paper uses the following formula for convolution using the above three types of graphs.

$$D_{l+1} = \sigma \left( \prod_{A \in A} f(A; \theta_i) D_l W_l \right). \quad (6)$$

$D_l \in R^{|V| \times P_l}$ ,  $D_{l+1} \in R^{|V| \times P_{l+1}}$  are the feature vectors of  $|V|$  regions in layers  $l$  and  $l+1$ ,  $\sigma$  represents the activation function,  $\prod$  represents the aggregation function,  $A$  represents graph set,  $f(A; \theta_i) \in R^{|V| \times |V|}$  represents aggregation matrix based on different samples represented by  $\theta_i$ , and  $W_l \in R^{P_l \times P_{l+1}}$  represents feature transformation matrix. For instance, if  $f(A; \theta_i)$  is a polynomial function of Laplacian matrix  $L$ , it will be ChebNet; if  $f(A; \theta_i)$  is the identity matrix, it will be fully connected network.  $f(A; \theta_i)$  is  $K$ -order polynomial function of graph Laplacian matrix  $L$ . Figure 2 represents an example that a central region transforms value by graph convolution. Suppose the input of the adjacent matrix is 0 or 1, and  $L_{ij}^k \neq 0$  represents  $v_i$  reaching  $v_j$  in  $k$ -hop. In convolution operation,  $k$  defines the size of the reception field when spatial features are extracted.

Figure 2 describes the graph convolution operation. Left: the centralized region is black, the adjacent area is yellow, and the peripheral area is blue. Middle: as the Laplacian degree of the graph increases, more areas change from blue to green. Right: the output layer is the sum of graph transformations in which the degree increases from 1 to  $K$ .

The spatial variables involved in this paper include distance, POI, and public transport connection, and variables such as intersections can also be considered in future research. It models spatial correlation by feature extraction of region relationship; when the reception is small, feature extraction is concentrated in the near region. Increasing the graph Laplacian degree or stacking multiple convolutional layers will increase the reception field and capture more global correlations.

**3.3. Temporal Modeling.** LSTM network is a variant of recurrent neural network (RNN), which was proposed to solve the long-term dependence problem. As can be seen from the cellular structure diagram of LSTM in Figure 3, LSTM adopts three kinds of gating mechanisms to solve the long-term dependence problem in traditional RNN.

The three types of LSTM gating units are forget gate  $f$ , input gate  $i$ , and output gate  $o$ . The forget gate is used to control the information to be discarded by the current cell state  $C^{(t)}$ . The state update formula of the forget gate is as follows:

$$f^{(t)} = \sigma(W_f[h^{(t-1)}, d^{(t)}] + b_f). \quad (7)$$

The input gate controls how much of the input  $d^{(t)}$  of the current network will remain in the cell state  $C^{(t)}$ , and the state update formula for the input gate is

$$\begin{aligned} i^{(t)} &= \sigma(W_i[h^{(t-1)}, d^{(t)}] + b_i), \\ \widetilde{C}^{(t)} &= \tanh(W_c[h^{(t-1)}, d^{(t)}] + b_c), \\ C^{(t)} &= f^{(t)} \times C^{(t-1)} + i^{(t)} \times \widetilde{C}^{(t)}. \end{aligned} \quad (8)$$

The output gate controls how much information can be output in the current cell state. The state update formula of the output gate is

$$\begin{aligned} o^{(t)} &= \sigma(W_o[h^{(t-1)}, d^{(t)}] + b_o), \\ h^{(t)} &= o^{(t)} \times \tanh(C^{(t)}). \end{aligned} \quad (9)$$

$\sigma$  represents the sigmoid function, which will generate the vector between  $[0, 1]$  according to the input.  $\widetilde{C}^{(t)}$  represents candidate cell information;  $W_f$ ,  $W_i$ ,  $W_o$ ,  $W_c$  represent the weight coefficient matrix in the LSTM cell state update process;  $b_f$ ,  $b_i$ ,  $b_o$ ,  $b_c$  represent the bias matrix in the state update process.

This paper introduces LSTM with a contextual gate mechanism to model the correlation between observations at different time, as shown in Figure 4.

First, the information in the relevant region is considered as contextual information and is convolved by a  $K$ -order graph convolution using the corresponding graph Laplacian matrix (10). The context gating mechanism is designed to perform graph convolution before pooling in order to make the pooling contain topological information.

$$\widehat{D}^{(t)} = \left[ D^{(t)}, F^{K'}(D^{(t)}) \right], \quad t = 1, 2, \dots, T. \quad (10)$$

Second, this paper uses global average pooling  $F_{\text{pool}}$  to aggregate the information of all nodes into one node:

$$z^{(t)} = F_{\text{pool}}(\widehat{D}^{(t)}) = \frac{1}{|V|} \sum_{i=1}^{|V|} \widehat{D}_i^{(t)}, \quad t = 1, 2, \dots, T. \quad (11)$$

Third, the weight  $s$  is generated by nonlinear transformation of vector  $z$  using attention operation (12),  $\delta$  and  $\sigma$  are the ReLU and sigmoid functions, respectively.

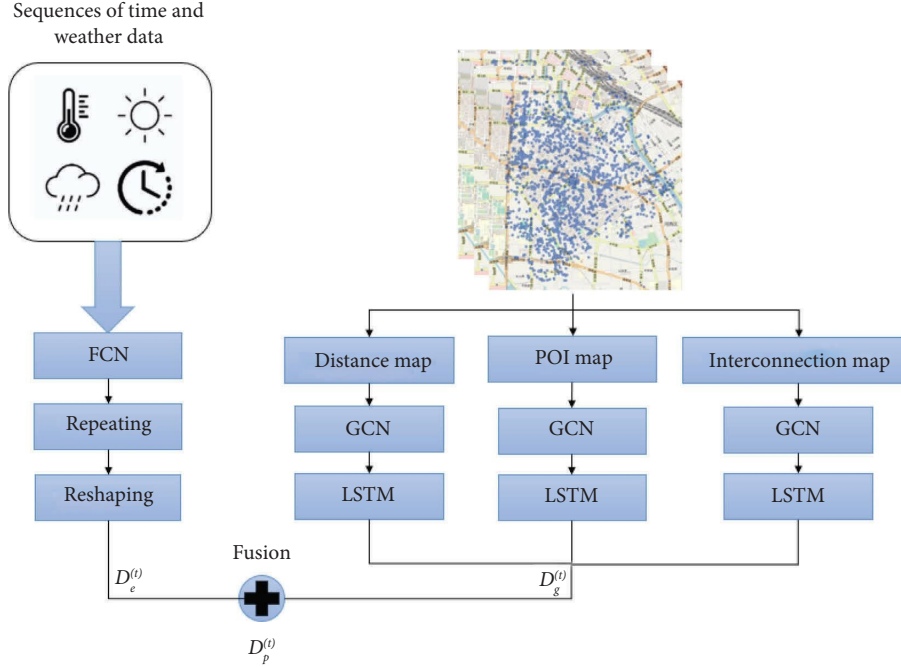


FIGURE 1: The structural framework of SGCNPM.

$$s = \sigma(W_2 \delta(W_1 z)). \quad (12)$$

Finally,  $s$  is applied to weight each time step:

$$\tilde{D}_g^{(t)} = D^{(t)} \circ s^{(t)}, \quad i = 1, 2, \dots, T. \quad (13)$$

After obtaining multiple graphs weighted by time steps, multiple graphs are fused into one graph using weight-sharing LSTM:

$$D_g^{(t)} = LSTM(\tilde{D}_i^{(1)}, \dots, \tilde{D}_i^{(T)}; W_3), \quad i = 1, \dots, |V|. \quad (14)$$

**3.4. Environment Modeling.** The model can effectively reflect the influence of environmental factors on demand for bike-sharing. The main structure is fully connected network implemented for time-series variables, including weather state, maximum temperature, minimum temperature, wind level, and air quality index. The model could output demand prediction  $D_e^{(t)}$ .

**3.5. Model Fusion.** The results of the two models are weighted and fused to obtain the demand prediction of bike-sharing, and the calculation formula is as follows:

$$D_p^{(t)} = w_g D_g^{(t)} + w_e D_e^{(t)}. \quad (15)$$

$D_p^{(t)}$  is the demand prediction of bike-sharing,  $D_g^{(t)}$  and  $D_e^{(t)}$  represent the output of two models, and  $w_g$  and  $w_e$  represent fusion weight, respectively.

## 4. Experiments and Results

Since bike-sharing entered Tianjin in February 2017, the development of the industry has gone through four stages: early stage, development stage, boom stage, and maturing stage. Initially, Tianjin also suffered from the problem of excessive and indiscriminate distribution of bike-sharing, with about 1 million bikes in the city. An accurate prediction of the demand for bike-sharing can help the Tianjin government to rationally release and dispatch bike-sharing, help bike-sharing better to play the role of green travel and slow transportation, and keep Tianjin among the advanced cities in the management of bike-sharing among companies in China.

In this section, this paper first preprocesses the data, predicts the bike-sharing demand with SGCNPM, and then compares the prediction performance of other models.

**4.1. Data Collection.** In this paper, three types of data collection work are involved: spatiotemporal variables, built environment variables, and weather variables.

**4.1.1. Spatiotemporal Variables.** The dataset utilized in this paper is extracted from Hellobike and Mobike, the top two bike-sharing service platforms in China, for two months, from May 1, 2019, to June 30, 2019. As shown in Figure 5, the studied site is in the Heping District, surrounded by solid black lines. The dataset is partitioned into 1-hour time intervals, and the investigated region is partitioned into 53 regions by  $500 \text{ m} \times 500 \text{ m}$  grids; this paper also studies grids connected to regions, so the studied region includes 53 grids, which are filled with yellow.

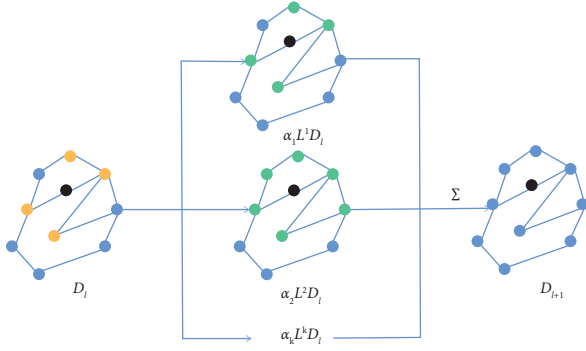


FIGURE 2: Graph convolution operation.

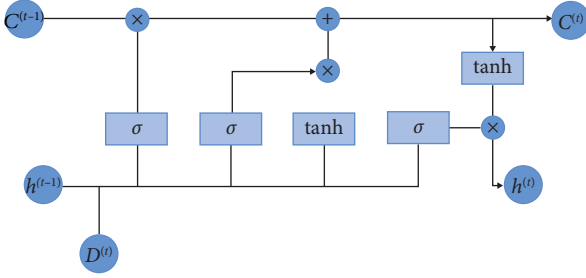


FIGURE 3: The cellular structure diagram of LSTM.

The demand dataset is divided into a 70% training set comprised of observations between May 1 and June 11, a 10% validation set consisting of the observations between June 12 and June 17, and a 20% test set comprised of observations between June 18 to June 30. Figure 6 shows the total demand of all grids on different days within the investigated training (before the red dash line), validating (between the red dash line and the green dash line), and testing (after the green dash line) period.

It can be observed from Figure 7, based on the training set, which shows the mean and variance of bike-sharing demand in different hours of the day, that the bike-sharing demand on both working days and rest days demonstrates a double-peak nature. However, the double peak on working days is steeper, and the peak on working days is higher than that on rest days. It is observed that demand is lower on rest days, which presents a challenge for short-term demand forecasting.

To verify whether the spatiotemporal variables exist in spatiotemporal correlations, this paper uses Pearson correlation to examine the correlations between the demand at the  $t$ -th time interval and spatiotemporal variables ahead of the  $t$ -th time interval, given by

$$\text{Corr}(X, Y) = \frac{E[(X - E(X))(Y - E(Y))]}{E[(X - E(X))^2]E[(Y - E(Y))^2]}, \quad (16)$$

where  $X$  and  $Y$  are two random variables with the same number of observations.

First, this paper calculates the Pearson correlations between the demand at  $t$  time in the grid  $(m, n)$  and demand at  $t - k$  time in the grid  $(m', n')$ , for all  $m, m' \in \{1, \dots, m\}$ ,  $n, n' \in \{1, \dots, N\}$ ,  $k \in \{1, 2, 3, 4\}$ . Second, this paper averages

the correlations partitioned by distances and look-back time intervals. The distance of grid  $(m, n)$  and  $(m', n')$  is denoted as the Euclidean distance between the central points of the two grids.

Figure 8 shows the average correlations between the dependent variables (the demand at  $t$  time in the grid  $(m, n)$ ) and the explanatory variables (the demand at  $t - k$  time in the grid  $(m', n')$ ). It can be observed from the figure that the average correlations decrease gradually with the increase of distance, which verifies that each area is spatially correlated with its neighbours. On the other hand, the smaller the look-back time intervals are, the more relevant the variables are. Pearson correlation confirms that the spatiotemporal variables have spatiotemporal dependencies.

## 4.2. Built Environment Variables

**4.2.1. Land Use Type.** This paper crawls the Baidu Maps' POI data through Python, which includes longitude, latitude, name, address, and administrative region. There are nine categories of POI: subway stations, parks, shopping centres, training institutions, office buildings, schools, well-known enterprises, housing estates, and comprehensive restaurants.

**4.2.2. Public Transport Accessibility.** This paper gathers public transportation data from Baidu Maps and investigates how many public transportation methods link one region to another, including subways and buses.

**4.3. Weather Variables.** This paper also gathers one-hour aggregated weather variables, including temperature, weather state, wind speed, and air quality index, during the same period from the China Meteorological Administration. This paper also gathers variables, including date, week, and holiday.

**4.4. Prediction Result.** The SGCNPM with full variables is trained on the training set and validated on the test set, respectively. Aggregation matrix  $f(A; \theta_i)$  selects the Chebyshev polynomial function with  $K$  equal 2. The model consists of five layers: an input layer, three hidden layers, and an output layer. Each hidden layer has eight hidden cells. To solve the overfitting problem, this paper introduces L2 parameter regularization. The model uses ReLU as the activation function of the graph convolution network. This paper predicts the demand for 24 hours in a day, and the output is a  $24 \times 53$  matrix. The paper analyzes the influence of epoch, time step, and batch size on the results in advance, as shown in Figure 9. Then, this paper sets the epoch, time step, and batch size to 100, 12, and 24, respectively, because these values provide the best prediction. There are 53 regions in this paper, and each node represents a map region, so this paper sets the node to 53.

The model is evaluated via the four measures of effectiveness: mean absolute error (MAE), mean absolute percentage error (MAPE), root mean squared error (RMSE), and coefficient of determination ( $R^2$ ), given by

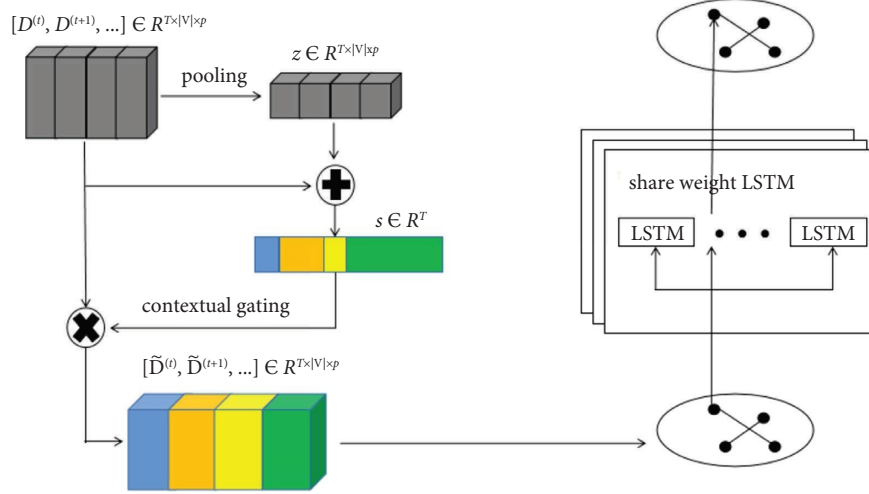


FIGURE 4: LSTM with a contextual gate mechanism.

$$\begin{aligned}
 \text{MAE} &= \frac{1}{n} \sum_{i=1}^n |y^{(i)} - \hat{y}^{(i)}|, \\
 \text{MAPE} &= \frac{1}{n} \sum_{i=1}^n \frac{|y^{(i)} - \hat{y}^{(i)}|}{y^{(i)}}, \\
 \text{RMSE} &= \sqrt{\frac{1}{n} \sum_{i=1}^n (y^{(i)} - \hat{y}^{(i)})^2}, \\
 R^2 &= 1 - \frac{\sum_{i=1}^n (y^{(i)} - \hat{y}^{(i)})^2}{\sum_{i=1}^n (y^{(i)} - \bar{y})^2},
 \end{aligned} \tag{17}$$

where  $y^{(i)}$  and  $\hat{y}^{(i)}$  are the  $i$ -th ground truth and predicted demand value, respectively,  $\bar{y}$  is the mean of  $\hat{y}^{(i)}$ , and  $n$  is the size of the test set. This paper uses MAE, RMSE, and  $R^2$  to measure the overall forecast accuracy of the entire test data and MAPE to measure the forecast performance of the model in high-demand areas and periods.

To investigate the fusion weight, the results under different weights are analyzed, and the bike-sharing demands in 53 regions are predicted. As shown in Figure 10, it is found that the model has the best prediction effect when  $w_g$  equals 0.4 and  $w_e$  equals 0.6.

This paper applies SGCNPM to forecast the demand for bike-sharing in 53 regions and find that region 7 has the best forecast effect and region 8 has the worst forecast effect; the results are shown in Figure 11 and Table 2. Figure 11 shows MAE of each region, region 7 has the smallest MAE, and region 8 has the largest MAE. Table 2 shows the average values of the predicted results of all regions, and MAE, MAPE, RMSE, and  $R^2$  are 8.209, 37.12%, 11.527, and 0.737, respectively.

Figure 12 shows the error heat maps, where deeper color and denser dots represent greater demand. It can be observed from the heat maps that the error during peak hours (e.g., 8–9 AM and 6–7 PM) is much higher than that during



FIGURE 5: The investigated region (marked with numbers).

0–1 AM and 11–12 PM, and the heat maps can accurately reflect the error. The combination of short-term demand forecast and visual analysis can help operators quickly identify areas with great errors and adjust them.

#### 4.5. Model Comparisons

**4.5.1. Comparison with Other Models.** Apart from the proposed models, other algorithms are tested. The algorithms include four traditional time-series forecasting models (i.e., HA, MA, ARIMA, and Holt exponential smoothing) and several learning/deep learning methods (i.e., ANN, LSTM, and GRU).

- (1) HA: The historical average model predicts future demand in the test set based on empirical statistics in the training set. For instance, the average demand during 7–8 AM in the grid  $(m, n)$  is estimated from all historical demand during 7–8 AM in the grid  $(m, n)$ .



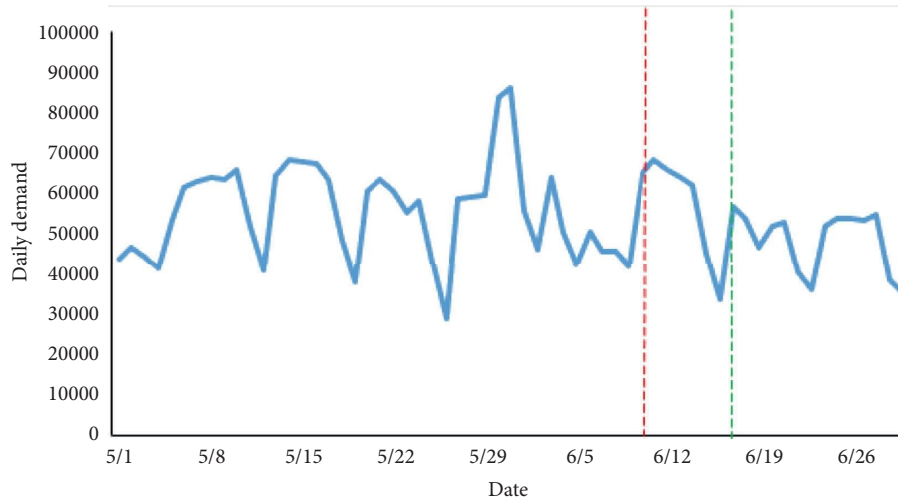


FIGURE 6: The total demand of all grids on different days.

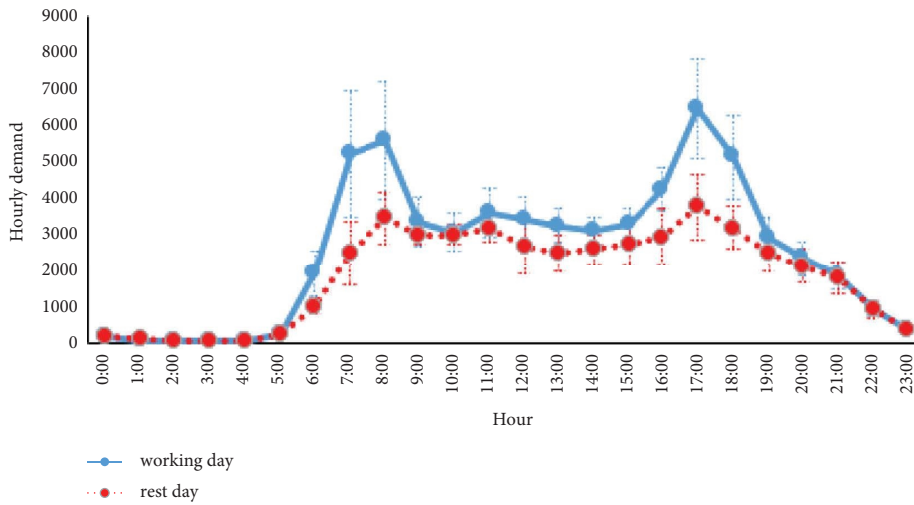


FIGURE 7: The mean and variance of bike-sharing demand at different hours of a day.

- (2) MA: The moving average model is widely used in time-series analysis. It predicts the future demand by the average of several recent historical demands. This paper uses the average of 12 previous demands in the grid  $(m, n)$  to predict the future demand in the grid  $(m, n)$ .
- (3) ARIMA: the autoregressive integrated moving average model integrates the autoregressive (AR), integrated (I), and MA parts and takes into account the trend, periodicity, and nonstationary characteristics of the dataset [33].
- (4) Holt exponential smoothing: Holt exponential smoothing model adds a trend smoothing coefficient  $\beta$  based on a simple exponential smoothing coefficient  $\alpha$ , which is also called the two-parameter smoothing method.
- (5) ANN: The artificial neural network [34, 35] uses all the variables of a specific grid  $(m, n)$ , including historical demand, travel time rate, hourly state, weekly state, and weather variables with

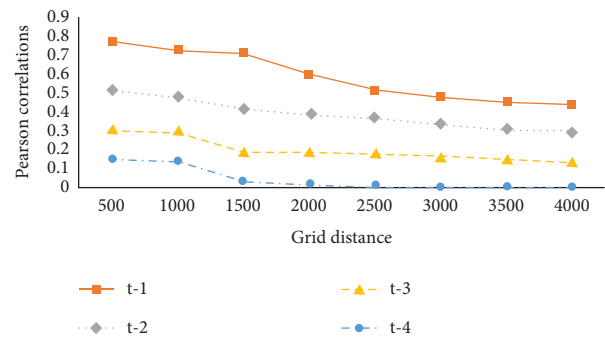


FIGURE 8: Pearson correlations.

retrospective time windows, to predict the future demand in the grid  $(m, n)$ . The neural network cannot distinguish the variables of different time, so it cannot capture the time correlation.

- (6) LSTM: In LSTM [36], all variables in the grid  $(m, n)$  are reconstructed into a matrix, where one axis is the time step (the size of which equals look-back time

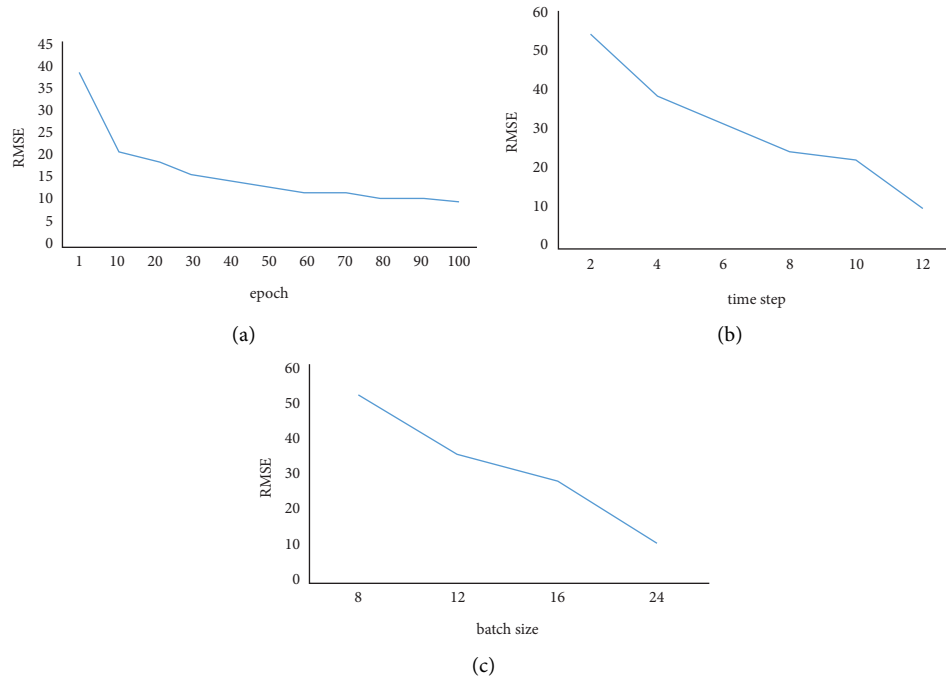


FIGURE 9: (a) Epoch. (b) Time step. (c) Batch size.

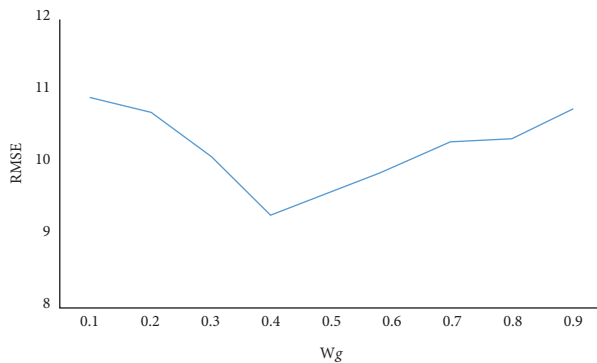


FIGURE 10: The prediction results under different weights.

window  $K=12$ ), and the other axis is the feature category. In this way, all the features used in SGCNPM are sent to LSTM for training. LSTM considers time dependencies but does not capture spatial dependencies.

- (7) GRU: gated recurrent unit is a kind of neural network which performs similar to LSTM but is computationally cheaper.

The aforementioned deep learning methods have the same input features (same category and look-back time windows) as the SGCNPM, and the four time-series models (HA, MA, ARIMA, and Holt exponential smoothing) make use of the same time series. Moreover, the deep learning models (ANN, LSTM, and GRU) are also trained with 100 epochs. Before model training and validation, all data are standardized to the range  $[0, 1]$  through the same standardization.

The research uses Python 3.7 with TensorFlow 1.14.0 [37], Keras [38], and scikit-learn [39] for comparing the models.

For all methods, this paper makes predictions on the validation set. Table 3 compares the results of different forecasting methods in twenty runs. This paper observes the following phenomena in the results. (1) The prediction results of deep learning methods, including ANN, LSTM, GRU, and SGCNPM, are superior to other models. (2) SGCNPM achieves the best results regarding all the metrics on the same dataset. (3) The deep learning methods take longer than other methods, and SGCNPM takes the longest time. (4) MAPE obtained by all methods is great.

**4.5.2. SGCNPM with Different Factors.** To verify the need to consider multiple factors, this paper rebuilds SGCNPM, which removes the weather variable, land use type variable, and public transport accessibility variable, respectively. The results are shown in Table 4. After removing the public transport accessibility variable, the prediction accuracy decreases the most, MAE increases from 8.209 to 11.223, and RMSE increases from 11.527 to 14.017. As shown in Table 4, removing any variable will lead to an increase in prediction error, reflecting the importance of each variable.

Furthermore, to analyze the accuracy of the SGCNPM in different periods, this paper predicts the demand for bike-sharing separately during the day and at night. The model is recalibrated. The results are shown in Tables 5 and 6. Several conclusions are put forward here.

It is necessary to consider all variables. The prediction results considering all variables are the best in the daytime or at night. The MAE, MAPE, RMSE, and  $R^2$  are 8.377, 20.21%, 10.648, and 0.738, respectively, in the daytime; the MAE, MAPE, RMSE, and  $R^2$  are 0.866, 32.19%, 1.203, and 0.811,

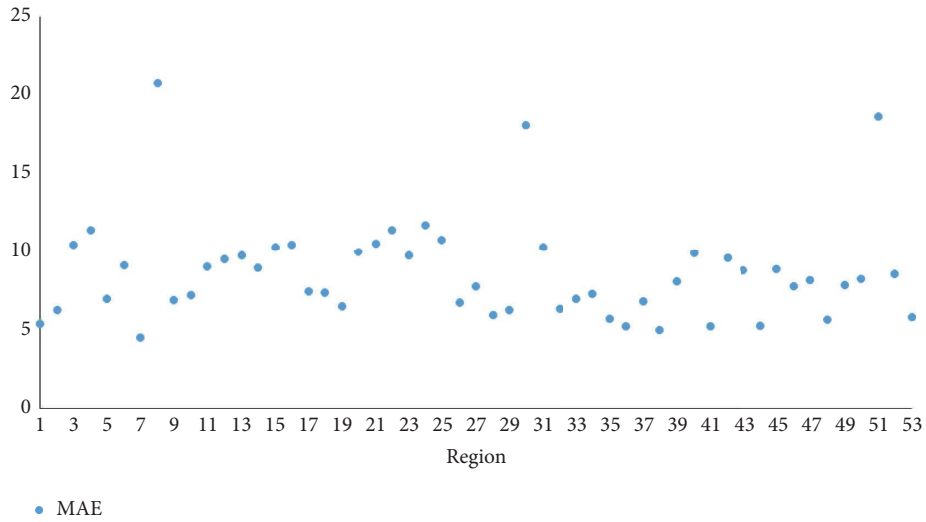


FIGURE 11: MAE of each region.

TABLE 2: Predictions for region 7 and region 8.

	Region	MAE	MAPE (%)	RMSE	$R^2$
Best	7	4.535	35.66	5.898	0.780
Worse	8	20.75	40.00	22.635	0.634
Average	[1, 53]	8.209	37.12	11.527	0.737

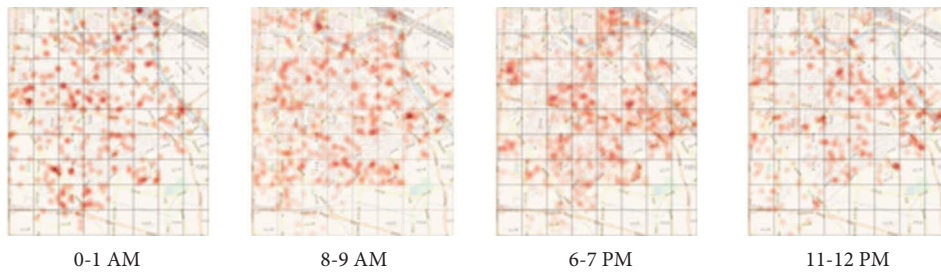


FIGURE 12: Heat maps of the errors.

TABLE 3: Result comparison of different models for bike-sharing demand forecasting.

Model	MAE	MAPE (%)	RMSE	$R^2$	Time (s)
HA	8.375	38.80	12.421	0.183	1
MA	14.830	478.34	17.370	-0.598	1
ARIMA	9.870	55.79	13.954	0.593	4
Holt exponential smoothing	10.000	56.87	14.174	0.580	4
ANN	13.333	590.60	17.042	-0.539	18
LSTM	10.900	91.01	14.573	-0.125	656
GRU	9.526	70.47	12.264	0.383	727
SGCNPM	8.209	37.12	11.527	0.737	1320

TABLE 4: Result comparison of removing any variable.

Removed variable	MAE	MAPE (%)	RMSE	$R^2$	Time (s)
Weather	10.903	70.22	13.877	0.762	1310
Land use type	8.460	93.88	11.228	0.851	1310
Public transport accessibility	11.223	123.76	14.017	0.792	1310
SGCNPM	8.209	37.12	11.527	0.737	1320

TABLE 5: Result comparison of removing any variable (after removing the data during 23:00–6:00).

Removed variable	MAE	MAPE (%)	RMSE	$R^2$	Time (s)
Weather	9.703	22.93	12.027	0.680	1310
Land use type	12.101	33.27	14.689	0.526	1310
Public transport accessibility	11.564	39.53	13.792	0.702	1310
SGCNPM	8.377	20.21	10.648	0.738	1320

TABLE 6: Result comparison of removing any variable (after removing the data during 6:00–23:00).

Removed variable	MAE	MAPE (%)	RMSE	$R^2$	Time (s)
Weather	1.446	58.84	2.193	0.517	1310
Land use type	0.877	32.38	1.253	0.644	1310
Public transport accessibility	1.262	75.12	1.449	0.789	1310
SGCNPM	0.866	32.19	1.203	0.811	1320

respectively, at night. The MAE and RMSE of the night are less than those of the daytime, while the MAPE is greater than that of the daytime because the bike-sharing demand is greater in the daytime than at night.

## 5. Conclusions

This paper first analyzes the travel behaviors of bike-sharing in the Heping District of Tianjin. SGCNPM forecasts the demand for bike-sharing in various urban areas, providing technical support for bike-sharing dispatching. SGCNPM is a fusion of two models, which are the built environment variable model and the weather environment variable model. The built environment variable model integrates distance map, POI map, and interconnection map. The weather environment variable model considers the influence of weather state, maximum temperature, minimum temperature, wind level, and air quality index. The model can better reflect the influence of spatiotemporal correlation, built environment, and weather on demand for bike-sharing. Through the comparison of forecasting results, this paper finds that the forecasting accuracy of SGCNPM is better than that of HA, MA, ARIMA, Holt exponential smoothing, ANN, LSTM, and GRU. In addition, this paper discovers that the forecast accuracy will decline regardless of any time when built environment variables or weather variables are missing from the model. It is demonstrated that the built environmental variables and weather variables are crucial for forecasting the demand for bike-sharing. Future research will deeply analyze how to allocate and dispatch bike-sharing based on bike-sharing demand forecast accurately and continue to improve the model.

## Data Availability

The demand data extracted from Hellobike and Mobike to support the results of this study have not been made available because of the requirements in the confidentiality agreement.

## Conflicts of Interest

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

## Acknowledgments

This study was supported by the Fundamental Research Funds for the Central Universities (2021PT207), the National Natural Science Foundation of China (52172312 and 71931003), and the Research Foundation for Youth Scholars of Beijing Technology and Business University.

## References

- [1] TQ. Gu, I. Kim, and G. Currie, "To be or not to be dockless: empirical analysis of dockless bikeshare development in China," *Transportation Research Part A Policy Pract*, vol. 119, pp. 122–147, 2019.
- [2] Beijing Municipal Commission of Transportation, *Public Notice on the Supervision of the Operation and Management of the Internet Rental Bicycle Industry in the First Half of 2019*, Beijing Municipal Commission of Transportation, 2019, [https://jtw.beijing.gov.cn/xxgk/tzgg/201907/t20190731\\_1279822.html](https://jtw.beijing.gov.cn/xxgk/tzgg/201907/t20190731_1279822.html).
- [3] E. Heinen, B. Van Wee, and K. Maat, "Commuting by bicycle: an overview of the literature," *Transport Reviews*, vol. 30, no. 1, pp. 59–96, 2010.
- [4] M. Hyland, Z. Hong, H. K. R. d F. Pinto, and Y. Chen, "Hybrid cluster-regression approach to model bikeshare station usage," *Transportation Research Part A: Policy and Practice*, vol. 115, pp. 71–89, 2018.
- [5] HT. Yu and Z. R. Peng, "Exploring the spatial variation of ridesourcing demand and its relationship to built environment and socioeconomic factors with the geographically weighted Poisson regression," *Journal of Transport Geography*, vol. 75, pp. 147–163, 2019.
- [6] E. Eren and V. E. Uz, "A review on bike-sharing: The factors affecting bike-sharing demand," *Sustain Cities Society*, vol. 54, 2020.
- [7] M. Kabak, M. Erbaş, C. Çetinkaya, and E. Özceylan, "A GIS-based MCDM approach for the evaluation of bike-share stations," *Journal of Cleaner Production*, vol. 201, pp. 49–60, 2018.
- [8] D. Sun and X. Q. Ding, "Spatiotemporal evolution of ride-sourcing markets under the new restriction policy: a case study in Shanghai," *Transportation Research Part A: Policy and Practice*, vol. 130, pp. 227–239, 2019.
- [9] V. E. Sathishkumar, J. Park, and Y. Cho, "Using data mining techniques for bike sharing demand prediction in metropolitan city," *Computer Communications*, vol. 153, pp. 353–366, 2020.
- [10] SY. Li, CG. Zhuang, ZZ. Tan, F. Gao, Z. Lai, and Z. Wu, "Inferring the trip purposes and uncovering spatio-temporal activity patterns from dockless shared bike dataset in

- Shenzhen, China,” *Journal of Transport Geography*, vol. 91, Article ID 102974, 2021.
- [11] S. Sohrabi, R. Paleti, L. Balan, and M. Cetin, “Real-time prediction of public bike sharing system demand using generalized extreme value count model,” *Transportation Research Part A: Policy and Practice*, vol. 133, pp. 325–336, 2020.
- [12] F. Reynaud, A. Faghih-Imani, and N. Eluru, “Modelling bicycle availability in bicycle sharing systems: a case study from Montreal,” *Sustainable Cities and Society*, vol. 43, pp. 32–40, 2018.
- [13] S. Hu, C. Xiong, Z. Liu, and L. Zhang, “Examining spatio-temporal changing patterns of bike-sharing usage during COVID-19 pandemic,” *Journal of Transport Geography*, vol. 91, Article ID 102997, 2021.
- [14] E. Collini, P. Nesi, and G. Pantaleo, “Deep learning for short-term prediction of available bikes on bike-sharing stations,” *IEEE Access*, vol. 9, pp. 124337–124347, 2021.
- [15] A. Mehdizadeh Dastjerdi and C. Morency, “Bike-sharing demand prediction at community level under COVID-19 using deep learning,” *Sensors*, vol. 22, no. 3, p. 1060, 2022.
- [16] X. Chang, J. Wu, H. Sun, G. HdA. Correia, and J. Chen, “Relocating operational and damaged bikes in free-floating systems: a data-driven modeling framework for level of service enhancement,” *Transportation Research Part A: Policy and Practice*, vol. 153, pp. 235–260, 2021.
- [17] W. L. Shang, J. Chen, H. Bi, Y. Sui, Y. Chen, and H. Yu, “Impacts of COVID-19 pandemic on user behaviors and environmental benefits of bike sharing: a big-data analysis,” *Applied Energy*, vol. 285, Article ID 116429, 2021.
- [18] Y. Yang, A. Heppenstall, A. Turner, and A. Comber, “A spatiotemporal and graph-based analysis of dockless bike sharing patterns to understand urban flows over the last mile,” *Computers, Environment and Urban Systems*, vol. 77, Article ID 101361, 2019.
- [19] Y. Ai, Z. Li, M. Gan et al., “A deep learning approach on short-term spatiotemporal distribution forecasting of dockless bike-sharing system,” *Neural Computing and Applications*, vol. 31, no. 5, pp. 1665–1677, 2019.
- [20] Y. Li and B. Shuai, “Origin and destination forecasting on dockless shared bicycle in a hybrid deep-learning algorithms,” *Multimedia Tools and Applications*, vol. 79, no. 7–8, pp. 5269–5280, 2020.
- [21] Y. Xu, D. Chen, X. Zhang et al., “Unravel the landscape and pulses of cycling activities from a dockless bike-sharing system,” *Computers, Environment and Urban Systems*, vol. 75, pp. 184–203, 2019.
- [22] A. Li, P. Zhao, Y. Huang, K. Gao, and K. W. Axhausen, “An empirical analysis of dockless bike-sharing utilization and its explanatory factors: case study from Shanghai, China,” *Journal of Transport Geography*, vol. 88, Article ID 102828, 2020.
- [23] J. Dong, B. Chen, LN. He et al., “A spatio-temporal flow model of urban dockless shared bikes based on points of interest clustering,” *ISPRS International Journal of Geo-Information*, vol. 8, no. 8, p. 345, 2019.
- [24] Q. Yan, K. Gao, LJ. Sun, and M. H. Shao, “Spatio-temporal usage patterns of dockless bike-sharing service linking to a metro station: a case study in shanghai, China,” *Sustainability*, vol. 12, no. 3, p. 851, 2020.
- [25] WW. Li, SH. Wang, X. Zhang, Q. Jia, and Y. Tian, “Understanding intra-urban human mobility through an exploratory spatiotemporal analysis of bike-sharing trajectories,” *International Journal of Geographical Information Science*, vol. 34, no. 12, pp. 2451–2474, 2020.
- [26] X. Li, Y. Xu, Q. Chen, L. Wang, X. Zhang, and W. Shi, “Short-term forecast of bicycle usage in bike sharing systems: a spatial-temporal memory network,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 8, pp. 10923–10934, 2022.
- [27] C. Zeng, C. Ma, K. Wang, and Z. Cui, “Predicting vacant parking space availability: a DWT-Bi-LSTM model,” *Physica A: Statistical Mechanics and Its Applications*, vol. 599, Article ID 127498, 2022.
- [28] C. Ma, G. Dai, and J. Zhou, “Short-term traffic flow prediction for urban road sections based on time series analysis and LSTM\_BILSTM method,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 6, pp. 5615–5624, 2022.
- [29] H. Ziheng, W. Dujuan, Y. Yunqiang, and L. Xiang, “A spatiotemporal bidirectional attention-based ride-hailing demand prediction model: a case study in beijing during COVID-19,” *IEEE Transactions on Intelligent Transportation Systems*, 2021.
- [30] K. Zhu, S. Zhang, W. Zhang, and Z. Zhang, “A novel hybrid deep learning model for taxi demand forecasting based on decomposition of time series and fusion of text data,” *Journal of Intelligent and Fuzzy Systems*, vol. 41, no. 2, pp. 3355–3371, 2021.
- [31] D. Kim, H. Shin, H. Im, and J. Park, “Factors Influencing Travel Behaviors in Bike-Sharing,” in *Proceedings of the Transportation Research Board 91st Annual Meeting*, New York, NY, USA, May 2012.
- [32] X. Geng, Y. Li, and L. Wang, “Spatiotemporal Multi-Graph Convolution Network for Ride-Hailing Demand Forecasting,” in *Proceedings of the 33rd AAAI Conference on Artificial Intelligence*, pp. 3656–3663, Virtually, May 2019.
- [33] G. E. P. Box and D. A. Pierce, “Distribution of residual autocorrelations in autoregressive-integrated moving average time series models,” *Journal of the American Statistical Association* 1970, vol. 65, no. 332, pp. 1509–1526, 1970.
- [34] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, “Learning Internal Representations by Error Propagation,” *Parallel Distributed Processing*, MIT Press, Cambridge, MH, USA, 1985.
- [35] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, “Learning representations by back-propagating errors,” *Cogn. Model.*, vol. 5, p. 1, 1988.
- [36] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [37] M. Abadi, A. Agarwal, and P. Barham, “TensorFlow: large-scale machine learning on heterogeneous systems,” 2015, <https://arxiv.org/abs/1603.04467>.
- [38] F. Chollet, “Keras,” 2015, <https://github.com/keras-team/keras>.
- [39] F. Pedregosa, G. Varoquaux, and A. Gramfort, “Scikit-learn: machine learning in Python,” *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.
- [40] J. Ke, H. Zheng, H. Yang, and X. M. Chen, “Short-term forecasting of passenger demand under on-demand ride services: a spatio-temporal deep learning approach,” *Transportation Research Part C: Emerging Technologies*, vol. 85, pp. 591–608, 2017.