

The station-free sharing bike demand forecasting with a deep learning approach and large-scale datasets



Chengcheng Xu*, Junyi Ji, Pan Liu

Jiangsu Key Laboratory of Urban ITS, Southeast University, Si Pai Lou #2, Nanjing 210096, China

Jiangsu Province Collaborative Innovation Center of Modern Urban Traffic Technologies, Southeast University, Si Pai Lou #2, Nanjing 210096, China

School of Transportation, Southeast University, Si Pai Lou #2, Nanjing 210096, China

ARTICLE INFO

Keywords:

Sharing mobility
Station-free bike sharing
Travel demand forecasting
Bike sharing
Deep learning
Long short-term memory networks
Big data

ABSTRACT

The station-free sharing bike is a new sharing traffic mode that has been deployed in a large scale in China in the early 2017. Without docking stations, this system allows the sharing bike to be parked in any proper places. This study aimed to develop a dynamic demand forecasting model for station-free bike sharing using the deep learning approach. The spatial and temporal analyses were first conducted to investigate the mobility pattern of the station-free bike sharing. The result indicates the imbalanced spatial and temporal demand of bike sharing trips. The long short-term memory neural networks (LSTM NNs) were then developed to predict the bike sharing trip production and attraction at TAZ for different time intervals, including the 10-min, 15-min, 20-min and 30-min intervals. The validation results suggested that the developed LSTM NNs have reasonable good prediction accuracy in trip productions and attractions for different time intervals. The statistical models and recently developed machine learning methods were also developed to benchmark the LSTM NN. The comparison results suggested that the LSTM NNs provide better prediction accuracy than both conventional statistical models and advanced machine learning methods for different time intervals. The developed LSTM NNs can be used to predict the gap between the inflow and outflow of the sharing bike trips at a TAZ, which provide useful information for rebalancing the sharing bike in the system.

1. Introduction

The sharing bike is a sustainable and environmentally friendly urban traffic mode. During the past decade, the bike sharing system has been widely deployed in numerous cities worldwide. In the early 2017, a new kind of station-free bike sharing has been deployed in a large scale in several cities in China (see Fig. 1). This new station-free bike sharing system is quite different from the conventional bike sharing system. Without any docking stations in this bike sharing system, the bike can be parked in any proper places. As shown in Fig. 1(a), the built-in GPS tracking module allow riders to find and rent nearby bikes by using their smartphone apps. When finding the nearby bikes, riders can unlock the intelligent lock of a bike by scanning the quick response (QR) code printed on the bike body.

Due to the high freedom and conveniences, the station-free bike sharing system has attracted large number of bike riders for commuting trips. It becomes one of the useful solutions for traffic congestions and last-mile problem in urban transportation systems. However, like the conventional bike sharing system, it also suffers from the problem of fluctuating spatial and temporal demand, leading to inefficient bike repositioning and high operating costs in the bike sharing rebalance (Pal and Zhang, 2017). The dynamic

* Corresponding author at: Jiangsu Key Laboratory of Urban ITS, Southeast University, Si Pai Lou #2, Nanjing 210096, China.

E-mail addresses: xuchengcheng@seu.edu.cn (C. Xu), liupan@seu.edu.cn (P. Liu).

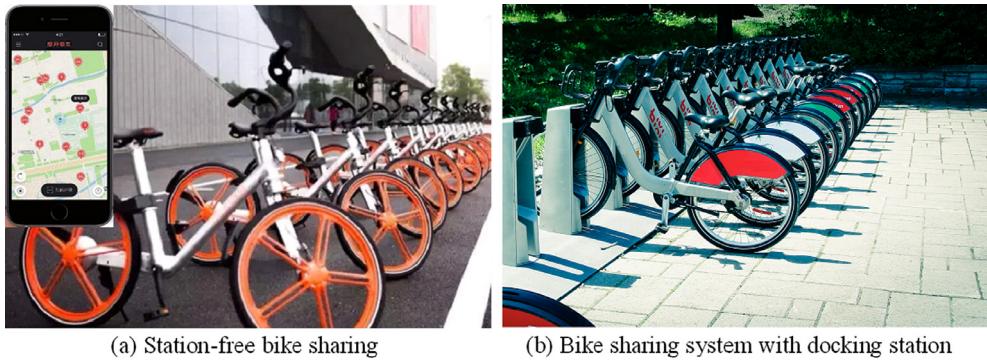


Fig. 1. Station-free bike sharing and conventional bike sharing with docking stations.

demand forecasting is critical important for rebalancing the sharing bikes in different areas (Liu et al., 2018). This study investigated the mobility pattern of the station-free bike sharing based on the OD matrices data from one of the bike sharing operation companies for several consecutive days. The dynamic demand forecasting models were developed to predict the travel demand in short time.

To improve the performance of bike sharing with docking stations, numerous studies have been conducted to understand the contributory factors affecting the demand of station-based bike sharing in recent years (Beecham and Wood, 2014; Gebhart and Noland, 2014; Zhao et al., 2014, 2015; Fishman et al., 2015; Campbell et al., 2016; El-Assi et al., 2017; Feng and Wang, 2017; Fournier et al., 2017). The current studies on travel demand of station-based bike sharing have generally focused on two major issues, including the investigations of bike sharing travel patterns (Gebhart and Noland, 2014; Beecham and Wood, 2014; Zhao et al., 2015; Fishman et al., 2015; Campbell et al., 2016) and the forecasting of bike sharing trip frequency (Zhao et al., 2014; Fournier et al., 2017; El-Assi et al., 2017; Feng and Wang, 2017). In the studies focused on the first issue, the travel survey and smart card data were used to investigate the travel patterns of station-based bike sharing, including the riders' choice preferences, travel time, trip duration, and trip purpose. For example, Campbell et al. (2016) developed a multinomial logit model to investigate the contributing factors to the choice of sharing bike based on a stated preference survey conducted in the four main urban districts of Beijing. The results suggested that trip distance, temperature, precipitation and poor air quality negatively affect the choice to switch from other transportation mode to sharing bike. Gebhart and Noland (2014) used the hourly weather data to evaluate the effects of weather conditions on the travel patterns of bike sharing. The results indicated that cold temperatures, rain, and high humidity levels reduce both the likelihood of using sharing bike and the duration of trips.

In the studies about the prediction of bike sharing trip frequency, different modeling methods and data were used to identify the contributory factors to the number of bike sharing trips (Zhao et al., 2014; Fournier et al., 2017; El-Assi et al., 2017; Feng and Wang, 2017). Zhao et al. (2014) investigated the effects of urban features and bike sharing system characteristics on daily sharing bike trip frequency. A partial least squares regression model was developed to link the daily trip number with urban population, government expenditure, the number of bike sharing members, the number of sharing bikes and the number of docking stations. El-Assi et al. (2017) evaluated the effects of built environment and weather on bike sharing demand in Toronto. The distributed lag model was used to link the daily station-level bike sharing trip number with land use, built environment and weather conditions. The results showed that road network configuration, bike infrastructure, and temperature significantly affect the bike trip frequency. Fournier et al. (2017) developed a sinusoidal model to predict the pattern of seasonal sharing bicycle demand. The evaluation results showed that the sinusoidal model is capable of estimating monthly average daily sharing bike trip frequency and average annual daily sharing bike trip frequency.

The findings from previous studies provide valuable insights in understanding of the contributing factors to the travel demand of station-based bike sharing. However, to the best of our knowledge, we are unable to find studies that have been conducted to investigate the travel demand of the station-free bike sharing. The dynamic demand forecasting model for station-free bike sharing is important because it provides useful information to develop effective and timely rebalance strategies to increase the operational efficiency of the station-free bike sharing system.

This study aimed to develop a citywide dynamic demand forecast model for station-free bike sharing system using deep learning approach with large-scale trip data. The station-free sharing bike trip data were collected at a citywide scale from one station-free sharing bike operation company. This study has the potential to contribute to the field of sharing bike by: (1) revealing the mobility pattern of the station-free bike sharing at a citywide scale; and (2) using deep learning approach to predict the travel demand of station-free bike sharing for large-scale road network at a citywide scale. So far, the development of station-free sharing bike demand forecast model has not been identified in previous studies. The results of this study have the potential to provide useful information for station-free sharing bike rebalance and to improve the operational efficiency of the station-free bike sharing system.

Recently, deep learning approaches have been increasingly used in different fields of transportation engineering (Jiang et al., 2014; Chen et al., 2017; Liu et al., 2018; Zhang et al., 2018; Wu et al., 2018; Dabiri and Heaslip, 2018; Kanarachos et al., in press). A number of studies have used deep learning approaches for travel demand forecasting, such the passenger demand of on-demand ride service and high-speed rail (Jiang et al., 2014; Chen et al., 2017; Ke et al., 2017). Ke et al. (2017) proposed a novel deep learning approach of convolutional long short-term memory network to forecast short-term passenger demand of on-demand ride service. The

comparison results suggested that the proposed method provides better predictive performance than both conventional time-series prediction models and state-of-art machine learning algorithms. Chen et al. (2017) used an ensemble learning approach to investigate the ride-splitting choices of on-demand ride services. Four types of on-demand ride services were considered, including Taxi Hailing Service, Express, Private Car Service, and Hitch. The results suggested that travel time, trip costs, trip length, waiting time fee, travel time reliability are the main factors affecting the ride-splitting choice of on-demand ride services.

In addition to the travel demand forecasting, the deep learning approaches have recently been used in the short-term traffic flow prediction (Lv et al., 2015; Duan et al., 2016; Polson and Sokolov, 2017; Liu and Chen, 2017; Zhao et al., 2017), and traffic speed prediction (Ma et al., 2015a, 2015b; Yu et al., 2017; Ma et al., 2017). For example, Lv et al. (2015) developed a short-term traffic flow forecasting model based on the stacked autoencoder model. The results showed that the proposed method for traffic flow prediction has better predictive performance than commonly used machine learning algorithms. Ma et al. (2017) proposed a convolutional neural network based method for traffic speed prediction. The proposed model learns large-scale, network-wide traffic speed as images. The results showed that the proposed method outperforms other prevailing algorithms by an average accuracy improvement of 42.9%.

In this study, the long short-term memory neural network (LSTM NN) was used to develop the dynamic demand forecast model for station-free bike sharing system. When modeling the time-series data in transportation engineering, the conventional artificial neural network (ANN) is one of the most commonly used artificial intelligence algorithms. However, ANN is unable to fully capture the characteristics of time-series data as ANN does not account for the temporal dependencies in the model structure. To overcome the limitation associated with ANN, the feed forward deep neural networks have been proposed, such as the recurrent neural network (RNN). By recurrently connecting hidden layers at different timestamps, RNN can accounts for the temporal dependencies and produce good predictive performance on time-series data. However, Traditional RNN is not suitable to fit time-series data with very long time lags. RNNs cannot well capture long time dependencies in time-series data due to the problems of vanishing and exploding gradients (Hochreiter and Schmidhuber, 1997). Moreover, the traditional RNNs are difficult to find optimal window size in modeling time-series data, as they rely on the predetermined time lags to learn the temporal sequence processing (Ma et al., 2015a, 2015b; Zhao et al., 2017). To overcome the limitations associated with RNN in modeling the time-series data with long-term dependencies, the LSTM NN was used to fit the time-series data of station-free sharing bike.

2. Data sources

The used data were collected from the downtown area of the Nanjing City which is the capital city of the Jiangsu province in China. Nanjing is a large-sized city that is located on the east coast of China. The population of Nanjing City was 8.27 million and the total area was 6597 km² in 2016. Fig. 2 illustrates the downtown area of Nanjing and the traffic analysis zone (TAZ) boundary for the downtown area. TAZ is the geographical division system for transportation planning (Xu et al., 2017). The TAZ boundary was collected from the transportation system planning document of Nanjing City. There are a total of 118 TAZs in the downtown area of the Nanjing City (see Fig. 2).

The station-free bike sharing data from one of the operation companies were collected from a large-scale multi-processing crawler developed by python. The multi-processing crawler was executed for about one month on a powerful workstation with 16 cores and 32 threads (two Intel(R) Xeon(R) CPUs E5-2620 v4 @ 2.10–3.00 GHz, and 32 GB RAM. To develop the web crawler, the open

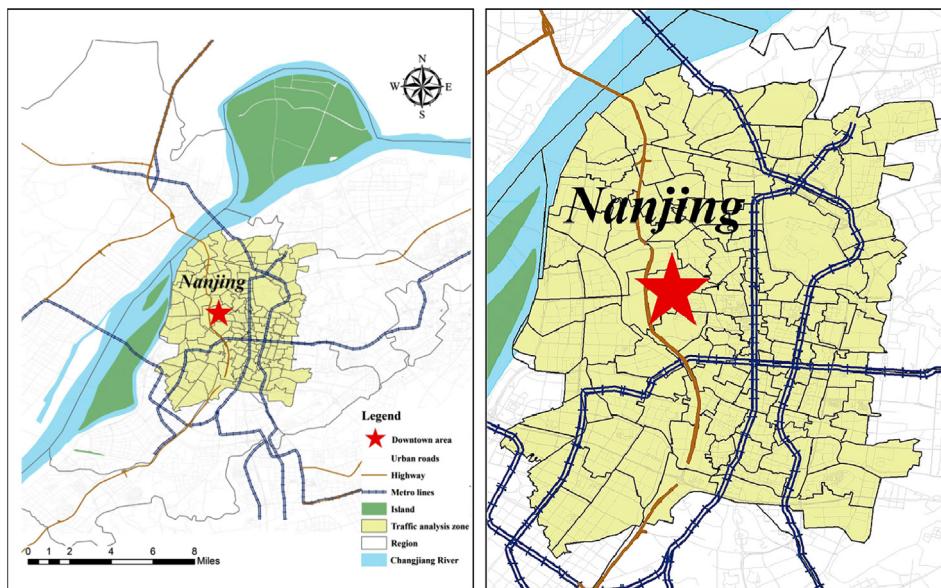


Fig. 2. The downtown area of Nanjing and the traffic analysis zone boundary.

application programming interface (API) of station-free bike sharing mobile application was first identified by using web debug tool which can capture network traffic on the Internet. Based on the identified API, the web crawler can simulate a request of finding bikes sent by a mobile application. The sever will then return the json formatted data of geo-location information of each stopped sharing bike within certain area. The downtown area of the Nanjing City was divided into 8800 grids with the equal area of 100×100 m. The developed crawler will continuously request the geo-location information of sharing bikes on these 8800 grids, and record the returned information.

Data between June 19th, 2017 and July 2nd, 2017 were used in the following analysis, as the extracted data during these 14 consecutive days have the lowest missing rate. A typical sharing bike geo-location information sample is given as {"2017/06/23 21:37:34", "object": {"bikeIds": "0250045914#", "distX": 118.73578455079466, "distY": 32.04531884191313}}, which records the time and location of each stopped sharing bike. The data set contained 0.718 billion geo-location information samples in the downtown area of Nanjing during the selected 14 days. The Hash partition algorithm was used to extract the time and GPS locations of the origin and destination point of each trip for each sharing bike. Hash partitioning algorithm is a method to separate out information and to spread them evenly in sub-tables. This algorithm evenly distributes rows among partitions, giving partitions approximately the same size. It is an easy-to-use alternative to distribute the data to space out the load. It can be used to increase the efficiency of matching queries when the ranges are not applicable.

The time, bike ID and GPS position were used to extract trip data from the raw geo-location information sample. The origin, destination, start time and end time associated with each trip were extracted by scanning whether the GPS position of one bike changed between two continuous samples. The GIS tool was then used to match the trip data with TAZs and time to generate the production and attraction data. The matched data were generated for four different time intervals, including the 10-min, 15-min, 20-min and 30-min intervals.

To increase the model predictive performance, the weather data, air quality, land use patterns were also collected. The hourly weather conditions data were obtained from Nanjing Meteorological Bureau. The hourly air quality data were collected from the National Environmental Monitoring center of China. The online point of interest (POI) data were extracted to reflect the land use pattern of each TAZ (Bao et al., 2017). The POI data were extracted from the Baidu map API.

3. Methodology

The LSTM NN was used to predict the bike sharing trip production and attraction at a TAZ in the next time interval based on the several past observations at this TAZ. Compared with the commonly used modeling techniques for time series data, the LSTM NN has the following advantages. First, as an advanced deep learning approach, the LSTM NN can better capture the non-linear relationship in time series data than traditional models (Ma et al., 2015a, 2015b). In addition, LSTM NN can overcome the limitation associated with RNN in terms of vanishing and exploding gradients when modeling the long time dependencies in time series data (Hochreiter and Schmidhuber, 1997; Zhao et al., 2017). Finally, previous studies have demonstrated that the LSTM NN has better performance than commonly used time-series modelling approaches (Ma et al., 2015a; Zhao et al., 2017).

The LSTM NN is an extension of RNN. To present the LSTM NN, we would first briefly introduce RNN. It is a recently developed deep learning method for time-series data modeling. The RNNs are capable of over-coming the previously inherent limitation of traditional ANN, i.e., not considering the temporal dependencies. As shown in Fig. 3(a), RNN accounts for the temporal correlation by recurrently connecting hidden layers at different timestamps (Simoncini et al., 2018). RNNs have good performance in modeling the nonlinear time series data in an effective way. However, due to the problems of vanishing gradient and exploding gradient, the

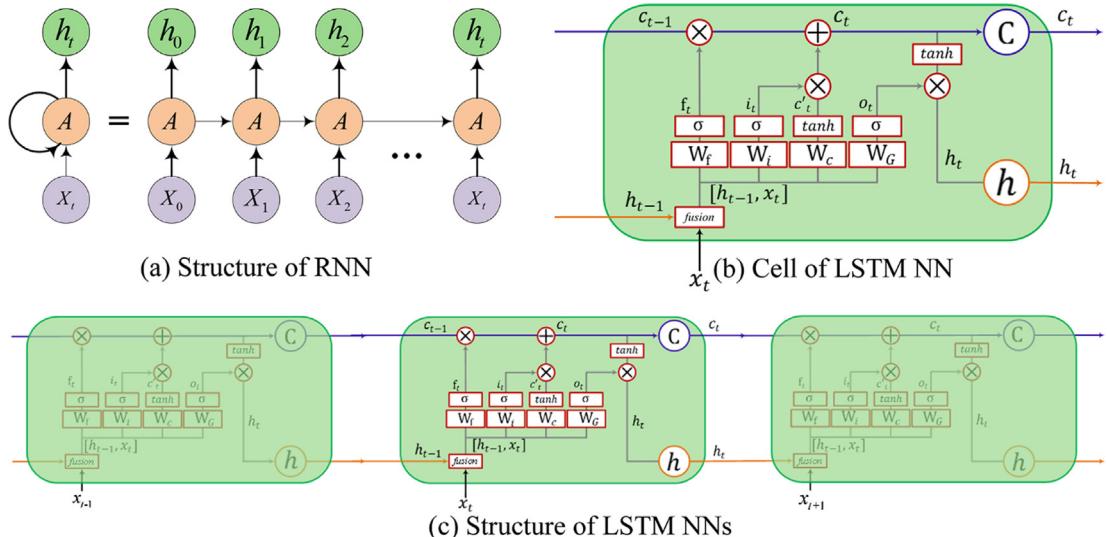


Fig. 3. Illustrations of the recurrent neural network and LSTM neural network.

traditional RNNs cannot well fit the time series with very long time lags (Hochreiter and Schmidhuber, 1997; Ma et al., 2015a, 2015b; Zhao et al., 2017).

LSTM NNs are specifically developed to overcome the limitations of RNNs in modeling the long-term dependencies. A typical LSTM NN is composed of an input layer, one or more hidden layers, and an output layer. One important component of the LSTM NN is the memory cell contained in the hidden layer(s). As shown in Fig. 3(b), each memory cell has three gates including a forget gate f_t , an input gate i_t and an output gate o_t . These three gates of memory cell are used to control and change the cell state C_t .

The forget gate f_t determines which information should be removed from the previous cell state C_{t-1} . The sigmoid was used as the activation function of forget gate. The current input data x_t and the output h_{t-1} of the memory cell at the last time interval ($t-1$) were used to calculate the activation value f_t of the forget gate as follows (Zhao et al., 2017; Fischer and Krauss, in press):

$$f_t = \text{sigmoid}(W_{fx}x_t + W_{fh}h_{t-1} + b_f) \quad (1)$$

where f_t is the activation value of forget gate; W_{fx} and W_{fh} are the weight matrices; b_f are the bias vector; x_t are the time-series data at current time interval t ; h_{t-1} is the output of the memory cell at the previous time interval ($t-1$); and sigmoid represents the activation function given by:

$$\text{sigmoid}(x) = \frac{1}{1 + e^{-x}} \quad (2)$$

The input gate i_t determines which information is put into the cell state. Two steps were conducted in this gate. More specially, a tanh layer is first used to generate candidate values C'_t that could be added to the state. Then, a sigmoid layer creates activation values i_t of the input gates based on current input data x_t and the output h_{t-1} of the memory cell at the previous time interval. The equations for these two steps are as follows (Ke et al., 2017; Zhao et al., 2017):

$$C'_t = \tanh(W_{Cx}x_t + W_{Ch}h_{t-1} + b_C) \quad (3)$$

$$i_t = \text{sigmoid}(W_{ix}x_t + W_{ih}h_{t-1} + b_i) \quad (4)$$

where W_{Cx} , W_{Ch} , W_{ix} and W_{ih} are the weight matrices; b_C and b_i are the bias vectors; and tanh represents the activation function given by:

$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (5)$$

The results of forget and input gates are then used to calculate the new cell states C_t . The new cell states C_t are calculated as (Fischer and Krauss, in press):

$$C_t = f_t \circ C_{t-1} + i_t \circ C'_t \quad (6)$$

where \circ denoting the Hadamard product. Eq. (6) indicates that the current cell states C_t are the combinations of long term memory C_{t-1} and the current memory C'_t . The forget gate f_t controls the amount of information from long term memory C_{t-1} stored in the current cell states. And the input gate determines the amount of information from the current memory C'_t stored in the current cell states.

The output gate o_t determines which information from the cell state is used to generate output h_t . More specially, the following two equations are used to calculate the final output h_t of the memory cell (Fischer and Krauss, in press):

$$o_t = \text{sigmoid}(W_{Gx}x_t + W_{Gh}h_{t-1} + b_G) \quad (7)$$

$$h_t = o_t \tanh(C_t) \quad (8)$$

where W_{Gx} and W_{Gh} are the weight matrices; b_G are the bias vectors; and h_t is the output vector of the LSTM layer.

The training procedure of LSTM NNs begins with randomly initializing weight matrices W_{fx} , W_{fh} , W_{Cx} , W_{Ch} , W_{ix} , W_{ih} , W_{Gx} , and W_{Gh} , as well as the bias vectors b_f , b_C , b_i , and b_G (Zhao et al., 2017). A batch of samples was then randomly selected from the whole sample. The dropout regularization within the recurrent layer was used to reduce the risk of overfitting and achieve better generalization (Fischer and Krauss, in press). At each of update step, a part of the input units is dropped randomly. An optimal low dropout value was used because greater dropout values lead to an obviously decrease in performance. The batch size and dropout value were selected based on the sensitivity analysis which will be discussed in the following section.

The LSTM NNs are trained by the Adam optimization algorithm (Kingma and Ba, 2015). It is an efficient stochastic optimization method that only requires first-order gradients with little memory requirement. The Adam optimizer has the advantages of high computation efficiency, straightforward implementation, low memory requirement, and being invariant to diagonal rescaling of the gradients (Kingma and Ba, 2015). It is appropriate for problems with large-scale data and parameters. Moreover, it is also suited for non-stationary objectives and problems with noisy and/or sparse gradients (Kingma and Ba, 2015). Based on the Adam optimizer, the weights and bias terms are trained to minimize the loss of objective function.

The developed LSTM NNs can be used to predict the bike sharing trip production and attraction at a TAZ. The predicting procedure begins with the standardizing the historical data. The observations of the selected variables were then used as inputs of the developed LSTM NN. After the predicted future production and attraction by LSTM NN were unstandardized, the mean absolute percentage error (MAPE) was used to compare the predict results with the original production and attraction data.

Table 1
Descriptive statistics of productions and attractions during the two weeks.

Descriptive statistics	Trip production	Trip attraction
Mean	19284.923	19284.729
Std. Dev.	12567.705	12514.255
Minimum	418	423
25% Percentile	10,070	10,115
Median	16,907	17,160
75% Percentile	26,226	26,158
Maximum	57,947	57,786
Range	57,529	57,363

4. Data analysis and results

4.1. Spatial and temporal trip patterns

Table 1 gives the descriptive statistics of the bike sharing trip productions and attractions at the 118 TAZs for the selected two weeks. The means of production and attraction are 19284.923 and 19284.729, respectively, indicating the considerable usage of sharing bike at each TAZ. The total area of Nanjing downtown area is about 100 km². The area of each TAZ is about 1 km² on average. The means of production and attraction suggested that there are a total of 1377 trip productions and 1378 trip attractions on average for each day in a TAZ. The standard deviations of 12567.705 and 12514.255 also indicate high variations of bike sharing trips across different TAZs. The average production over 118 TAZs for each day ranges from 30 to 4139. The average trip attraction over 118 TAZs for each day ranges from 30 to 4128.

The spatial distributions of bike sharing trip productions and attractions over TAZs were examined. As shown in Figs. 4(a) and

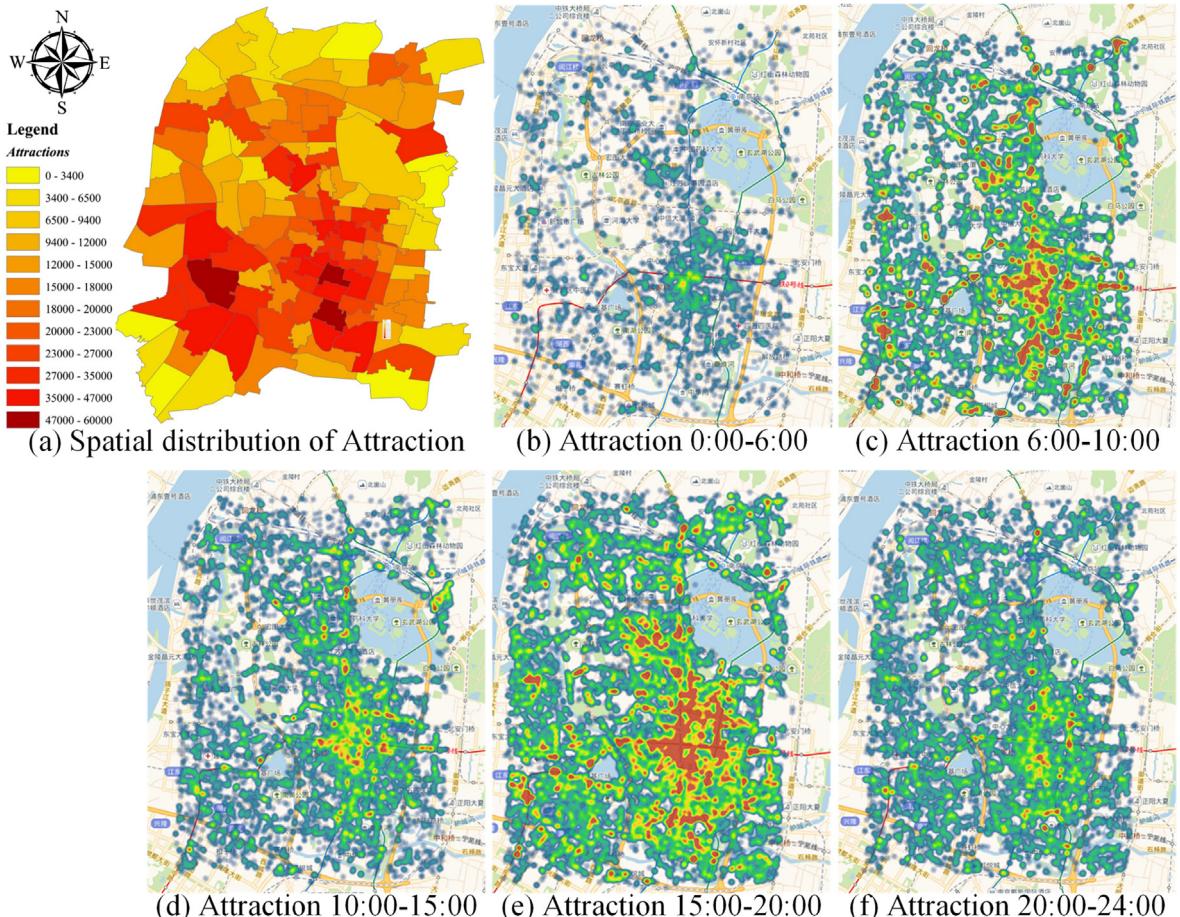


Fig. 4. The spatial and temporal distributions of attractions over traffic analysis zones.

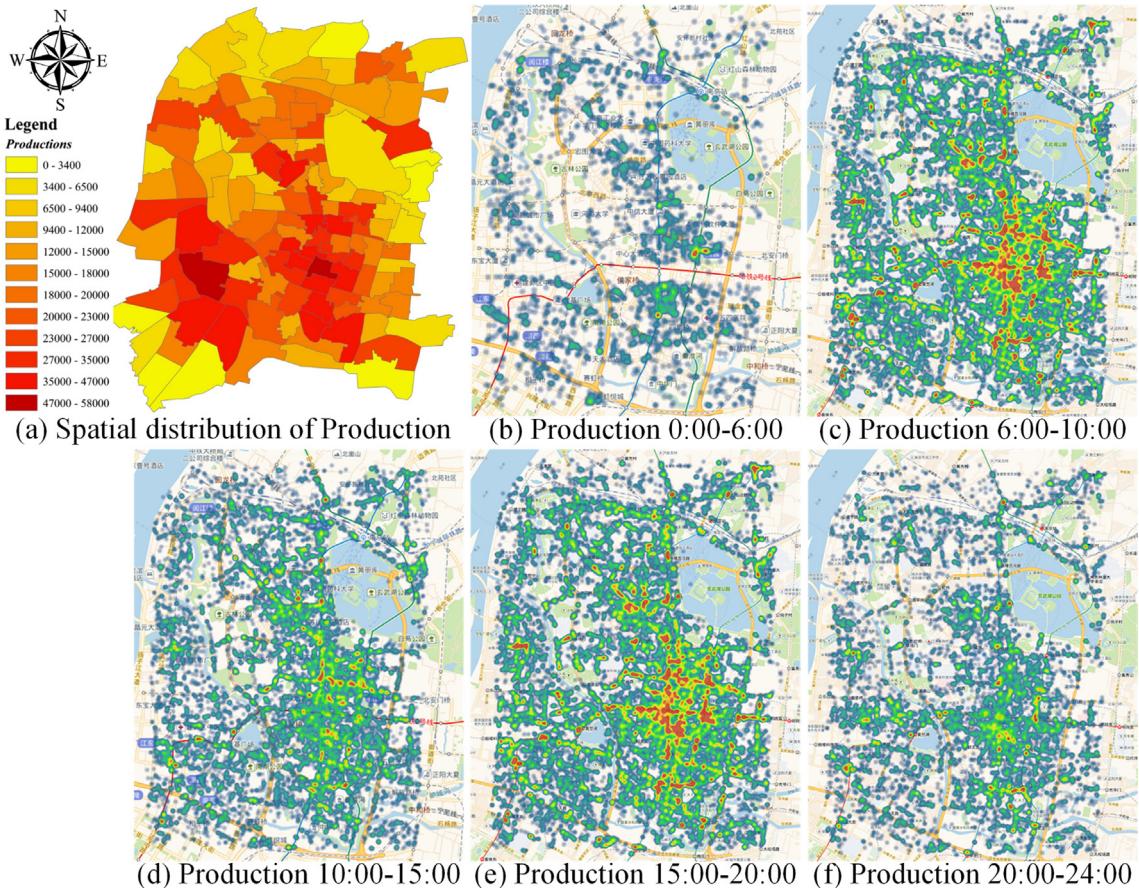


Fig. 5. The spatial and temporal distributions of productions over traffic analysis zones.

5(a), the distributions of bike sharing trip productions and attractions are quite different across the 118 TAZs. Comparing the production and attraction distributions of bike sharing trips in Fig. 4(a) and 5(a), it can be found that the distribution of trip production is generally negatively associated with that of the trip attraction. The TAZ with larger trip attractions generally have smaller trip productions. This result indicates the imbalanced spatial demand of bike sharing trips. In other words, the distribution of sharing bikes in the Nanjing downtown area will become significant skewed to several TAZs without rebalance operations. To keep a high level of service of the bike sharing system, considerable efforts are needed to rebalance the station-free sharing bikes.

Figs. 4 and 5 also illustrate the temporal distributions of bike sharing trip productions and attractions. It can be found that the distributions are quite different across different time period. In addition to the common peak periods of 7:00 am to 10:00 am, and 17:00 pm to 18:00 pm, considerable number of trips occurred in midday time between 11:00 am to 13:00 pm.

4.2. Development of LSTM NNs

The bike sharing trip generation model and trip attraction model were developed for the whole 118 TAZs. The trip generation and attraction models for each TAZ were developed for 4 different time intervals, including 10-min, 15-min, 20-min and 30-min intervals. For each time interval dataset, the data samples of the first 10-days were used as the training dataset, and the data samples of the last 4-day were used as the validation dataset.

Previous study suggested that weather conditions, air quality data, and land use patterns are important contributing factors to travel demand of station-based bike sharing (El-Assi et al., 2017). Accordingly, these factors are expected to also affect the station-free sharing bike demand. These data were included as the exogenous variables in the LSTM NNs (see Fig. 6). We compared the predictive performance of the LSTM NNs with sharing bike trip data only, and the fusion data of sharing bike trip, weather condition, air quality, and land use pattern data together. The comparison results of the LSTM NNs for the 30-min interval were given in Table 2. The MAPE was used to measure the predictive performance. As expected, the inclusions of weather, air quality, and land use patterns data lead to considerable improvement in prediction accuracy of the LSTM NNs for 30-min interval. Similar results can be found for other time intervals. Accordingly, all these variables can be used as the contribution factors in the LSTM NNs.

The weather data contain the detailed weather parameters, including temperature, dew point, humidity, wind speed and pressure, as well as the overall weather conditions, including clear weather, cloudy weather, light rain, mediate rain and heavy rain. The

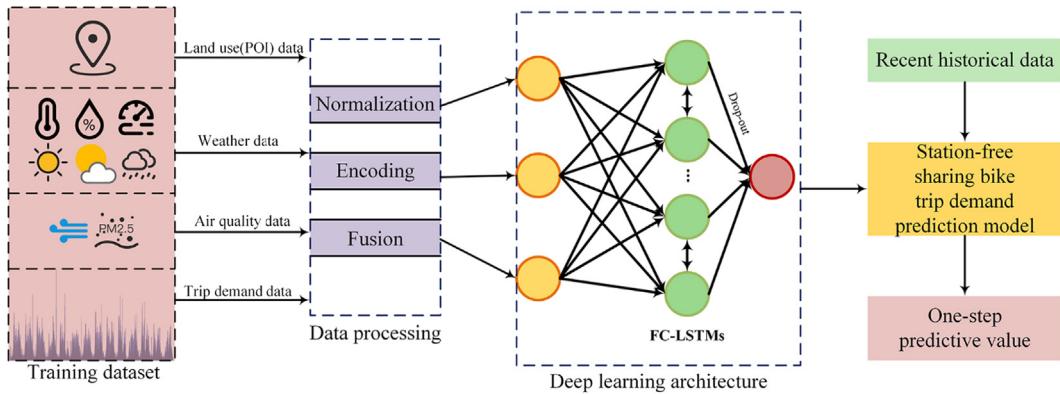


Fig. 6. LSTM NNs development and feature selection.

Table 2

Predictive performance of different models for 30-minute interval.

Models	Production	Attraction
LSTM (with trip data only)	28.374	25.875
LSTM (with trip and weather data)	14.365	17.704
LSTM (with trip and air quality data)	20.474	30.671
LSTM (with trip and land use data)	15.891	19.348
LSTM (with full data)	12.542	16.979

available weather data were collected at every 1 h, which is the commonly used time aggregation interval for weather data (Xu et al., 2018). The air quality data include the centration of PM 2.5 and the air quality index (AQI) which is an index measuring the overall air conditions. The land use patterns of each TAZ were measured by the extracted POI data at this TAZ.

With regard to the fusion process, since the land use patterns are temporally static variables, the variables of land use patterns for each TAZ at different time were measured by the percentages of different POIs at this TAZ. Both weather condition and air quality data are spatially static variables and collected at every one hour. The weather and air quality variables for each observation at each TAZ were extracted based on the time of each observation. The observations during one hour were assigned with the same values of weather and air quality variables.

In addition to the weather data, air quality data and land use patterns, the past observations at several nearest TAZs were also included as the input variables. The purpose of doing so was to account for the spatial correlation of production and attraction at TAZs. The centroid distant between 118 TAZs were calculated using GIS tool. The several nearest TAZs were selected based on the centroid distant between TAZs. Different numbers of selected nearest TAZs were used to develop LSTM NNs, and the prediction accuracies are then tested. It was found that the prediction accuracies of LSTM NNs did not have further improvement when the number of selected nearest TAZs is greater than 5. Accordingly, the previous interval observations at the nearest 5 TAZs were also included to account for the spatial correlation when developing LSTM NNs.

To optimize the model structure and increase predictive performance, the sensitivity analysis was conducted to tune four parameters of LSTM NN, including the number of training epochs, batch size, number of nodes, and dropout rate. The parameter tuning of LSTM NN for attraction with 10-min aggregation interval was used as an example to illustrate the process of parameter tuning. The parameter tunings for other time intervals followed the same procedure. Fig. 7(a) illustrates the relationship between loss and number of epoch. It was found that 80 epochs were sufficient to obtain a constant minimum loss. The Adam optimization algorithm was considered converged when reaching the number of epoch reach 80. As shown in Fig. 7(b), the batch size of 600 was used since this value leads to the best root mean square error (RMSE). Similarly, the number of nodes and dropout rate were set to be 60 and 0.07, respectively (see Fig. 7(c) and (d)). Table 3 gives the parameters of LSTM NNs for the four time intervals.

The LSTM NNs in this study were trained by using the TensorFlow deep learning tool in the Linux OS environment. TensorFlow is an open source software library of machine learning developed by the Google Brain Team (Abadi et al., 2016). It has been widely used for conducting deep learning in different fields. The API was used to specify the LSTM NNs. The LSTM NNs were trained on a workstation with 16 cores.

4.3. Predictive performance of LSTM NN

To better test the relative predictive performance of LSTM NN, seven methods were used, including one-step forecast, historical average (HA), autoregressive integrated moving average (ARIMA), extreme gradient boosting (XGBoost), support vector machine (SVM), artificial neural network (ANN), and RNN. The one-step forecast, HA and ARIMA are the commonly used statistical methods. The XGBoost, SVM, ANN and RNN are the recently developed artificial intelligent algorithms. A short introduction of these methods

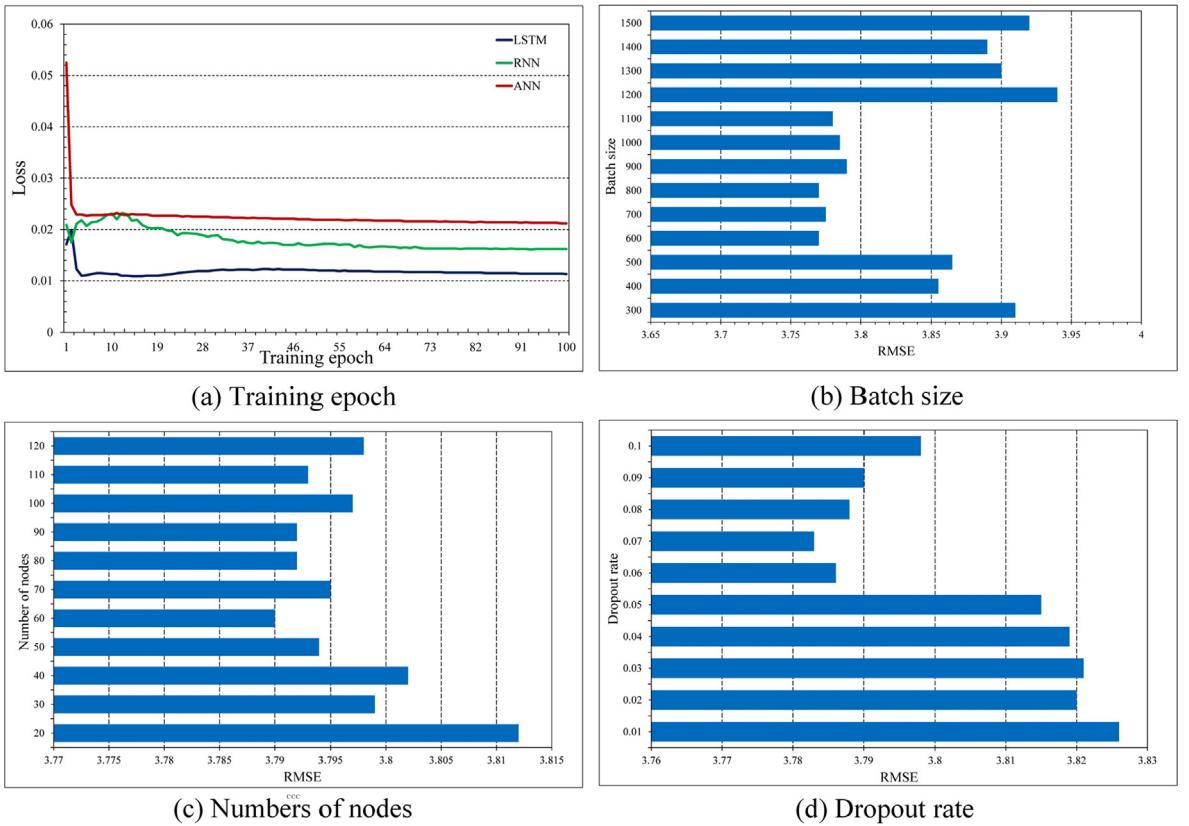


Fig. 7. Parameter tuning and sensitivity analysis of LSTM parameters.

Table 3

The parameters and estimation results of LSTM NNs for different time intervals.

Aggregation level	Attraction				Production			
	Batch size	Number of nodes	Dropout rate	MAPE	Batch size	Number of nodes	Dropout rate	MAPE
10 min	600	60	0.07	46.499	600	70	0.09	49.687
15 min	500	80	0.13	33.802	400	50	0.07	33.159
20 min	1200	40	0.14	34.649	900	50	0.10	20.820
30 min	600	70	0.08	16.979	1000	80	0.05	12.542

Table 4

Estimation results of the ARIMA models for different time intervals.

Aggregation level	Attraction				Production			
	p	d	q	MAPE	p	d	q	MAPE
10 min	0	1	3	73.751	1	2	2	83.812
15 min	0	1	14	78.971	3	1	9	73.160
20 min	1	1	10	52.320	1	1	3	50.102
30 min	1	2	4	40.575	0	1	7	58.229

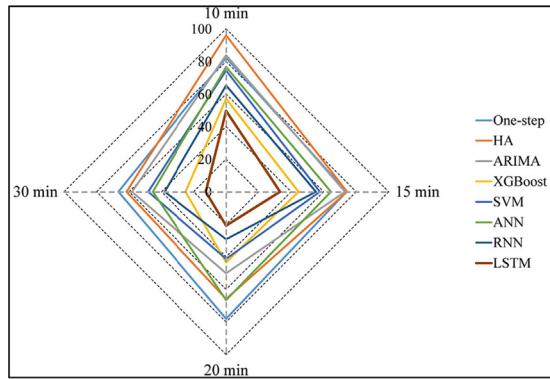
is given as follows:

- (1) **One-step forecast:** this method predict the value of production and attraction at the next time interval as the current observed value.
- (2) **HA:** the future demand is predicted as the average value at the same time period and the same TAZ during the past days.
- (3) **ARIMA:** this method is one of the most commonly used statistical regressions for time-series data modeling (Ilbeigi et al., 2016). In the ARIMA model developed in this study, the future demand is assumed to be a linear function of several past observations,

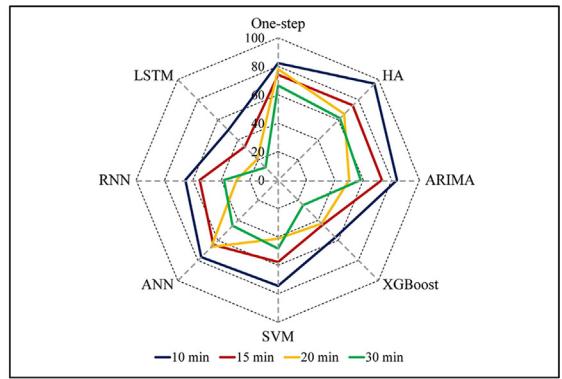
Table 5

Predictive performance of different models on the validation sample.

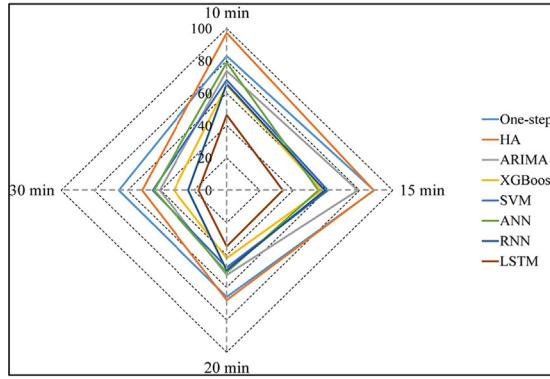
Models	Production				Attraction			
	10 min	15 min	20 min	30 min	10 min	15 min	20 min	30 min
One-step	82.250	74.125	78.017	66.568	82.957	88.745	65.866	65.259
HA	95.961	74.385	65.867	61.604	97.164	88.508	67.805	51.185
ARIMA	83.812	73.160	50.102	58.229	73.751	78.971	52.320	40.575
XGBoost	57.091	44.560	43.389	24.965	65.128	57.158	41.900	31.941
SVM	74.487	57.677	41.218	48.221	68.144	60.916	47.681	44.786
ANN	76.649	64.490	66.556	45.391	78.896	55.229	50.717	44.007
RNN	65.341	55.227	29.200	38.065	65.690	59.074	49.444	23.254
LSTM	49.687	33.159	20.820	12.542	46.499	33.802	34.649	16.979



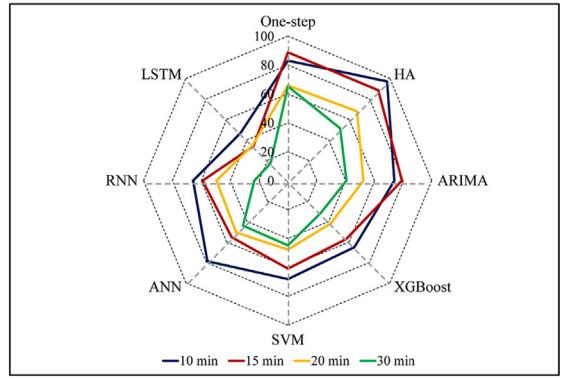
(a) Prediction accuracy of production at different intervals



(b) Prediction accuracy of production of different models



(c) Prediction accuracy of attraction at different intervals



(d) Prediction accuracy of attraction of different models

Fig. 8. Comparison of prediction accuracy of different models.

exogenous variables and random errors. The used parameters p , d and q were given in the following Table 4.

- (4) **XGBoost:** it is one of the recently developed machine learning models based on an implementation of gradient boosted decision trees (Chen and Guestrin, 2016). The gbtree was selected as the booster of the model. The parameter eta of 0.3 was used for different time intervals with the purpose of preventing overfitting problem. The maximum depth of tree and subsample were set to be 6 and 0.8 respectively.
- (5) **SVM:** It is an advanced machine learning model based on the statistical learning theory and the structural risk minimization principle. Numerous recent studies suggested that SVM can achieve high prediction performance. The parameters of embedding dimension m and penalty factor C were set as 8 and 1000 respectively for different time intervals. The parameter σ of the radial basis function was set as 2.3.
- (6) **ANN:** it is one of the most commonly used artificial intelligent algorithms based on interconnected group of nodes. This method has been widely applied to classifications and regression analyses in transportation engineering (Vlahogianni et al., 2005). To make the results comparable to those of LSTM NNs, the used parameters of ANN were the same to those of LSTM NNs for different time intervals.
- (7) **RNN:** this method is an advanced deep learning approach for time-series data modeling. It is an extension of ANN by recurrently connecting hidden layers at different timestamps (Hochreiter and Schmidhuber, 1997). The used parameters of ANN were the

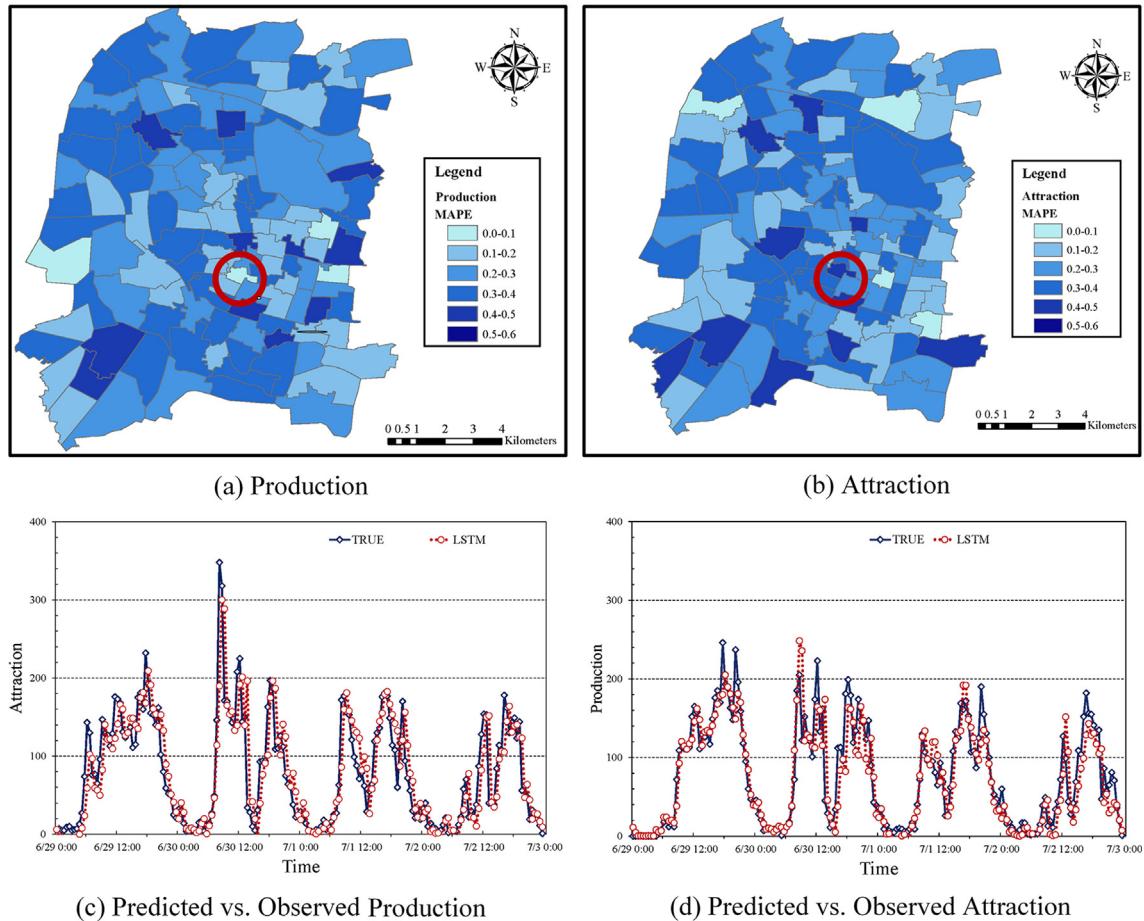


Fig. 9. Predicted versus observed productions and attractions.

same to those of LSTM NNs for different time intervals, with the purpose of making the results of RNN comparable to those of LSTM NNs.

The ARIMA models were estimated by the SPSS software package. For development of an ARIMA model, the input time series data are needed to be stationary. The stationary time-series data should have the same statistical behavior at each time step. The ARIMA model based on non-stationary time series data will generate significant errors (Washington et al., 2003). Readers are recommended to consult Washington et al. (2003) for full explanation of the requirement of stationarity in the time series analysis. The differencing was used for the original production and attraction data to ensure the requirement of stationarity. The autocorrelation and partial autocorrelation analyses were then conducted to test the stationarity of the production and attraction data after differencing.

To identify the best combinations of the autoregressive term order p and moving average term order q , the ARIMA models were developed for different combinations of parameter p and q . The parameter p and q were set from 0 to 15. The Akaike's Information Criterion (AIC) was used to find the best ARIMA model. The white noise tests were conducted for the developed ARIMA models to make sure there is no pattern remaining. Table 4 gives the estimation results of the ARIMA models and predictive performance on the validation data sample for different time intervals.

The XGBoost, SVM, ANN and RNN were trained by using the correspondence open source package for Python. All the 7 benchmarks were trained by using the same training samples and exogenous variables to LSTM NNs training. Table 5 and Fig. 8 compare the predictive performance of LSTM NNs with those of 7 benchmarks on the validation data sample. The MAPE was used to measure the predictive performance. For all the four time intervals, the LSTM NNs produce significantly higher prediction accuracy than those of the conventional statistical methods (see Fig. 8(a) and (c)). Taking ARIMA as an example, the average differences in MAPE between LSTM NNs and ARIMAs for production and attraction data are 37.3% and 28.4%, respectively, indicating that LSTM NNs can increase the prediction accuracy of production and attraction data by an average of 37.3% and 28.4% compared with the ARIMA models. The predictive performances of LSTMs NN are also better than those of four machine learning models. The LSTM NNs increase the prediction accuracy of production and attraction data by an average of 17.9% and 16.4% compared with RNN. Comparing the prediction accuracy of LSTM NNs for different time intervals, it can be found that the predictive performance of LSTM NN increases with an increase in the aggregation time interval (see Fig. 8(b) and (d)). The reason for this result is that the sharing bike

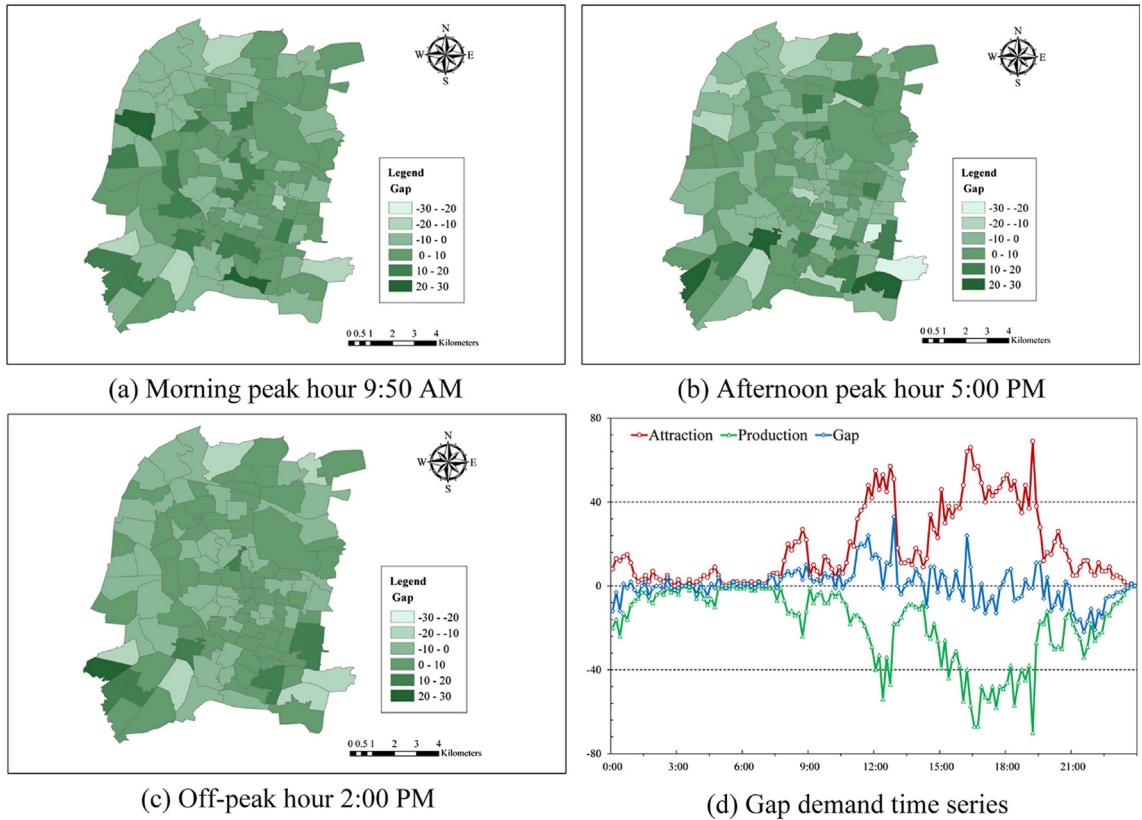


Fig. 10. The predicted demand gap for the 10-min time interval.

data aggregated at the 10-min interval have greater data noises and more useless fluctuation information than the longer time intervals. The data aggregated at shorter time intervals are more difficult to be predicted. Previous studies about short-term traffic flow forecast have also found similar results that prediction accuracy of traffic flow data aggregated at a shorter time interval is worse than a longer time interval (Chen et al., 2012; Xu et al., 2016).

For illustrative purposes, Fig. 9 was developed to demonstrate the prediction accuracy of the LSTM NNs over the 118 TAZs for 30-min time interval. Fig. 9(a) and (b) illustrate the MAPEs of the predicted production and attraction at each of the 118 TAZs. In general, the LSTM NN model provides reasonably accurate forecasts of both production and attraction data over the 118 TAZs. Fig. 9(c) and (d) further illustrate the predicted production and attraction of the LSTM NNs and the original observations for one selected TAZ marked with a red circle in the Fig. 9(a) and (b). As expected, the predicted production and attraction by LSTM NN are very close to the original observations, indicating that the predictive performances of LSTM NNs are satisfactory.

4.4. Demand gap prediction using the demand forecasting models

The demand models of station-free bike sharing system can be used to improve the operation efficiency of the system. One of the practical application is to provide useful information for rebalancing. The developed trip production and attraction forecast models can be used to predict the demand gap, which is defined as the difference between the inflow and outflow of the sharing bike trips. Such information can be used to rebalance the station-free sharing bike among various TAZs. For example, Fig. 10(a) illustrates the spatial distribution of the predict demand gap by the LSTM NNs for 10-min across the 118 TAZs at 9:50 AM. By investigating Fig. 10(a), it is intuitive to identify which TAZs have continuously decreasing sharing bikes and which TAZs have increasing sharing bikes. Based on the identified imbalance pattern of spatial demand, specific strategy can be used by the operation company to move the bikes from over-supply areas to over-demand areas.

Fig. 10(b) and (c) illustrate the demand gap distributions at the time of 2:00 PM and 5:00 PM. It can be found that the imbalance patterns of spatial demand in these two periods are obviously different from that in Fig. 10(a). Accordingly, different strategies should be used to rebalance the sharing bike at different time periods. Fig. 10(d) illustrates the time series of the predicted demand gap at one selected TAZ that is marked with a red circle in Fig. 10(a). From Fig. 10(d), it is intuitive to find when the demand gap will happen. With these information about demand gap, reasonable strategies can be developed to improve the rebalance efficiency. These information can help to determine the time and area to relocation bikes, the number of bikes needed to be relocated and the route of relocating bikes. The latest study also suggested that the zone-based demand models are the basis for the rebalance of station-free

system (Caggiani et al., 2018). Since the rebalance strategy is not the main objective of this study, additional efforts are needed to develop rebalance strategy in future studies.

5. Conclusion and discussion

This study investigated the mobility pattern of the station-free bike sharing at a citywide scale, and developed dynamic demand forecasting models to predict the travel demand of station-free sharing bike using the deep learning approach. The station-free bike sharing data from one of the operation companies were collected from the downtown area of the Nanjing City for two weeks. The spatial and temporal analyses were first conducted to investigate the mobility pattern of the station-free bike sharing in the Nanjing downtown area. The results indicate imbalanced spatial and temporal demand of bike sharing trips.

The LSTM NN models were developed to predict the bike sharing trip production and attraction at a TAZ in the next time interval based on the several past observations at this TAZ. The weather data, air quality condition and land use characteristics were included as exogenous variables when training LSTM. Four time intervals were used in the model development, including the 10-min, 15-min, 20-min and 30-min intervals. The validation results suggested that the developed LSTM NN models have reasonable good prediction accuracy in trip productions and attractions for different time intervals. 7 benchmarks were also developed for different time intervals to test the relative predictive performance of the LSTM NNs. The comparison results suggested that the LSTM NN models provide better prediction accuracy than commonly used statistical models and machine learning algorithms. Finally, the applications of the LSTM NN for identifying demand gap were examined. The results suggested that the LSTM NNs can be used to predict the demand gap, which provide useful information for rebalancing the sharing bike in the system.

Acknowledgement

This research was sponsored by the National Natural Science Foundation of China (Grant No. 51508093), Projects of International Cooperation and Exchange of the National Natural Science Foundation of China (No. 51561135003), and Natural Science Foundation of Jiangsu Province (BK20171358). The authors would like to thank the editor and the reviewers for their constructive comments and valuable suggestions to improve the quality of this article.

References

- Abadi, M., Agarwal, A., Barham, P., 2016. Tensorflow: Large-scale machine learning on heterogeneous distributed systems. arXiv preprint arXiv:1603.04467.
- Bao, J., Xu, C., Liu, P., Wang, W., 2017. Exploring bikesharing travel patterns and trip purposes using smart card data and online point of interests. *Networks Spatial Econ.* 17 (4), 1231–1253.
- Beecham, R., Wood, J., 2014. Exploring gendered cycling behaviours within a large-scale behavioural dataset. *Transport. Plann. Technol.* 37 (1), 83–97.
- Caggiani, L., Camporeale, R., Ottomanelli, M., Szeto, W.Y., 2018. A modeling framework for the dynamic management of free-floating bike-sharing systems. *Transport. Res. Part C: Emerg. Technol.* 87, 159–182.
- Campbell, A.A., Cherry, C.R., Ryerson, M.S., Yang, X., 2016. Factors influencing the choice of shared bicycles and shared electric bikes in Beijing. *Transp. Res. Part C* 67, 399–414.
- Chen, C., Wang, Y., Li, L., Hu, J., Zhang, Z., 2012. The retrieval of intra-day trend and its influence on traffic prediction. *Transport. Res. Part C: Emerg. Technol.* 22, 103–118.
- Chen, T., Guestrin, C., 2016. Xgboost: A scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, pp. 785–794.
- Chen, X.M., Zahiri, M., Zhang, S., 2017. Understanding ridesplitting behavior of on-demand ride services: an ensemble learning approach. *Transport. Res. Part C: Emerg. Technol.* 76, 51–70.
- Dabiri, S., Heaslip, K., 2018. Inferring transportation modes from GPS trajectories using a convolutional neural network. *Transport. Res. Part C: Emerg. Technol.* 86, 360–371.
- Duan, Y., Lv, Y., Liu, Y.L., Wang, F.Y., 2016. An efficient realization of deep learning for traffic data imputation. *Transport. Res. Part C: Emerg. Technol.* 72, 168–181.
- El-Assi, W., Mahmoud, M., Habib, K., 2017. Effects of built environment and weather on bike sharing demand: a station level analysis of commercial bike sharing in Toronto. *Transportation* 44, 589–613.
- Fischer, T., Krauss, C., 2018. Fischer and Krauss, in press. Deep learning with long short-term memory networks for financial market predictions. *Eur. J. Oper. Res.* 270 (2), 654–669.
- Fournier, N., Christofa, E., Knodler, M., 2017. A sinusoidal model for seasonal bicycle demand estimation. *Transp. Res. Part D* 50, 154–169.
- Feng, Y., Wang, S., 2017. A Forecast for Bicycle Rental Demand Based on Random Forests and Multiple Linear Regression. Presented at the International Conference on Information Systems, 2017.
- Fishman, E., Washington, S., Haworth, N., Watson, A., 2015. Factors influencing bike share membership: an analysis of Melbourne and Brisbane. *Transp. Res. Part A* 71, 17–30.
- Gebhart, K., Noland, R.B., 2014. The Impact of weather conditions on Bikeshare trips in Washington, D. C. *Transportation* 41 (6), 1205–1225.
- Hochreiter, S., Schmidhuber, J., 1997. Long short-term memory. *Neural Comput.* 9 (8), 1735–1780.
- Ilbeigi, M., Ashuri, B., Joukar, A., 2016. Time-series analysis for forecasting asphalt-cement price. *J. Manage. Eng.* 33 (1), 04016030.
- Jiang, X., Zhang, L., Chen, X.M., 2014. Short-term forecasting of high-speed rail demand: a hybrid approach combining ensemble empirical mode decomposition and gray support vector machine with real-world applications in China. *Transport. Res. Part C: Emerg. Technol.* 44, 110–127.
- Kanarachos, S., Christopoulos, S.R.G., Chroneos, A., 2018;al, in press. Smartphones as an integrated platform for monitoring driver behaviour: the role of sensor fusion and connectivity. *Transport. Res. Part C: Emerg. Technol.* (in press).
- Ke, J., Zheng, H., Yang, H., Chen, X.M., 2017. Short-term forecasting of passenger demand under on-demand ride services: A spatio-temporal deep learning approach. *Transport. Res. Part C: Emerg. Technol.* 85, 591–608.
- Kingma, D., Ba, J., 2015. Adam: a method for stochastic optimization. In: Proceeding of International Conference on Learning Representations.
- Liu, L., Chen, R.C., 2017. A novel passenger flow prediction model using deep learning methods. *Transport. Res. Part C: Emerg. Technol.* 84, 74–91.
- Liu, Y., Szeto, W.Y., Ho, S.C., 2018. A static free-floating bike repositioning problem with multiple heterogeneous vehicles, multiple depots, and multiple visits. *Transport. Res. Part C: Emerg. Technol.* 92, 208–242.
- Lv, Y., Duan, Y., Kang, W., Li, Z., Wang, F.Y., 2015. Traffic flow prediction with big data: a deep learning approach. *IEEE Trans. Intell. Transp. Syst.* 16 (2), 865–873.
- Ma, X., Tao, Z., Wang, Y., Yu, H., Wang, Y., 2015a. Long short-term memory neural network for traffic speed prediction using remote microwave sensor data.

- Transport. Res. Part C: Emerg. Technol. 54, 187–197.
- Ma, X., Yu, H., Wang, Y., Wang, Y., 2015b. Large-scale transportation network congestion evolution prediction using deep learning theory. *PLoS One* 10 (3), 1–17.
- Ma, X., Dai, Z., He, Z., Ma, J., Wang, Y., Wang, Y., 2017. Learning traffic as images: a deep convolutional neural network for large-scale transportation network speed prediction. *Sensors* 17 (4), 818.
- Pal, A., Zhang, Y., 2017. Free-floating bike sharing: solving real-life large-scale static rebalancing problems. *Transport. Res. Part C: Emerg. Technol.* 80, 92–116.
- Polson, N.G., Sokolov, V.O., 2017. Deep learning for short-term traffic flow prediction. *Transport. Res. Part C: Emerg. Technol.* 79, 1–17.
- Simoncini, M., Taccari, L., Sambo, F., Bravi, L., Salti, S., Lori, A., 2018. Vehicle classification from low-frequency GPS data with recurrent neural networks. *Transport. Res. Part C: Emerg. Technol.* 91, 176–191.
- Vlahogianni, E.I., Karlaftis, M.G., Golias, J.C., 2005. Optimized and meta-optimized neural networks for short-term traffic flow prediction: a genetic approach. *Transp. Res. Part C* 13 (3), 211–234.
- Washington, S., Karlaftis, M., Mannerling, F., 2003. Statistical and Econometric Methods for Transportation Data Analysis. Chapman & HALL/CRC.
- Wu, Y., Tan, H., Qin, L., Ran, B., Jiang, Z., 2018. A hybrid deep learning based traffic flow prediction method and its understanding. *Transport. Res. Part C Emerg. Technol.* 90, 166–180.
- Xu, C., Li, Z., Wang, W., 2016. Short-term traffic flow prediction using a methodology based on autoregressive integrated moving average and genetic programming. *Transport* 31 (3), 343–358.
- Xu, C., Li, H., Zhao, J., Chen, J., Wang, W., 2017. Investigating the relationship between jobs-housing balance and traffic safety. *Acc. Anal. Prevent.* 107, 126–136.
- Xu, C., Wang, Y., Liu, P., Wang, W., Bao, J., 2018. Quantitative risk assessment of freeway crash casualty using high-resolution traffic data. *Reliab. Eng. Syst. Saf.* 169, 299–311.
- Yu, H., Wu, Z., Wang, S., Wang, Y., Ma, X., 2017. Spatiotemporal recurrent convolutional networks for traffic prediction in transportation networks. *Sensors* 17, 1501.
- Zhang, Z., He, Q., Gao, J., Ni, M., 2018. A deep learning approach for detecting traffic accidents from social media data. *Transport. Res. Part C Emerg. Technol.* 86, 580–596.
- Zhao, J., Wang, J., Deng, W., 2015. Exploring Bikesharing travel time and trip chain by gender and day of the week. *Transp. Res. Part C* 58, 251–264.
- Zhao, J., Deng, W., Song, Y., 2014. Ridership and effectiveness of bikesharing: the effects of urban features and system characteristics on daily use and turnover rate of public bikes in China. *Transp. Policy* 35, 253–264.
- Zhao, Z., Chen, W., Wu, X., Chen, P.C., Liu, J., 2017. LSTM network: a deep learning approach for short-term traffic forecast. *IET Intel. Transport Syst.* 11 (2), 68–75.