Neural Computing and Applications
https://doi.org/10.1007/s00521-018-3470-9

S.I.: EMERGING INTELLIGENT ALGORITHMS FOR EDGE-OF-THINGS COMPUTING

A deep learning approach on short-term spatiotemporal distribution forecasting of dockless bike-sharing system

Yi Ai[1,2] · Zongping Li[1,2] · Mi Gan[1] · Yunpeng Zhang[3] · Daben Yu[1,2] · Wei Chen[1,2] · Yanni Ju[1,2]

Received: 17 January 2018 / Accepted: 27 March 2018

Abstract
Dockless bike-sharing is becoming popular all over the world, and short-term spatiotemporal distribution forecasting on system state has been further enlarged due to its dynamic spatiotemporal characteristics. We employ a deep learning approach, named the convolutional long short-term memory network (conv-LSTM), to address the spatial dependences and temporal dependences. The spatiotemporal variables including number of bicycles in area, distribution uniformity, usage distribution, and time of day as a spatiotemporal sequence in which both the input and the prediction target are spatiotemporal 3D tensors within one end-to-end learning architecture. Experiments show that conv-LSTM outperforms LSTM on capturing spatiotemporal correlations.

Keywords Dockless bike-sharing system · Short-term spatiotemporal distribution forecasting · Deep learning · Convolutional long short-term memory network (conv-LSTM)

1 Introduction

In China, the dockless bike-sharing system, e.g., Mobike, Ofo, and Bluegogo, show a rapid growth trend that spreads rapidly from major cities to small- and medium-sized cities. The way it works is simple enough in theory. Users download an application that tells them where to find a cycle, which they unlock by scanning a QR code on their phones or using a combination they are sent. Unlike traditional rental services, however, which require bikes to be returned to a fixed docking station, riders are free to leave the bikes wherever their journey ends. The traditional bike-sharing systems fail to address the connectivity issue, which refers to the distance between the subway or a bus stop and a passenger's starting point or final destination as well as the short-distance trip. Ofo and Mobike answer precisely this need by enabling users to take and leave their bikes wherever is convenient. The user looks for nearby bikes or geo-localizes them with his smartphone, then scans the QR code of the bike he has selected. For Mobike, the QR code instantly unlocks the bike. Ofo users must additionally enter a four-number code sent to their phones via the application to unlock the bicycle.

These brands are the product of a whole host of new startups, aggressively competing for territory and investment. As a result, the two biggest operators—Mobike and Ofo, have occupied a large domestic market and high popularity [1].

The two leading bike-sharing companies Mobike and Ofo are aware of how convenient bikes can be. Taking the example of Beijing, 93% of travels less than 5 km are quicker done by bike and public transport than with the car. Drawing on this convenience aspect, their business models rely on bikes that can be ridden and left anywhere in the city (or a defined area) thanks to a free, easy-to-use application and an efficient geo-localization system. Ofo's and Mobike's model is therefore a leading business model in China, especially with regard to being cost-effective and time-saving [2].

✉ Mi Gan
migan@swjtu.edu.cn

Yi Ai
aiyi@my.swjtu.edu.cn

[1] School of Transportation and Logistics, Southwest Jiaotong University, Chengdu, China

[2] Comprehensive Transportation Key Laboratory of Sichuan Province, Chengdu, China

[3] College of Technology, University of Houston, Houston, USA

Published online: 07 April 2018

The scale is simply stunning. In less than a year, Mobike alone has flooded the streets of over 160 Chinese cities with what is thought to be more than a million new bikes. Since April 2017, the company has placed more than 100,000 of their trademark orange-and-silver bikes in big cities like Shanghai, Beijing, Guangzhou, and Chengdu. Already backed by the Chinese internet giant Tencent, a recent deal with Apple supplier Foxconn has doubled Mobike's production capacity. In summer 2017 Mobike, after having launched in Manchester, has begun operations in Florence and Milan, and it will start in London in September [3–5]. Sapporo is the newest city to be launched by Mobike, on August 23, 2017 [6]. Mobike launched in Bandar Setia Alam, Shah Alam, Selangor, Malaysia on September 7, 2017 with collaboration with SP Setia.

Then there's Ofo—which started in 2015 as a Peking University project and now claims 30 million users in 46 cities for its bright yellow bikes [7]. They are parked up by the hundred outside shopping malls and metro stations. Furthermore, Ofo is cooperating with Alipay, China's largest online payment platform, to launch in the market of USA, UK, Austria, etc.

The operators make the size and location of the launch as well as the redistribution according to the supply and demand relationship. Meanwhile, dynamic changes in bike distribution in different regions characterized by user's OD behavior will affect the efficiency of bike-sharing system to a great extent, so as to affect people's travel behavior choice. Therefore, conducting a comprehensive short-term forecast over different spatial zones is of great importance to the operator, who can adjust the launch and redistribution plan to incline to the zones with more potential passenger demands, and improve the utilization rate of bikes. The short-term prediction is a prediction of changes in market development in a short time (usually within a day or a short period of time). This kind of forecast activity is large and frequent in the business operation. Short-term prediction can help enterprises to understand the market dynamics in time, grasp the favorable opportunity of market change, and improve the level of management decision.

There are abundant researches focused on forecasting short-term distribution under different urban traffic system such as bus [8, 9], metro [10], taxi [11, 12] as well as the emerging on-demand ride service platform for private car sharing [13]. These studies on transportation forecasting can provide valuable insights into the online ride-hailing market for dockless sharing bike. A series of mathematic models had been developed to spell out the interaction between supply and demand. For the demand, it can be affected by waiting time, fare, etc., while for the supply, vehicle quantity, distribution, and fare may be the factors affecting the supply level. However, some new features triggered by dockless bike-sharing system have emerged in the supply and demand relationship. The spatiotemporal distribution is determined by density of dockless bikes and distribution uniformity, i.e., how far to find an available sharing bike. Furthermore, the regional distribution is also considered to be dynamic along with demand realizing; for example, the over-demand in one region will lead to the decline in the number of shared bike in this region, which causes more serious shortage. These problems raise a strong need for a precise forecasting of short-term distribution. It helps the operator implement launch and redistribution mechanism. These strategies not only make supply and demand more balanced during the different time period but also help improve the bike utilization rate and reduce the search time to find a bike.

However, forecasting short-term spatiotemporal distribution is very challenging due to the characteristics of uneven spatial and time distributions [13].

1. Spatial dependences: The spatiotemporal distribution triggered by dockless bike-sharing in one specific zone was not merely related to density of dockless bikes and distribution uniformity of this zone, but endogenously dependent on the zonal attribute of adjacent areas and even the whole network. Generally, the mutual circulation of bikes is easier among adjacent zone, which raises the stronger influences than nonadjacent zone. These requirements indicate that the short-term distribution is dependent on spatial attribute.
2. Time dependences: Spatiotemporal distribution is considered to be accompanied by a strong time dependences. It is expected to be regular peaks and troughs at different times of day. Furthermore, the short-term distribution is also related to the demand of the previous time point.
3. Demand dependence: The spatiotemporal distribution in the zone is also affected by the actual user demand, which will lead to dynamic changes in the distribution of sharing bike.

In the current study, little direct experience proposes solutions to this multidimensional dependency in short-term distribution forecasting. Fortunately, deep learning (DL) approaches provides an availability solution to short-term passenger demand for traffic flow prediction [13]. In these researches, long short-term memory (LSTM) neural network and convolutional neural network (CNN) are proposed to capture the short-term dependences and spatiotemporal correlations [14, 15]. Furthermore, an innovation based on integrated CNN and LSTM in one end-to-end DL structure, named the convolutional LSTM (conv-LSTM), has provided a new approach to the short-term passenger demand forecasting, which showed a better performance than traditional prediction model [13, 15].

Recently, a novel deep learning (DL) approach, the fusion convolutional long short-term memory network (FCL-Net), is used to address these three dependences within one end-to-end learning architecture, which better captures the spatiotemporal characteristics and correlations of explanatory variables [13]. It can be considered as an ideal model for this problem.

In this paper, we will employ this DL structure based on conv-LSTM to capture the superposition dependences for the dockless bike-sharing system. We attempt to develop a one end-to-end trainable model to learn the spatiotemporal variables (number of bikes in area, distribution uniformity, and usage distribution) and time series variables (time of day). The number of bikes in area and distribution uniformity corresponds to temporal and spatial dependence, and time of day corresponds to temporal dependence, while usage distribution has both temporal and spatial correlation and demand correlation. The main contributions of this paper are as follows:

1. The DL approach characterizes the spatiotemporal properties of the predictors and captures both the spatiotemporal features of spatiotemporal variables and the temporal features of non-spatial time series variables. Simultaneously, it coordinates them in one end-to-end learning structure for the short-term distribution.
2. We obtained some valuable information related to bike-sharing users by data mining and extract the potential user behavior characteristics. Sensitivity analyses about these predictors are carried out.
3. Through the data mining based on application and network platform, a large number of vehicle usage data based on the spatiotemporal properties and regional distribution of bikes are introduced to the DL model to forecast short-term distribution.

2 Literature review

A review of the literature shows that the development of bike-sharing system has gone through four stages [16]: The first generation of shared bikes can be traced back to the late 1960s in Europe, where the "White Bicycle System" is found in Amsterdam, Netherlands. The second generation is famous for the "Coin Deposit Systems" required users to insert a refundable deposit to unlock it. Unfortunately, the two generations of bike-sharing systems are a failure due to restrictions on use and lack of flexibility. The third generation, with docking stations and smart technology for bicycle check-in and check-out, has popularized on a large scale in Lyon, France, in 2005. With the rapid development of vehicle networking technology, the fourth generation has launched in China in which travelers can use mobile phones to locate idle bikes around and scan QR codes to unlock them as well as flexible parking. This system, called dockless bike-sharing system, has grown exponentially around the world.

Lately, the reported dockless bike-sharing's operation has aroused widespread concern. A consistent help toward a better regulation of bike-sharing systems lies in forecasting their demands. Forecasting the demand of the dockless bike-sharing system is beneficial for users, which is able to inform them about the bikes distribution in one zone at a certain time in the future and better plan their trips. Similar benefits are also present for the operators, and the short-term demand forecasting is also able to conduct the launch and redistribution plan to improve the customer's experience [17, 18]. Travelers choose a trip mode among walking, bike-sharing system, and public transportation according to time cost, physical cost, and expense cost, while the operators are designed to pursue higher usage and revenue and attract more users. There is a common ground between their goals, that is to meet more demand as much as possible. Thus the short-term demand forecasting is expected to be of a great importance to bike-sharing systems.

Many valuable researches about the prediction of short term have been carried out [19], such as dynamic flow distribution, historical mean, regression analysis, time series, support vector machine, neural network. With regard to short-term forecast, the linear prediction approaches are considered as a basic method, where the prediction is based on linear hypothesis and the previous trends to forecast future trends, such as the historical average model [20]. On the other hand, the nonlinear predictions were also widely used, and they are designed to capture the nonlinear features with a higher computing complexity [21]. Furthermore, the SVM with machine learning technique has been successfully used in data recognition. Some results show that the SVM in traffic prediction have better performed in accurately predicting time series data of nonlinear, non-stationary, and not predefined processes [22]. However, the SVM has been proven to be slightly worse in accuracy to the entire data with noise or information incompleteness [23]. In order to improve the recognition accuracy of complex systems, neural network approaches are proposed to capture local characteristics of short-term demand. Implementing the comparison between the Markov algorithm and neural network, the neural network shows a better adaptability for blocks with low predictability, while the Markov predictor performed better with high predictability [24].

In recent studies, the DL algorithms have been further applied in traffic prediction with better capability to extract distinguished characteristic from a huge amount of data.

The DL prediction approach provides an accurate recognition to time series with recurrent and non-recurrent traffic conditions. Some exogenous features can also be well extracted, such as weather, factor distribution. For instance, a DL structure based on day of time was proposed in Lv et al. [25], which indicates that DL can be effectively used in traffic forecasts. Huang et al. [26] proposed a deep belief networks with multitask learning to predict short-term traffic flow. However, these DL algorithms do little work on spatial and temporal dependence. To address this problem, the convolutional neural network (CNN) is developed to capture spatial correlations. Graham [27] developed a spatially sparse convolutional neural network for processing spatially sparse inputs, and the results indicated that the CNNs perform well to extract space features. Furthermore, it is noteworthy that the long short-term memory (LSTM) has manifested the extraordinary performance on time series data recognition. For example, Ma et al. [28] developed a long short-term memory neural network (LSTM) for travel speed prediction, where a comparative study suggests that the LSTM receives the best performance. However, neither CNN nor LSTM performs well to capture spatiotemporal characteristic. To overcome this shortcoming, a combined CNN and LSTM in one end-to-end structure have been proposed by Ke et al. [13] to capture spatial and temporal dependences simultaneously. This novel deep learning (DL) approach addresses the spatial dependences, temporal dependences, and exogenous dependence within multiple convolutional long short-term memory layers. Investigation results indicate that the model performs better than traditional approaches.

In this paper, we attempt to improve the structure of the convolutional long short-term memory network in order to address the spatial dependences, time dependences, and demand dependence.

3 Preliminary

The spatiotemporal distribution forecasting for the dockless bike sharing is essentially considered as a time series prediction problem. We can derive the valuable information about the future distribution from the historical distribution. In general, the attributes of time of day have a dominant influence on the spatiotemporal distribution, since it reflects the periodic characteristics of travel behavior. For example, the region distribution will show a rapid variation during the peak commuting. Furthermore, usage distribution also has an impact on dockless bike distribution.

In this section, we first interpret the notations of the variables based on temporal and spatial and then give an explicit framework for the short-term forecasting of the spatiotemporal distribution of dockless bike sharing.

According to the method proposed by Ke et al. [13], the urban area will be partitioned into $I \times J$ grid areas with equal size, and the interval of temporal series variables will be considered as 1 h in this paper. Based on the partitioning, we will define the property variables as follows:

1. Number of bikes in area

The average density of dockless bikes reflects a supply level during tth time interval $(t, t + \Delta t)$ in this grid area. In this paper, the average density of dockless bikes is denoted by $m_{(i,j)}^{t,\Delta t} \in M^t$. To reduce the impact of aggregation point with a large number of dockless bikes on the gird area, we set the maximum visible number of dockless bikes within 300 m radius to be 200.

2. Distribution uniformity

Distribution uniformity can reflect the concentration and dispersion of the distribution of bikes in the region. The more uniform the distribution will be, the more useful it will be for the users to find the available dockless bikes. The search radius for dockless bikes is usually set to 300 m in these apps of bike-sharing system. The grid area will be divided into five smaller square areas on average where each side is 600 m. The distribution uniformity is defined as the standard deviation for these square areas, i.e., $(D_t)_{i,j} = d_t^{i,j} = \sqrt{\sum \left(d_t^{i,j,k} - \overline{d}\right)^2}$, where $d_t^{i,j,k}$ represents the number of dockless bikes in kth square area of grid area (i, j), and $\overline{d}$ denotes the average number of dockless bikes in this grid area.

3. Usage distribution

The usage count of dockless bikes at the tth time slot (e.g., 1 h) in grid area (i, j) is denoted by $u_{(i,j)}^t$. The distribution of usage in all $I \times J$ grids at the tth time slot is defined as $u_{(i,j)}^t \in U^t$.

4. Time of- day

By preliminary processing of experimental data, it implies that the spatiotemporal distribution is characterized by time series. Each day can be intuitively divided into peak period, off-peak period, and sleep period, where the peak period is also divided into non-commuter peak, morning peak, and evening peak in order to embody the phenomenon of tidal traffic. The non-commuter peak period refers to the other peak periods divided by the morning peak and evening peak period. We define the variable δ^t as the attribute of time of day, given by

$$\delta^t = \begin{cases} 1, & t \text{ belongs to sleep period} \\ 2, & t \text{ belongs to off - peak period} \\ 3, & t \text{ belongs to non - commuter peak period} \\ 4, & t \text{ belongs to morning peak period} \\ 5, & t \text{ belongs to evening peak period} \end{cases}$$

3.1 Spatiotemporal variables

According to spatial and temporal dependence shown by variables, we have reason to believe that there is a spatiotemporal correlation among variables, e.g., number of bikes in area, distribution uniformity, and usage distribution.

3.2 Time series variables

The variables related to time of day can be clearly considered as a time series variable.

With the aforementioned definition of the explanatory variables, we can give the historical observations and preknown information $\{M^i, U^i, D^i, \delta^t | i = 1 \ldots t-1\}$, forecast $\{M^t, D_t\}$.

4 Methodology

In this paper, we employ a DL architecture—conv-LSTM to capture the spatial dependences, temporal dependences in short-term forecasting. In this section, we first present a brief review of the traditional LSTM and CNN, then introduce the DL architecture and training algorithm of conv-LSTM.

4.1 LSTM

The LSTM, a special RNN, the connection between units is organized by timestamps so as to better keep memory for a period of time by adding memory cell. The memory cell is controlled by the input gate, the forgetting gate, and the output gate. The sequence of input vectors pass to the memory cell tensors through the input gate, some information will be selectively obliterated by the forgetting gate and the processed information will be activated for the next input in memory cell. Finally, a hidden sequence will be outputted to by $T\psi$ iterations through the output gate. The inner structure of conv-LSTM is illustrated in Fig. 1. Input and information from previous state begins with a set of input are as follow:

$$\boldsymbol{i}_t = \sigma(\mathbf{w}_{xi}\boldsymbol{x}_t + \mathbf{w}_{ir}\boldsymbol{h}_{t-1} + \mathbf{w}_{ci} * \boldsymbol{c}_{t-1} + \boldsymbol{b}_i) \quad (1)$$

$$\boldsymbol{f}_t = \sigma(\mathbf{w}_{xf}\boldsymbol{x}_t + \mathbf{w}_{hf}\boldsymbol{h}_{t-1} + \mathbf{w}_{cf} * \boldsymbol{c}_{t-1} + \boldsymbol{b}_f) \quad (2)$$

$$\boldsymbol{c}_t = \boldsymbol{f}_t * \boldsymbol{c}_{t-1} + \boldsymbol{i}_t * \tanh(\mathbf{w}_{xc}\boldsymbol{x}_t + \mathbf{w}_{hc}\boldsymbol{h}_{t-1} + \boldsymbol{b}_c) \quad (3)$$

$$\boldsymbol{o}_t = \sigma(\mathbf{w}_{xo}\boldsymbol{x}_t + \mathbf{w}_{oh}\boldsymbol{h}_{t-1} + \mathbf{w}_{co} * \boldsymbol{c}_t + \boldsymbol{b}_o) \quad (4)$$

$$\boldsymbol{h}_t = \boldsymbol{o}_t * \tanh \boldsymbol{c}_t \quad (5)$$

The input $\boldsymbol{x} = (\boldsymbol{x}_1 \ldots \boldsymbol{x}_{t-1})$ is a T timestamp vector sequence and $\boldsymbol{h} = (\boldsymbol{h}_1 \ldots \boldsymbol{h}_{t-1})$ is the output vector sequence. $\boldsymbol{x}_t$ can be a one-dimensional vector or scalar, while $\boldsymbol{h}_t$ can be given different dimension. $\mathbf{w}_{xi} \sim \mathbf{w}_{co}$ are the weighted parameter matrices which conduct a linear transformation between the vector, $\boldsymbol{b}_i \sim \boldsymbol{b}_o$ are the intercept parameters. The operator '*' is Hadamard product, σ and tanh are the two nonlinear activation functions given by

$$\sigma = \frac{1}{1 + e^{-x}} \quad (6)$$

$$\tanh = \frac{e^x - e^{-x}}{e^x + e^{-x}}. \quad (7)$$

LSTM network can better capture the complex relationships between the inputs and outputs, where each layer regards the output from the last layer as the input of the next layer (Figs. 2, 3, 4). In this paper, each LSTM cell can be considered as a function $g^L : R^{T \times K} \rightarrow R^{T \times K'}$, where T is the length of time sequences, K' is the length of one output vector. The multiple LSTM cell in different layers will learn more characteristics and therefore improve the recognition rate.

However, a deep LSTM is not an ideal model for the spatiotemporal distribution of sharing bike with spatial characteristics in this paper, because it fails to capture the spatial dependences for these spatiotemporary variables.

4.2 Conv-LSTM

To overcome this shortcoming of spatial dependences, Conv-LSTM network, which combines CNN and LSTM in one end-to-end DL architecture, is proposed [29]. Unlike the LSTM, The conv-LSTM transforms all the inputs, intermediate variable, hidden states, and various gates from 2D matrices to 3D tensors. The input gate tensors, output gate tensors, forget gate tensors, the input tensors, hidden tensors, and memory cell tensors are denoted as 3D tensors $\boldsymbol{I}_t$, $\boldsymbol{O}_t$, $\boldsymbol{F}_t$, $\boldsymbol{X}_t$, $\boldsymbol{H}_t$, $\boldsymbol{C}_t \in R^{I \times J \times K}$. Besides, the Hadamard product converts to the convolutional operator, and thus the weight matrices $\mathbf{W}_{xi}, \mathbf{W}_{xf}, \mathbf{W}_{xc}, \mathbf{W}_{hc}, \mathbf{W}_{xo}, \mathbf{W}_{or}, \mathbf{W}_{co}$ will serve as convolutional flitters being used as the exploration of local correlations, which are replicated across the tensors with shared weights.

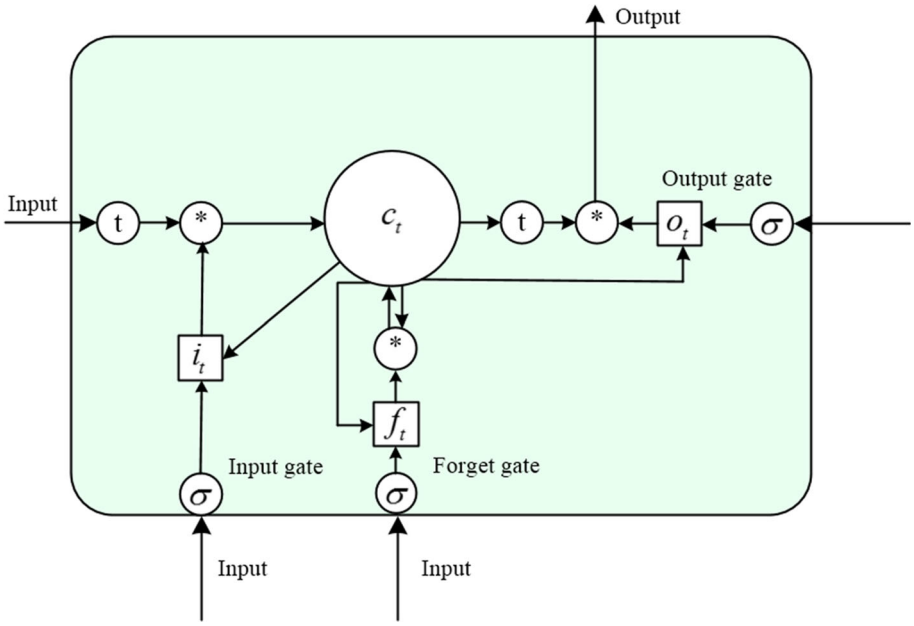
The structure of the conv-LSTM is to transform all the inputs, memory cell values, hidden states, and various gates to 3D tensors as follow:

$$\boldsymbol{I}_t = \sigma(\mathbf{W}_{xi} \circ \boldsymbol{X}_t + \mathbf{W}_{ir} \circ \boldsymbol{H}_{t-1} + \mathbf{W}_{ci} * \boldsymbol{C}_{t-1} + \boldsymbol{B}_i) \quad (8)$$

$$\boldsymbol{F}_t = \sigma(\mathbf{W}_{xf} \circ \boldsymbol{X}_t + \mathbf{W}_{hf} \circ \boldsymbol{H}_{t-1} + \mathbf{W}_{cf} * \boldsymbol{C}_{t-1} + \boldsymbol{B}_f) \quad (9)$$

Fig. 1 Dockless sharing bike: Mobike and ofo

Fig. 2 Structure of LSTM



$$\boldsymbol{C}_t = \boldsymbol{F}_t * \boldsymbol{C}_{t-1} + \boldsymbol{I}_t * \tanh(\mathbf{W}_{xc} \circ \boldsymbol{X}_t + \mathbf{W}_{hc} \circ \boldsymbol{H}_{t-1} + \boldsymbol{B}_c) \tag{10}$$

$$\boldsymbol{O}_t = \sigma(\mathbf{W}_{xo} \circ \boldsymbol{X}_t + \mathbf{W}_{oh} \circ \boldsymbol{H}_{t-1} + \mathbf{W}_{co} * \boldsymbol{C}_t + \boldsymbol{B}_o) \tag{11}$$

$$\boldsymbol{H}_t = \boldsymbol{O}_t * \tanh \boldsymbol{C}_t \tag{12}$$

Similar to LSTM, each conv-LSTM layer can map a sequence of input tensors. After T iterations, each conv-LSTM cell is denoted as a function $g^{CL} : R^{T\times I\times J\times K} \rightarrow R^{T\times I\times J\times K'}$, where T is the length of time sequences, I, J is the dimensions of rows and columns. The operator $\circ$ denotes the convolutional operator and $\mathbf{W}_{xi},\mathbf{W}_{xf},\mathbf{W}_{xc},\mathbf{W}_{hc},\mathbf{W}_{xo},\mathbf{W}_{or},\mathbf{W}_{co}$ serve as convolutional flitters.

In this section, we employ a convolutional LSTM network (conv-LSTM) in one end-to-end DL architecture and thus integrate spatiotemporal variables and non-spatial time series variables into short-term distribution forecasting for bike-sharing system. This model is end-to-end DL architecture, where are specially designed for non-spatial time series variables (time of day) and spatiotemporal variables. The corresponding repeating and transformation functions are added in the structure to fuse the two categories of variables.

4.2.1 Structure for spatiotemporal variables

In this paper, the number of bikes in area, distribution uniformity, and usage distribution are defined as spatiotemporal variables. By considering that the historical the number of bikes in area, distribution uniformity, and usage distribution influence the future distribution (i.e., the future number of bikes in area, distribution uniformity in grid area (i,j)) in different ways, the separate architectures will be used, each variable correspond to a set of independent stacked conv-LSTM layers. Assume the lookback time windows are K_d,K_u,K_v, the number of stacked conv-LSTM layers are l_d,l_u,l_v, respectively, for number of bikes in area, distribution uniformity and usage distribution. The

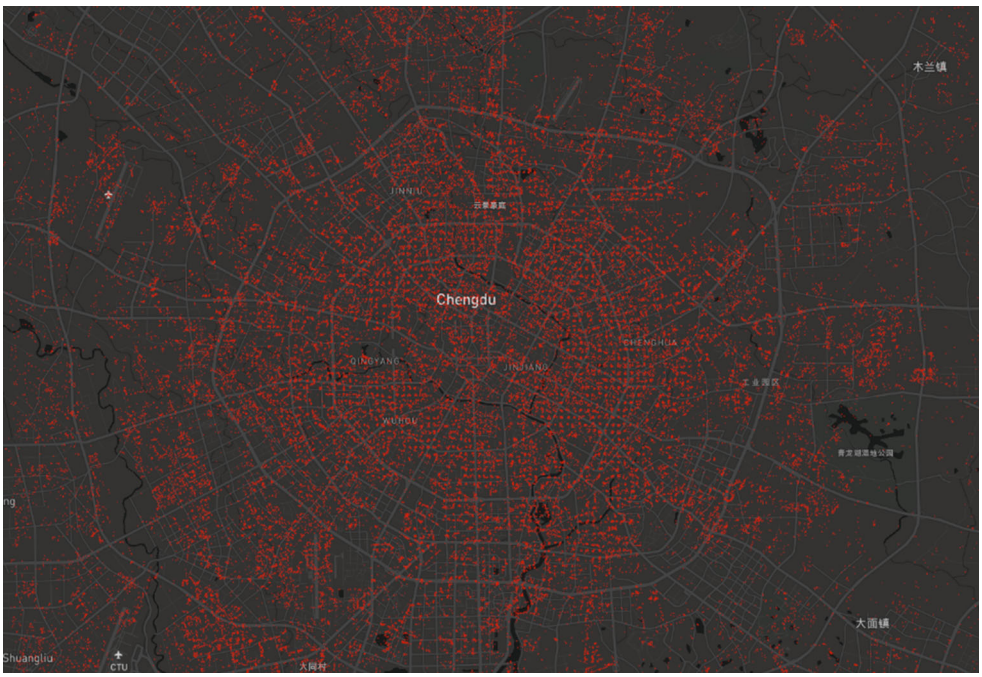


Fig. 3 Bike distribution for Mobike

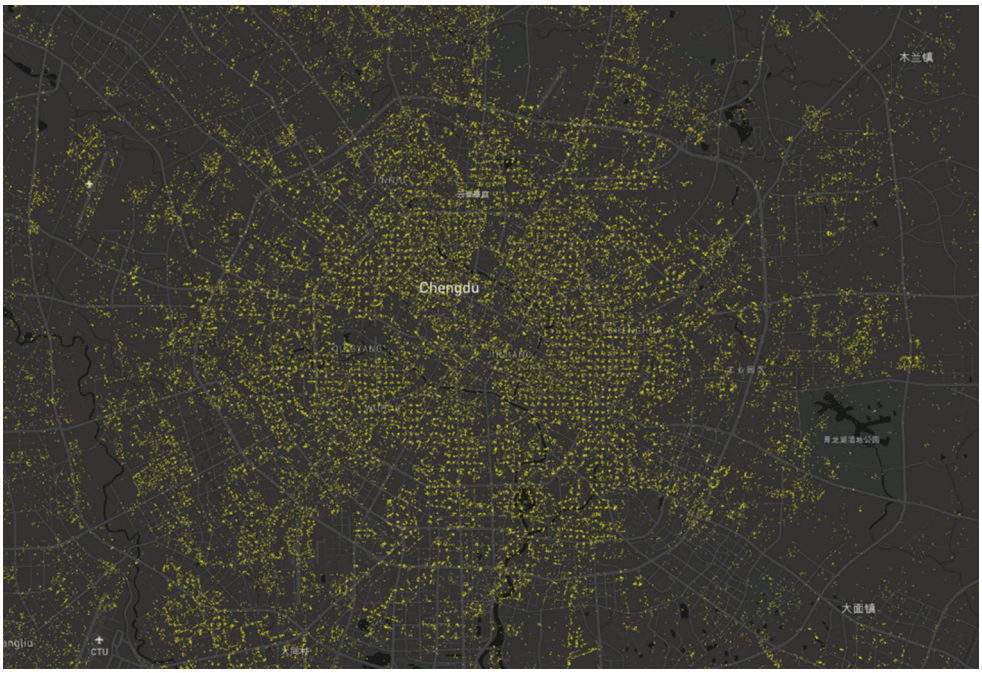


Fig. 4 Bike distribution for Ofo

formulations of the separate architecture are given, respectively:

$$(\boldsymbol{D}^{(l_d)}_{t-K_d}, \boldsymbol{D}^{(l_d)}_{t-K_d+1}), \ldots (\boldsymbol{D}^{(l_d)}_{t-2}, \boldsymbol{D}^{(l_d)}_{t-1}) = g^{CL}_{l_d} \ldots g^{CL}_{1}(\boldsymbol{V}_{t-K_d}, \boldsymbol{V}_{t-K_d+1}, \ldots, \boldsymbol{V}_{t-1}) \quad (13)$$

$$\boldsymbol{X}^{d}_{t} = \sigma(\mathbf{W}_{dx} * \boldsymbol{D}^{(l_d)}_{t-1} + \boldsymbol{B}_{d}) \quad (14)$$

$$(\boldsymbol{U}^{(l_u)}_{t-K_u}, \boldsymbol{U}^{(l_u)}_{t-K_u+1}), \ldots, (\boldsymbol{U}^{(l_u)}_{t-2}, \boldsymbol{U}^{(l_u)}_{t-1}) = g^{CL}_{l_u} \ldots g^{CL}_{1}(\boldsymbol{Z}_{t-K_u}, \boldsymbol{Z}_{t-K_u+1}, \ldots, \boldsymbol{Z}_{t-1}) \quad (15)$$

$$\boldsymbol{X}^{u}_{t} = \sigma(\mathbf{W}_{ux} * \boldsymbol{U}^{(l_u)}_{t-1} + \boldsymbol{B}_{u}) \quad (16)$$

$$(\boldsymbol{H}^{(l_r)}_{t-K_r}, \boldsymbol{H}^{(l_r)}_{t-K_r+1}), \ldots, (\boldsymbol{H}^{(l_r)}_{t-2}, \boldsymbol{H}^{(l_r)}_{t-1}) = g^{CL}_{l_h} \ldots g^{CL}_{1}(\boldsymbol{Q}_{t-K_h}, \boldsymbol{Q}_{t-K_h+1}, \ldots, \boldsymbol{Q}_{t-1}) \quad (17)$$

$$\boldsymbol{X}^{h}_{t} = \sigma(\mathbf{W}_{hx} * \boldsymbol{H}^{(l_h)}_{t-1} + \boldsymbol{B}_{h}) \quad (18)$$

where $\boldsymbol{D}^{(l_d)}_{t-K_d}$, $\boldsymbol{U}^{(l_u)}_{t-K_u}$, $\boldsymbol{H}^{(l_h)}_{t-K_h}$ are the output hidden tensors in the highest level layers of the architectures of number of bikes in area, distribution uniformity and usage

distribution, respectively. $\mathbf{W}_{dx}, \mathbf{W}_{ux}, \mathbf{W}_{hx}$ are convolutional operators to explore the spatial correlations of the highest level output tensors, while $\boldsymbol{B}_d, \boldsymbol{B}_u, \boldsymbol{B}_h$ are the intercept parameters. The high-level components $\boldsymbol{X}_t^d, \boldsymbol{X}_t^u, \boldsymbol{X}_t^h$ can be achieved by above structure.

4.2.2 Structure for time series variables

The time of day can be considered as non-spatial time series variables, which produce the high-level components $\boldsymbol{X}_t^p$:

$$(\boldsymbol{p}_{t-K_p}^{(l_p)}, \boldsymbol{p}_{t-K_p+1}^{(l_p)}), \ldots (\boldsymbol{p}_{t-2}^{(l_p)}, \boldsymbol{p}_{t-1}^{(l_p)}) = g_{l_p}^{L} \ldots g_1^{L}(\boldsymbol{\alpha}_{t-K_p}, \boldsymbol{\alpha}_{t-K_p+1}, \ldots, \boldsymbol{\alpha}_{t-1}) \quad (19)$$

$$\boldsymbol{X}_t^p = g^{\mathrm{T}}(g^R(\sigma(W_{px}\boldsymbol{p}_t^{(l_p)} + \boldsymbol{b}_p) \quad (20)$$

where $\boldsymbol{p}_{t-K_p}^{(l_p)}$ are the output hidden vectors in the highest LSTM layers. g^R are the repeating function $g^R(x; M, N) : R \rightarrow R^{M \times N \times 1}$ and note that g^{T} is applied to transfer 2D matrices to 3D tensors.

4.2.3 Fusion of spatiotemporal and time series variables

Considering that the high-level components have different impacts on the prediction, the Hadamard products are served to integrate these components:

$$\boldsymbol{X}_t = \mathbf{w}_d * \boldsymbol{X}_t^d + \mathbf{w}_u * \boldsymbol{X}_t^u + \mathbf{w}_h * \boldsymbol{X}_t^r + \mathbf{w}_p * \boldsymbol{X}_t^p$$

The weighted and intercept parameters can be deeply learnt by means of minimizing the mean squared error between the estimated and real value, The objective function of the architecture is as follow:

$$\min_{w,b} \|\bar{\boldsymbol{X}} - \boldsymbol{X}\|_2^2 + \beta \|\mathbf{W}\|_2^2$$

$\beta \|\mathbf{W}\|_2^2$ is an L2 norm regularization term in order to avoid over-fitting issues. $\mathbf{W}\psi$ represents the set of weighted parameters.

The training steps of the conv-LSTM are as follows:

Input number of bikes in area $\{\boldsymbol{D}_1, \ldots, \boldsymbol{D}_n\}$ in training dataset
Distribution uniformity $\{\boldsymbol{U}_1, \ldots, \boldsymbol{U}_n\}$ in training dataset
Usage distribution $\{\boldsymbol{H}_1, \ldots, \boldsymbol{H}_n\}$ in training dataset
Time of day $\{\boldsymbol{p}_1, \ldots, \boldsymbol{p}_n\}$ in training dataset
lookback windows: K_d, K_u, K_h, K_p
Output conv-LSTM with learnt parameters
Initialize a null set: $\boldsymbol{L}$
for time series $1 \leq t \leq n$ **do**
$\{\boldsymbol{D}_{t-K_p}, \ldots, \boldsymbol{D}_{t-1}\} \rightarrow \boldsymbol{E}_t^d$
$\{\boldsymbol{U}_{t-K_p}, \ldots, \boldsymbol{U}_{t-1}\} \rightarrow \boldsymbol{E}_t^u$
$\{\boldsymbol{V}_{t-K_p}, \ldots, \boldsymbol{V}_{t-1}\} \rightarrow \boldsymbol{E}_t^v$
$\{\boldsymbol{p}_{t-K_p}, \ldots, \boldsymbol{p}_t\} \rightarrow \boldsymbol{E}_t^p$
A training dataset $\{\{\boldsymbol{E}_t^d, \boldsymbol{E}_t^u, \boldsymbol{E}_t^v, \boldsymbol{E}_t^p\}, \boldsymbol{D}_t, \boldsymbol{U}_t\}$ is put into $\boldsymbol{L}$
end for
Initialize all the weighted and intercept parameters
repeat
Randomly extract a batch of samples $\boldsymbol{L}^r$ from $\boldsymbol{L}$
Minimizing the objective function to achieve the parameters within $\boldsymbol{L}^r$
until convergence criterion met
end procedure

5 Experiments and results

5.1 Data preprocessing

The data mining of the large data of cycling of Chengdu City has been raised by APP of dockless sharing bike system. Chengdu (CTU) is the biggest city in the southwest of China, and it held about 16,000,000 people until 2017. According to the latest statistics, more than 600,000 dockless sharing bikes are in service around the city of Chengdu. And these bikes will upload the location information when be called for a service or for stopping a trip, and the data center will receive the data from one bike several times in 1 day.

As shown in Fig. 5, it is obvious that there is little difference in the riding distance distribution between the two brands of dockless bike sharing. The riding distance is mainly concentrated in the range of 800–2800 m, where the range of 800–2000 m is the peak interval for riding behavior. The average riding distance of dockless bike-sharing system is about 2300 m. The situation to the riding time distribution is similar to that of the riding distance, where the general trend is basically the same between the two types of dockless bike-sharing. The peak interval of riding time is between 10 and 25 min. The average riding time is about 20 min (see Fig. 6).

The frequency of use in time of day shows a multipeak phenomenon which is related to morning peak, non-commuter peak, and evening peak. The morning peak of frequency appears between 7:00 a.m. and 8:00 a.m. The next peak is between 12:00 a.m. and 2:00 p.m., which can be considered the peak in the non-commuter period. The evening peak period broke out at 5:00 pm. After 9 pm, the bike use basically enters the dormancy period (see Figs. 7, 8).

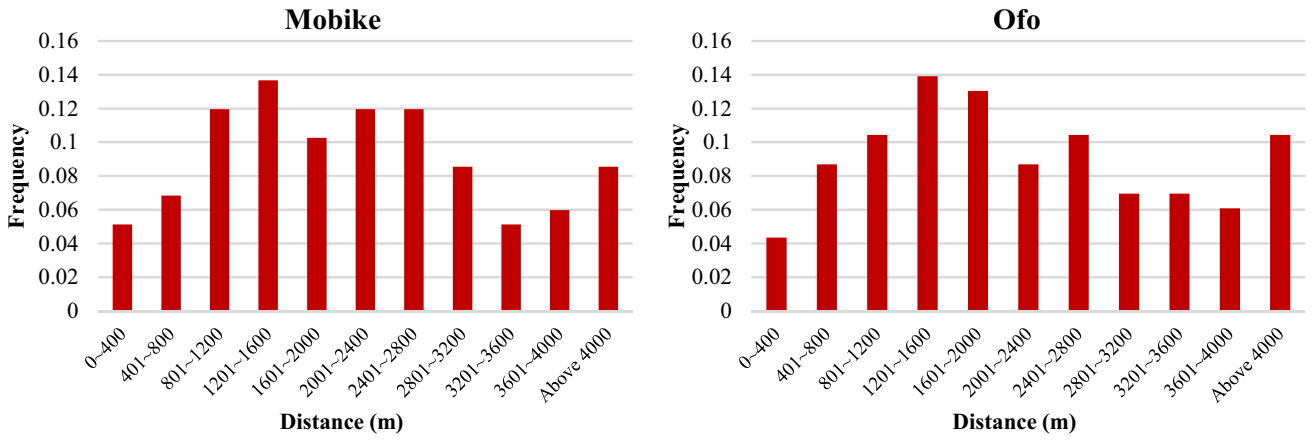


Fig. 5 Riding distance distribution

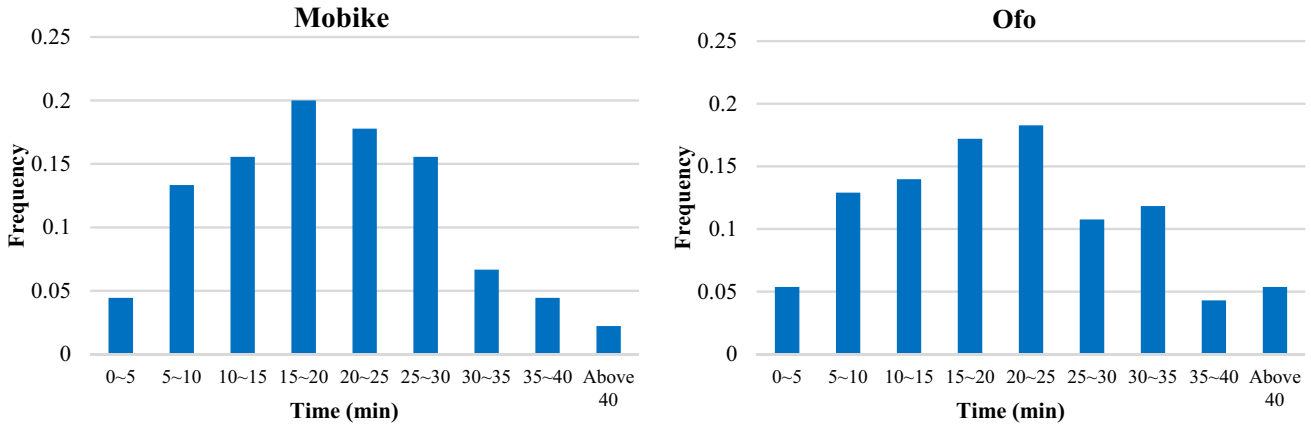


Fig. 6 Riding time distribution

5.2 Results

The shared bike will be collected for maintenance and reset in some special positions for service for a long time. And the places nearby subway station and beltway of city are often chosen for shared bikes' restarting a new work period. The interval of dataset for temporal series variables is set to 1 h, and the region is partitioned into 7 × 7 grids, which the size of each grid is 4 km × 4 km.

The conv-LSTM are trained on the training dataset and validated on the test dataset, respectively. Meanwhile, the LSTM network, which is only fed with the historical

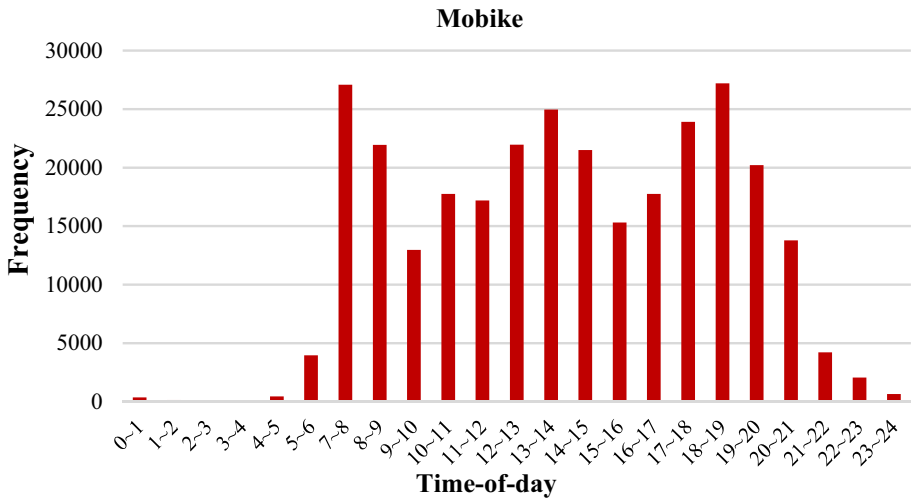


Fig. 7 Frequency of use in time of day for Mobike

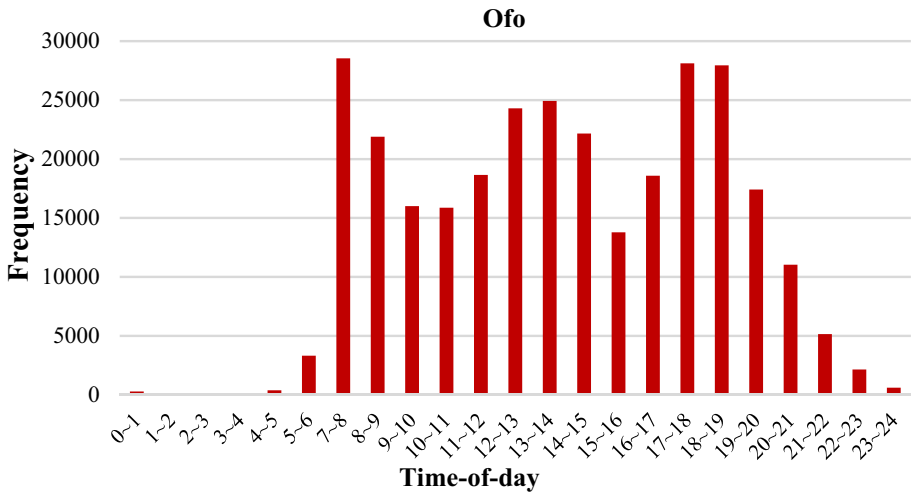


Fig. 8 Frequency of use in time of day for Ofo

Table 1 Predictive performance comparison

Model	RMSE	MAE
LSTM	0.0341	0.0189
Conv-LSTM	0.0314	0.0173

distribution dataset, is trained and tested in the same way. The definitions of the above-mentioned two models are shown as follows:

LSTM: We employ the structure as the unconditional future predictor model with two 2048-node LSTM layers. The LSTM embraces all the variables, including number of bikes in area, distribution uniformity, usage distribution, time of day, with lookback time window $K = 8$, of grid (i,j), to predict future demand intensity in grid (i,j).

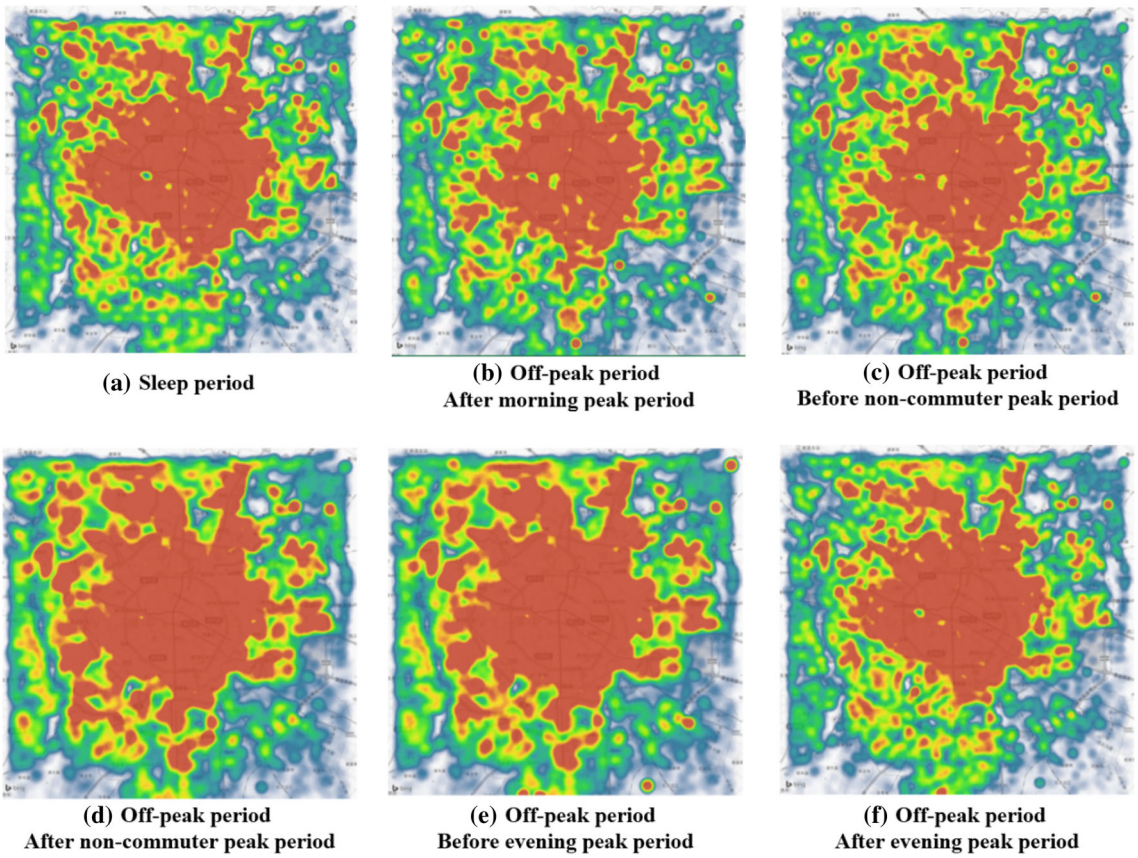


Fig. 9 Actual distribution on map. **a** Sleep period, **b** off-peak period after morning peak period **c** off-peak period before non-commuter peak period, **d** off-peak period after non-commuter peak period, **e** off-peak period befor evening peak period, **f** off-peak period after evening peak period

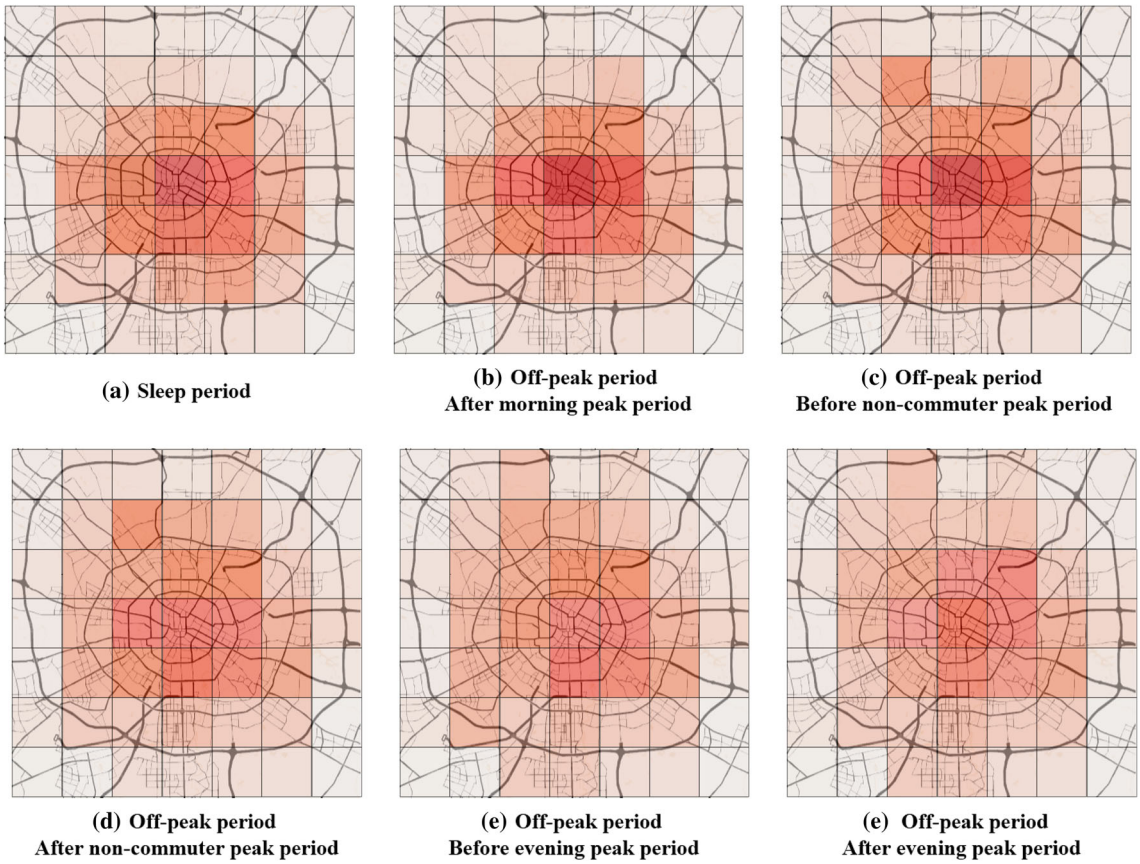


Fig. 10 Predicted distribution on map. **a** Sleep period, **b** Off-peak period after morning peak period, **c** off-peak period before non-commuter peak period, **d** off-peak period after non-commuter peak period, **e** off-peak period befor evening peak period, **f** off-peak period after evening peak period

However, LSTM considers temporal dependences better than capture spatial dependences.

Conv-LSTM: we use the same structure as [29], where we set the patch size to 4 × 4 with each 64 × 64 frame represented by a 16 × 16 × 16 tensor. We test three variants of our model with different number of layers. The model utilizes historical observations of number of bikes in area, distribution uniformity, usage distribution, and time of day to predict future number of bikes in area and distribution uniformity. The architecture of conv-LSTM is introduced in Sect. 4.2.

Root-mean-squared error (RMSE) and mean absolute error (MAE) calculations will be conducted to evaluate the effectiveness of the model. Furthermore, number of bikes in area, distribution uniformity, usage distribution, and time of day are standardized to the range [0, 1].

Table 1 shows the predictive performance comparison results between LSTM and conv-LSTM. It can be found that the conv-LSTM outperforms LSTM.

Figure 9 shows the heat maps of the ground truth of bikes distribution on some landmark moments, where the deeper color implies a larger distribution intensity. An interesting scene can be found that bikes are gathered from the city to the center of the city during the non-commuter peak period (noon peak period). Furthermore, different rush hours will lead different distribution characteristics of dockless bike-sharing system. By contrast, Fig. 10 shows predicted results that the good fitting effect is realized by conv-LSTM to actual distribution in Fig. 9, which implies that the conv-LSTM is better to capture the spatiotemporal characteristics of bike distribution. Figure 11 shows the comparison of the truth and predicted uniformity on different grid, and it also shows good predictive performance.

The combination of short-term distribution and uniformity forecasting as well as visualization helps bike-sharing operators and government to detect and forecast grids with oversupply and overfull demand and design a more reasonable redistribution strategy to avoid these imbalanced conditions.

6 Conclusions

In this paper, we employ the deep learning approach, conv-LSTM with an end-to-end architecture, for short-term distribution forecasting under dockless bike-sharing

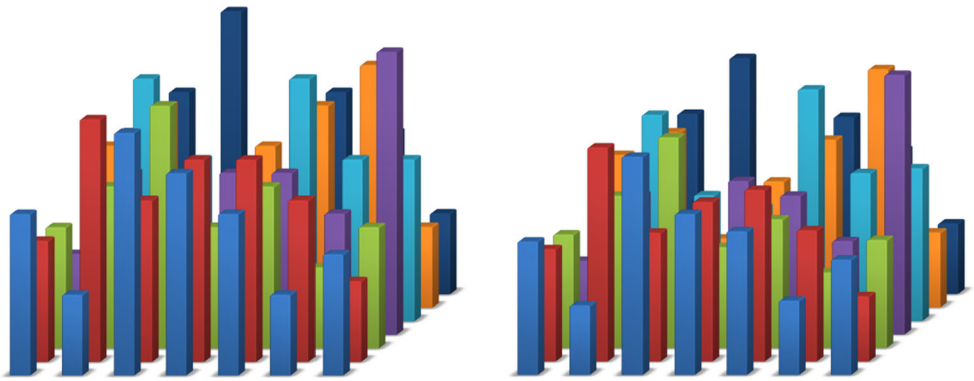


Fig. 11 Comparison of the truth and predicted uniformity on different grid

system. The architecture is fused by the spatiotemporal variables including number of bikes in area, distribution uniformity, usage distribution, and time of day. We trained LSTM and conv-LSTM with the same dataset, the results of which show that the conv-LSTM significantly outperform the LSTM. The comparison indicating that the employ DL approach performs better at capturing the spatiotemporal characteristics in short-term distribution forecasting.

This paper explores short-term distribution forecasting under dockless bike-sharing system via a novel spatiotemporal DL approach within an end-to-end architecture. Accurate real-time bike distribution forecasting can provide advice for user to better plan travel strategies and meet the travel needs. Furthermore, it also contributes to the implementation of a dynamic redistribution strategy to make bikes more balanced distributed according to demand. However, the model cannot explain the effect of supply and demand relationship on bike distribution and uniformity through a theoretical framework. In the future, we expect to build a theoretical model on this problem. Furthermore, we will also explore the relationship between the distribution of bicycles and the weather, season and events.

We declare that we do not have any commercial or associative interest that represents a conflict of interest in connection with the work submitted.

References

1. http://www.cityweekend.com.cn/shanghai/article/mobike-vs-mobike-lite-vs-ofo
2. http://daxueconsulting.com/mobike-and-ofo-bike-sharing/
3. Pidd H, Lavelle D (2017) Chinese bike-share scheme launches in rainy Manchester. The Guardian. Retrieved 6 July 2017
4. Mobike bicycle-hire scheme to be launched in Greater Manchester. BBC. 13 June 2017. Retrieved 6 July 2017
5. Cox, C (2017) Everything you need to know about the new Mobike service coming to Manchester and Salford. Manchester Evening News
6. Mobike launches in Sapporo. Shanghai Daily. Shanghai Daily. 23 August 2017. Retrieved 24 August 2017
7. https://en.wikipedia.org/wiki/Ofo_(bike_sharing)
8. Bai Y, Sun Z, Zeng B, Deng J, Li C (2017) A multi-pattern deep fusion model for short-term bus passenger flow forecasting. Appl Soft Comput 58:669–680
9. Ma Z, Xing J, Mesbah M, Ferreira L (2014) Predicting short-term bus passenger demand using a pattern hybrid approach. Transp Res Part C 39(39):148–163
10. Wei Y, Chen MC (2012) Forecasting the short-term metro passenger flow with empirical mode decomposition and neural networks. Transp Res Part C 21(1):148–162
11. Yang H, Leung CW, Wong SC, Bell MG (2010) Equilibria of bilateral taxi–customer searching and meeting on networks. Transp Res Part B 44:1067–1083
12. Yang H, Yang T (2011) Equilibrium properties of taxi markets with search frictions. Transp Res Part B 45:696–713
13. Ke JT, Zheng HY, Yang H et al (2017) Short-term forecasting of passenger demand under on-demand ride services: a spatio-temporal deep learning approach. Transp Res C Emerg Technol 85:591–608
14. Wu Y, Tan H (2016) Short-term traffic flow forecasting with spatial–temporal correlation in a hybrid deep learning framework. arXiv preprint arXiv:1612.01022
15. Zhang J, Zheng Y, Qi D (2016) Deep spatio-temporal residual networks for citywide crowd flows prediction. arXiv preprint arXiv:1610.00081
16. Shaheen S, Zhang H, Martin E, Guzman S (2011) China's Hangzhou public bicycle: understanding early adoption and behavioral response to bikesharing. Transp Res Rec J Transp Res Board 2247:33–41
17. Caggiani L, Ottomanelli M (2013) A dynamic simulation based model for optimal fleet repositioning in bike-sharing systems. Proc Soc Behav Sci 87:203–210
18. Caggiani L, Ottomanelli M (2012) A modular soft computing based method for vehicles repositioning in bike-sharing systems. Proc Soc Behav Sci 54:675–684
19. Zeng D, Xu J, Gu J, Liu L, Xu G (2008) Short term traffic flow prediction based on online learning SVR. In: Workshop on power electronics and intelligent transportation system, p 616–620
20. Smith B, Demetsky M (1997) Traffic flow forecasting: comparison of modeling approaches. J Transp Eng 123(4):261–266
21. Sun Y, Leng B, Guan W (2015) A novel wavelet-SVM short-time passenger flow prediction in Beijing subway system. Neurocomputing 166:109–121
22. Sapankevych N, Sankar R (2009) Time series prediction using support vector machines: a survey. IEEE Intell Syst 4(2):24–38
23. Yu H, Yang J, Han J, Li X (2005) Making SVMs scalable to large data sets using hierarchical cluster indexing. Data Min Knowl Disc 11(3):295–321

24. Zhao K, Khryashchev D, Freire J, Silva C, Vo H (2016) Predicting taxi demand at high spatial resolution: approaching the limit of predictability. In: IEEE international conference on big data
25. 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
26. Huang W, Song G, Hong H, Xie K (2014) Deep architecture for traffic flow prediction: deep belief networks with multitask learning. IEEE Trans Intell Transp Syst 15:2191–2201
27. Graham B (2014) Spatially-sparse convolutional neural networks. Comput Sci 34(6):864–867
28. Ma X, Tao Z, Wang Y, Yu H, Wang Y (2015) Long short-term memory neural network for traffic speed prediction using remote microwave sensor data. Transp Res Part C Emerg Technol 54:187–197
29. Shi X, Chen Z, Wang H, Yeung DY, Wong WK, Woo WC (2015) Convolutional LSTM network: a machine learning approach for precipitation nowcasting. In: Advances in neural information processing systems, p 802–810