IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received February 11, 2020, accepted April 10, 2020, date of publication April 14, 2020, date of current version May 19, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2987934

A Deep Learning Based Multi-Block Hybrid Model for Bike-Sharing Supply-Demand Prediction

MIAO XU, HONGFEI LIU, AND HONGBO YANG
School of Transportation, Jilin University, Changchun 130022, China

Corresponding author: Hongfei Liu (hongfeiliu@jlu.edu.cn)

This work was supported by the National Key Research and Development Program of China under Grant 2018YFB1601600, and in part by the Joint Laboratory for Future Transport and Urban Computing of Amap.

ABSTRACT As a new type of short distance commuting, the station-free sharing bike effectively alleviates urban traffic congestion. Thus, they are deployed in a large scale in many cities. However, various complex factors, including spatial, temporal, and other external information, result in serious imbalance of supply and demand between regions, which makes accurate prediction a challenging issue. In this study, our primary objective is to accurately forecast supply and demand by leveraging multi-source datasets. Based on the visual analysis about spatial-temporal characteristics of GPS data in Shanghai, we presented the innovative methods of Area of Interest grading and Traffic Analysis Zone division of bike-sharing, and revealed the distribution characteristics of sharing bike trips. A multi-block hybrid model where three blocks were separately modeled according to different data types was proposed. Moreover, eight state-of-the-art models and two variant models were developed as benchmarks to compare and evaluate the proposed model. The results suggested that MBH outperforms ten baselines with the highest accuracy. In addition, we conducted practical application of prediction results to validate that the proposed model could provide effective information for scheduling and rebalancing of bike-sharing system.

INDEX TERMS Bike-sharing, data visualization, deep learning, supply-demand prediction, spatial-temporal analysis.

I. INTRODUCTION

With the development of sharing economy, the station-free sharing bike has emerged as a novel and zero-emission short-distance commute way for urban residents. As of 2018, the number of station-free sharing bike users in China had reached 235 million. Comparing with the traditional sharing bike with docking station, this new type of sharing bike can be parked anywhere designated and can be unlocked by anyone who scans the Quick Response code on the bike. However, convenience and freedom also bring limitations. The imbalanced distribution of bike-sharing system (BSS) is more likely to occur due to the fluctuation of spatial-temporal demand. Some areas with excessive parking resulted in a large amount of ''invalid demand'', others with short supply could not meet users' needs, thereby increasing the operating costs of rebalancing.

At present, management departments mainly use large trucks to transport sharing bike and update the whole system continuously. However, this solution is time-consuming and laborious. Moreover, it has significant space-time delays. To solve this problem, the best way is to build prediction model based on big data and machine learning, which has been verified through many existing researches in the intelligent transportation systems (ITS) [1]–[5].

In this paper, we studied the bike-sharing supply-demand prediction problem, which related to predict the number of bikes rented and returned for a region in future by using historical multi-source datasets. In recent years, urban computing has been widely studied in the field of ITS [6]–[7]. Meanwhile, many works have been done on urban traffic prediction in terms of bike-sharing system. At first, how to analyze the correlation between multi-source data and prediction results is one challenging issue for accurate prediction. Some researchers did a lot of studies on the influencing factors of bike-sharing system. For instance, Gebhart and Noland [8] analyzed the influence of weather factors on the use of sharing bike. Bachand-marleau *et al.* [9] investigated social economy and spatial factors, and further analyzed their influence on the usage frequency. Besides, the univariate

The associate editor coordinating the review of this manuscript and approving it for publication was Jenny Mahoney.

regression algorithms and multivariate regression algorithms were used by Ashqar *et al.* [10] to model available bikes at each station and at the spatially correlated stations of each region, respectively. While these studies failed to capture the complex spatial-temporal features, they did clarify the importance of external factors in prediction model.

Furthermore, the data received are subject to constant changes in time and space as the complexity of transportation system itself. How to explore the fluctuation of spatial-temporal data and model the complex nonlinear relationship is another challenging issue. Fortunately, recent advances in deep learning modeling have motivated many researches on developing methods for traffic prediction [11]. The study of traffic forecasting model is fallen into three main directions: spatial, temporal and spatial-temporal feature capturing. On the one hand, convolutional neural network (CNN) showed its powerful ability in image and video recognition [12]. Inspired by this, some researchers treated the traffic information as an image to capture the spatial features by using CNN. Furthermore, Lin *et al.* [13] proposed a novel Graph Convolutional Neural Network (GCNN) that could learn hidden heterogeneous pairwise correlations between stations to predict station-level hourly demand. On the other hand, recurrent neural network (RNN) was widely used to address time series dataset. For example, a sequential learning task was well done in [14]. Further, Ma *et al*. [15] used long short-term memory neural network (LSTM) which could address the problem of back-propagated error decay of RNN for traffic speed prediction. Later, many researchers improved the structure of models by combining one neural network with another to achieve higher efficiency. For instance, in order to forecast passenger demand under on-demand ride service platform, a fusion convolutional long short-term memory network (FCL-Net) was proposed in [16] to capture spatial dependencies, temporal dependencies, and exogenous dependencies simultaneously. Afterwards, Yao *et al.* [17] presented Deep Multi-View Spatial-Temporal Network (DMVST-Net) framework by connecting CNN and LSTM, which can model both spatial and temporal relations.

The above-mentioned studies provide valuable insights for traffic demand prediction. However, the prediction model for bike-sharing system demand is not yet sound enough, and most of them are limited to a single model or rarely processed exogenous data according to different categories. We here propose a novel multi-block hybrid (MBH) model to consider both spatial-temporal feature and external context dependencies simultaneously in supply-demand prediction of BSS. The major contributions of this study are summarized as follows:

1) The novel MBH model characterized spatial, temporal, and spatial-temporal properties of multi-source data, captured three types of features by using CNN, GRU-Net, and ConvGRU-Net respectively, and coordinated them in a multi-block structure by feature level fusion.
2) Considering that the supply-demand of bike-sharing varied dynamically in different regions and time periods, we conducted visual analysis to understand the trip patterns of bike-sharing in Shanghai and utilized map meshing method to redefine TAZ that conforms to the distribution characteristics of bike-sharing.
3) As to external factors, both meteorological and geographical data were taken into account in our model. More specifically, we proposed a novel quantification method for POI data which used AOI grading method to represent the overlay distribution effect of six types POI. And we conducted regression analysis to explore the influence of temperature, wind speed, humidity, and other factors on the prediction results.
4) We carried out visual analysis of the prediction results for the first time and put forward corresponding feasibility suggestions, which verified the practical significance of this study.

The rest of this paper is organized as follows. Section II presents a comprehensive overview of related works. Section III describes the preliminary preparation of this study, including preprocessing and analysis of multi-source data. Section IV presents the architecture of multi-block hybrid model and the specific neural network for each block. We conduct experiments and discuss the results in Section V, and apply the prediction results in Section VI. Conclusions are drawn and future works are indicated in Section VII.

II. RELATED WORK

During the past decades, considerable efforts have been devoted to forecasting traffic related data, such as vehicle speed, traffic volume, taxi service, and the emerging bike-sharing supply or demand. The main research works for these different types of traffic prediction are the same which can be categorized into two aspects: data processing and prediction model. In this section, we will discuss the related work on the above research contents.

A. DATA PROCESSING

In this part, we believe that data processing includes external factors analysis and data preprocessing.

Numerous studies have been conducted to understand the contributory factors affecting the demand of sharing bike in recent years. Specifically, bike stations [10], the complex topological dependencies of road conditions [18], and POIs [19] were analyzed as spatial factors. As to temporal factors, Ashqar *et al*. [20] developed a method to quantify the effect of weather conditions on the prediction of bike station counts. Wu *et al*. [21] believed that weather may account for abnormal traffic drops. Other external context factors such

as social economy [9], traffic events [22], and cycling behaviors about gender [23] were also considered to analyze the hidden correlations. Moreover, as the basis of research work, data preprocessing appeared in almost every literature. Representatively, Zheng *et al*. [6] gave a detailed presentation of urban data and mentioned typical techniques used in urban computing.

While the above researches show that prediction can be improved by various factors, they still lack methods for quantifying nonlinear spatial-temporal correlations. In our study, we conducted regression analysis of meteorological factors and quantified POI by AOI grading to reveal implicit relationship between external factors and sharing bike usage. Further, the influencing factors were modeled separately according to the spatial-temporal type of data.

B. PREDICTION MODEL

Recent advances in deep learning have shown promising results in complex nonlinear relationships modeling. The current applications for traffic prediction in ITS mainly involve three directions: image processing [24]–[26], time series data processing [27]–[30], and multi-source fusion modeling [31]–[37].

Since the early 2000s, CNN has achieved success in object detection, image recognition, and image classification [24]. Computer vision has also expanded in transportation. Ma *et al*. [25] applied CNN to learn traffic as images and predicted traffic speed with a high accuracy. Further, Zhang *et al*. [26] proposed a convolution-based residual network to capture features on the images of traffic flow. The traditional methods to deal with time series data include ARIMA [38], RF [39], and SVM [40]. Recent studies further explored the utilities of advanced deep learning models. The first widely used model RNN was primarily utilized in natural language processing as studied in [28]. Further, to address the problem of back-propagated error decay of RNN, Sepp and Jürgen [29] popularized the application of LSTM. Xu *et al*. [30] used LSTM to develop a dynamic demand forecasting model for bike-sharing. A simpler model GRU used in [32] achieved faster computational efficiency than others. However, these studies were simply spatial or temporal modeling, and none of them considered both aspects simultaneously. Therefore, the CNN and GRU-Net in our model were only used to capture geographical and meteorological features, respectively.

Recently, several studies have tried to improve prediction performance through aggregated models. For instance, Yang *et al*. [31] combined forest model and mobility model to evaluate sharing bike users' demand. Shi *et al*. [32] used Trajectory GRU to handle spatial and temporal dependency. Wang *et al*. [34] presented a hybrid model which connected layerwise structure and markov transition matrix to forecast traffic flow. These studies provided a valuable reference for spatial-temporal modeling. In this paper, we proposed a ConvGRU-Net that could capture spatial-temporal features and improve the computational efficiency.

In summary, the advantage of our proposed method compared with existing literatures is that we considered both spatial relation and temporal dependency in a novel multi-block model, which built each block based on different data types.

III. PRELIMINARY

A. DATA SOURCES

This study uses the multi-source data collected from Shanghai which include the following three types: bike-sharing GPS data, geographic data, and meteorological data.

A trip information records ID number, longitude, latitude, timestamp, and lock status of each sharing bike. Users can scan the QR code to use bike. When unlocking, the lock status is marked as 0, and when closing, it is 1. We cleared out redundant data and deleted invalid information through Python programming. Further, we divided GPS data of Shanghai into 15 min, 30 min, 45 min, and 60 min intervals respectively to build four datasets. Therefore, we can guarantee more comparative experiments with limited databases and analyze the impact of time interval, which would be described specifically in Section V.

The geographic data in this study include administrative division, road network, and land use information. Arcgis Shape Files depicting the road networks attributes and municipal districts were obtained from the transportation system planning document. More specifically, considering the population density and bike-sharing usage of each district in Shanghai, we selected the area as shown in the black rectangular region in Fig. 1(a) for subsequent research. Then, we collected POI data in selected region from a crawler developed by Python 3.5.

In order to increase our model prediction performance, the meteorological data were collected from Weather Underground website which provides history weather observations for every half hour of a day. The specific information we selected includes weather condition, temperature, humidity, and wind speed.

B. PREPROCESSING

In this section, we further processed multi-source data by different methods in order to achieve accurate prediction.

1) TAZ DIVISION

There are many regionalization methods in terms of different semantic and granularities meanings, such as partition space by grid [26] or TAZ (Traffic Analysis Zone) [30]. However, the object discussed in this paper is sharing bike which mainly solves the last-kilometer problem in urban transportation. Based on this, we applied the map meshing method to redefine the TAZ that meets the characteristics of sharing bike. As shown in Fig. 1(b), we divided the selected region into 63×60 grids by using the Create_Fishnet Tool of ArcMap10.2. The length and width of each grid correspond to 1.021 kilometers and 1.016 kilometers on the actual map approximately. Therefore, the gird with an average area of 1 square kilometer is in line with the bike-sharing travel scope, which can effectively reflect the urban bike-sharing conditions.

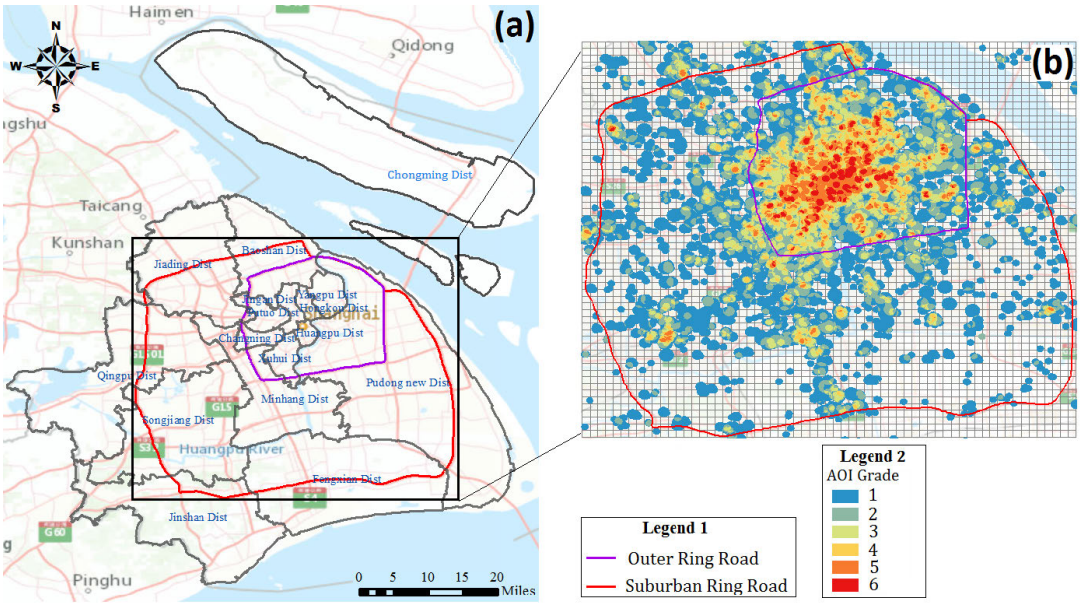


FIGURE 1. Regions in Shanghai. (a) Administrative map. (b) AOI grade distribution heat map. The grade number in Legend 2 represents the POI types covered by a region.

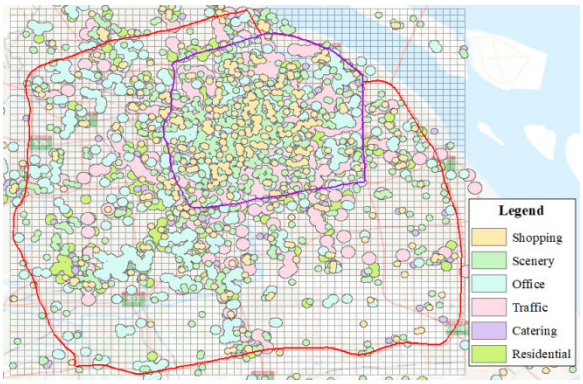


FIGURE 2. Diagram of POIs' buffer area. The 6 types of buffers produce distinct overlaps, especially in the central area.

TABLE 1. POI dataset collected in this paper.

POI type	Category	Buffer radius /m
Traffic	bus station, subway station	1000
Residential	community, flat, hotel	750
Office	university, school, company	750
Scenery	park, attractions	500
Catering	restaurant, café, snack bar	500
Shopping	plaza, store, supermarket	500

2) AOI GRADING

Urban POIs often directly affect residents' trip patterns, and each has its specific impact area and focus group. We selected six types of POI data and created buffers to represent each impact area through Proximity_Buffer Tool. The details of that are shown in Table 1. However, the six types of impact areas shown in Fig. 2 have distinct overlaps due to the different distribution density of POI, especially in the central area. We found that different land uses influence each other, and that the influence is regional, not a single point. The impact of land use on bike-sharing demand could not be quantified by using POI alone.

Hence, we came up with the conception of Area of Interest (AOI) which replaced point effects with region effects to represent the influence degree of geographic information data on a specific region. The AOI was utilized to express the effect of joint influence. More specifically, we used the Overlay_Union Tool to take intersection of six types of buffers, and then used the Symbology Option in Properties Layer to set the AOI grading distribution result shown in Fig. 1(b), which can directly reflect the POI density and obtain AOI grade in each TAZ.

In this part of study, there are three details should be noted. First, some urban POIs with low public awareness, which have little impact on residents' trip patterns, were not discussed. Further, the POI with important living functions and high public influence were chosen and divided into six types. Second, to ensure accuracy, we used the Geodesic Buffer algorithm in ArcMap to create circular buffer of POI. Afterwards, we set buffer radius according to the actual impact area of bike-sharing near different POI types in Shanghai. Third, the higher the AOI grade is, the denser the POI distribution is, and the greater the usage of bike-sharing is. The AOI grade of each TAZ was defined based on the maximum value of all AOI grades contained in the grid.

3) SPATIAL-TEMPORAL DATA VISUALIZATION

After TAZ division and AOI grading, we found that the trip of urban bike-sharing has obvious tidal phenomenon in a short time and periodic changes in a long time through

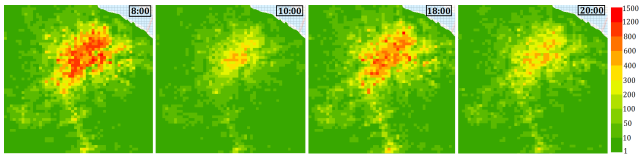


FIGURE 3. Spatial and temporal trip patterns (on Wednesday, Aug 26, 2018). In the upper right corner of maps, the grids outside coastline are padded by 0.

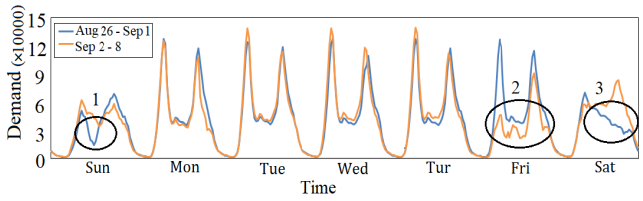


FIGURE 4. Effects of weather condition on bike-sharing demand (Mark 1: Aug 26; Mark 2: Sep 7; Mark 3: Sep 1).

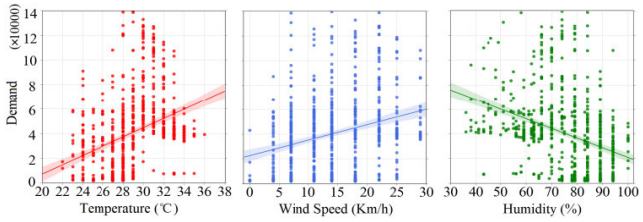


FIGURE 5. Linear regression of meteorological factors.

visual analysis [41]. In order to provide some intuitions, we present an example of bike-sharing demand for a working day in Fig. 3. The early peak and late peak appear at 8:00 and 18:00 respectively, which coincides with the peak time of urban trip. Moreover, spatial distribution is similar to the AOI grade distribution in Fig. 1(b), and the density is decreasing from central area to suburbs. Though there is a serious imbalance between different districts, supply and demand of bike-sharing in nearby grids may affect each other.

4) EXTERNAL FACTORS ANALYSIS

External factors, such as weather, land use patterns, and events, often have complex effects on traffic conditions. Compared with vehicles, open-air bikes are more susceptible to weather conditions. Fig. 4 and Table 2 illustrate the impact of adverse weather conditions on bike-sharing trips. Although the temporal dependency follows daily and weekly pattern, there are obvious non-periodic fluctuations in the corresponding time intervals. Taking mark 2 in Fig. 4 as an example, even on weekday, the early peak trip value reached its lowest due to rain, and the trip value gradually returned to normal until weather turned around.

In addition, we did a correlation analysis on other meteorological factors by using Seaborn which is a Python data visualization library based on Matplotlib. Fig. 5 shows the linear regression analysis results of temperature, wind speed, and humidity with demand. Although the influence of three

TABLE 2. Adverse weather conditions at different time.

Date	Time intervals	Condition
Aug 26	9:00-9:30;11:00-11:30; 12:30-13:00	Light Rain Shower
Sep 1	15:30-16:00;17:00-17:30	Light Rain with Thunder
	15:00-15:30;16:30-17:00	Heavy T-Storm
Sep 7	4:30-8:30;15:00-15: 30	Light Rain Shower
	9:00-10:00;11:00-11:30; 14:30-15:00;20:30-21:30	Light Rain
	12:30-14:00;16:00-16:30	Fog

Weather conditions not indicated are normal in all time intervals.

meteorological factors on bike-sharing demand is relatively scattered, it can be clearly seen that temperature and wind speed are positively correlated with demand, while humidity is negatively correlated with demand.

Weather conditions not indicated are normal in all time intervals.

Therefore, the effect of external factors in prediction model should not be neglected. To enhance prediction performance of our proposed model, meteorological data (temperature, humidity, wind speed, weather condition), metadata (hour, DayOfWeek), and AOI grade were added.

IV. METHODOLOGY

A. FORMULATION OF PROBLEM

In order to rebalance bike-sharing, we need to accurately predict the supply and demand within each grid through mathematical modeling. Thus, we fixed some notations and defined sharing bike supply-demand predication problem.

Definition 1 (Trip): A trip $Tr_k = (B_{id}, B_t, B_l, B_s)$ is a bike usage record, where B_{id} is the ID number of sharing bike, B_t is the timestamp, B_l donates the location including latitude and longitude, and B_s is the lock status of bike.

Definition 2 (Space and Time): The grid partitioned in Fig. 1(b) was defined as a spatial matrix $\boldsymbol{L}$ of $M \times N$, where a grid cell $l_{m\times n}$ donates a predication region. For simplicity, we use t instead of time intervals mentioned below.

Definition 3 (Supply and Demand): Let C be a set of historical record, the supply and demand at the time interval t in a grid $l_{m\times n}$ were defined respectively as:

$$y_t^{s,m,n} = \sum_{Tr\in C} |\{B_t \in t \wedge B_l \in l_{m\times n} \wedge B_s = 1\}| \quad (1)$$

$$y_t^{d,m,n} = \sum_{Tr\in C} |\{B_t \in t \wedge B_l \in l_{m\times n} \wedge B_s = 0\}| \quad (2)$$

Therefore, the supply-demand gap can be defined as:

$$G_t^{m,n} = y_t^{s,m,n} - y_t^{d,m,n} \quad (3)$$

Definition 4 (External Factors): In this study, the meteorology information and AOI grade were taken as external factors of our model. In terms of meteorology, we neglected the influence of geographical features, only considering the time series, and took the value every hour. On the contrary, POI was only related to geographical location. Thus, we defined a combined feature matrix.

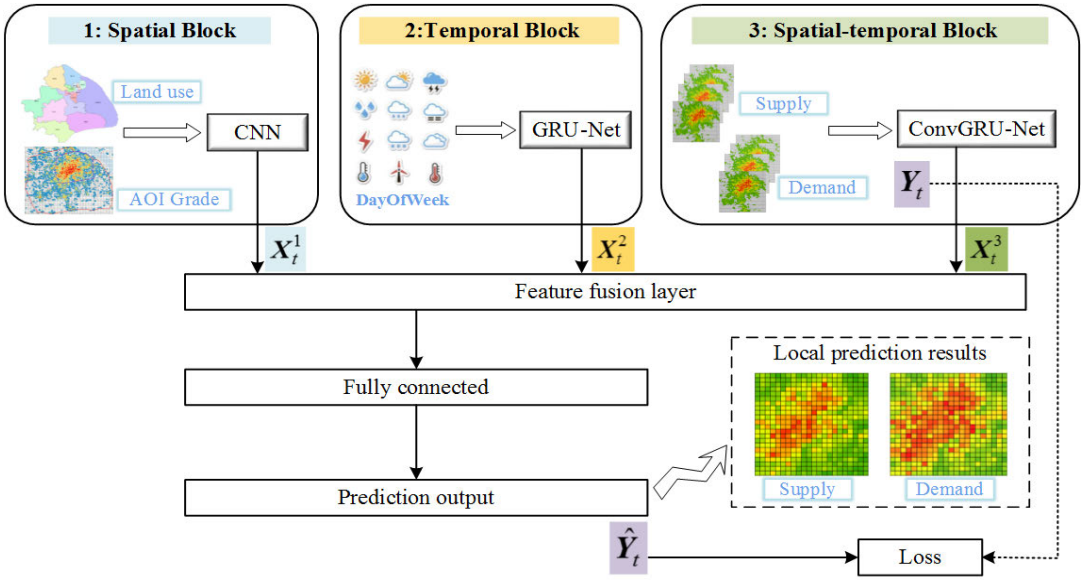


FIGURE 6. The architecture of MBH model.

FIGURE 7. Inner structure of CNN. FC: Fully-connected.

$\boldsymbol{E}_t = (W_t, T_t, S_t, H_t)$ where W_t, T_t, S_t, and H_t stand for the weather condition, temperature, wind speed, and humidity in period t respectively. The AOI grade of $l_{m\times n}$ was defined as $\boldsymbol{A}_{m,n}$.

Definition 5 (Predication Problem): The supply-demand predication aims to predict the supply-demand gap at time interval t, given a set of historical trip record $Tr = \{Tr_1, Tr_2, \cdots, Tr_k\}$ until time interval t-1.

B. THE STRUCTURE OF MBH MODEL

We constructed a multi-block hybrid (MBH) model for predicting supply-demand of bike-sharing, which involves three deep learning networks: the convolutional neural network (CNN), the gated recurrent neural network (GRU-Net), and the convolutional gated recurrent neural network (ConvGRU-Net). As shown in Fig. 6, each neural network is applied to one block, which is described as follows.

1) BLOCK1: SPATIAL MODELING WITH CNN

A city usually contains many districts whose population density and economic development level vary greatly. This determines different land use patterns and ultimately leads to uneven spatial distribution of bike-sharing. In this study, we conducted spatial modeling analysis for land use pattern. This type of variable is only spatially varied but temporally static during our study period. Thus, we built a CNN combination network to capture spatial feature.

The input is a feature map shown in Fig. 1b, of which grid cells have value of corresponding AOI grade. Fig. 7 shows the CNN structure used in this block. The extraction of spatial dependencies are performed mainly by the convolutional layer and pooling layer. We fed $\boldsymbol{A}^{m,n}$ which is described in Definition 4 to first convolution layer, then the k^{th} output can be calculated by:

$$\boldsymbol{A}^{m,n,k} = f(\boldsymbol{A}^{m,n,k-1} * \boldsymbol{W}^k + \boldsymbol{b}^k) \tag{4}$$

where $*$ donates the convolutional operator, f is an activation function such as the ReLU (Rectified Linear Unit) $f(x) = \max(0, x)$ we used in this study; $\boldsymbol{W}^k, \boldsymbol{b}^k$ are the parameters in the k^{th} layer.

In order to capture more spatial features, multiple kernels of a same size are set and scanned in the convolution layer simultaneously. Then, pooling layer is connected to convolution layer for reducing the spatial size of input feature maps and improving the robustness of the extracted features. Max pooling is a commonly used pooling method which takes the maximum value within a sliding region of given filter size [24]. Each pooling layer reduces dimension of the feature maps output from previous convolution layer, and obtains a vector of certain length as input of the next fully-connected layer. Finally, the captured spatial features of land use are catenated into a feature vector as the output.

2) BLOCK2: TEMPORAL MODELING WITH GRU-NET

In this block, we used gated recurrent unit (GRU) to capture the temporal sequential dependency of meteorology data, which is a variant of recurrent neural network (RNN) to address the exploding and vanishing gradient issue. The GRU is similar to Long short-term memory neural network (LSTM) in that it regulates information flow through

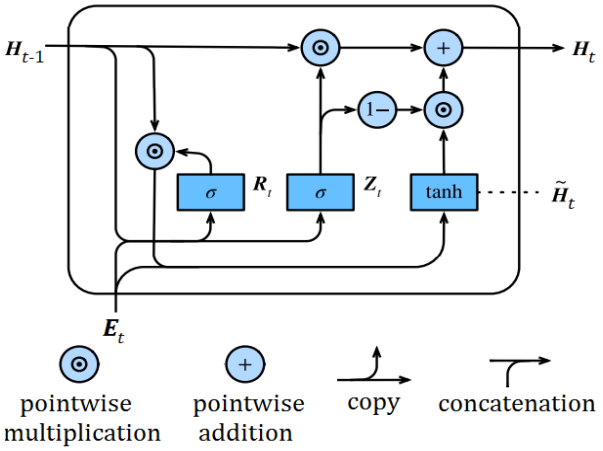


FIGURE 8. Inner structure of a GRU [42].

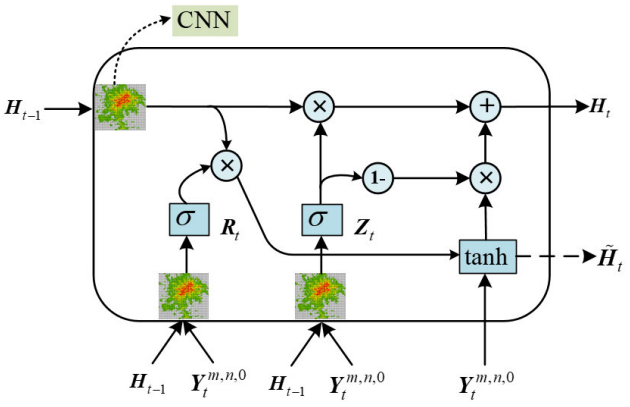


FIGURE 9. Inner structure of ConvGRU. Compared with GRU, the inner operator of ConvGRU is convolutional calculation.

sequence chain by the gate structure. However, it demonstrates more competitive performance and simpler structure than the standard LSTM. Previous work about short-term traffic speed prediction has also demonstrated the superiority of GRU-based model [33].

The typical structure of GRU is shown in Fig. 8, which only has two gates: reset gate $\boldsymbol{R}_t$ and update gate $\boldsymbol{Z}_t$. The

former captures short-term dependencies in time series, and the latter captures long-term dependencies. Both inputs are the current time step input $\boldsymbol{E}_t$ which is described in Definition 4 and the hidden state $\boldsymbol{H}_{t-1}$ of previous time step, and output is calculated by a fully- connected layer where the activation function is sigmoid function. The core idea is defined by (5) and (6) where $\boldsymbol{R}_t$ and $\boldsymbol{Z}_t$ make the range of each element [0, 1] through sigmoid function. Equation (7) calculates the candidate hidden state $\tilde{\boldsymbol{H}}_t$, in which the reset gate controls the flow of $\boldsymbol{H}_{t-1}$ by multiplying elements. Finally, update gate $\boldsymbol{Z}_t$ is combined with hidden state $\boldsymbol{H}_{t-1}$ and candidate hidden state $\tilde{\boldsymbol{H}}_t$ to obtain hidden state of time t.

$$\boldsymbol{R}_t = \sigma(\boldsymbol{E}_t \boldsymbol{W}_{xr} + \boldsymbol{H}_{t-1}\boldsymbol{W}_{hr} + \boldsymbol{b}_r) \quad (5)$$

$$\boldsymbol{Z}_t = \sigma(\boldsymbol{E}_t \boldsymbol{W}_{xz} + \boldsymbol{H}_{t-1}\boldsymbol{W}_{hz} + \boldsymbol{b}_z) \quad (6)$$

$$\tilde{\boldsymbol{H}}_t = \tanh(\boldsymbol{E}_t \boldsymbol{W}_{xh} + (\boldsymbol{R}_t \odot \boldsymbol{H}_{t-1})\boldsymbol{W}_{hh} + \boldsymbol{b}_h) \quad (7)$$

$$\boldsymbol{H}_t = \boldsymbol{Z}_t \odot \boldsymbol{H}_{t-1} + (1 - \boldsymbol{Z}_t) \odot \tilde{\boldsymbol{H}}_{t-1} \quad (8)$$

where $\boldsymbol{W}_\cdot$ denotes the weight matrix and $\boldsymbol{b}_\cdot$ indicates the bias. The operator $\odot$ denotes pointwise multiplication. σ is the sigmoid function. The output of each GRU layer is the hidden state at each time step.

3) BLOCK3: SPATIAL-TEMPORAL MODELING WITH CONVGRU-NET

In bike-sharing supply-demand prediction, the usage of bikes varies both temporally and spatially. This makes it impossible to extract temporal and spatial dependency simultaneously by using a single type neural network. To address this issue, we used a network named ConvGRU-Net, which combines CNN layer and GRU-Net layer to deal with spatial and temporal dependency. Specifically, as shown in Fig.9, the input of GRU-Net layer is the output of CNN layer. Note that the structure of neural networks we used here is similar with what we mentioned in block 1 and block 2.

The inputs of first spatial layer are some image-like matrices expressed as tensor $\boldsymbol{Y}_t \in \mathbb{R}^{2\times M\times N}$, which have both supply and demand channels (i.e., Definition 3), for each time interval t. Here the same CNN was used in block 1. After feeding $\boldsymbol{Y}_t^{m,n,0}$ into k convolutional layers, the output in time interval t is $\boldsymbol{Y}_t^{m,n,k}$ which forms the input of next temporal layer in our proposed network. Further, we stacked multiple ConvGRU layers to better capture spatiotemporal features among bike-sharing usage data in this study.

4) BLOCK FUSION AND LOSS FUNCTION

We then fused the output of three blocks. The spatial output extracted from CNN, the temporal output extracted from GRU-Net, and the spatial-temporal output extracted from ConvGRU-Net were concatenated into a dense vector in feature fusion layer. Using a parametric-matrix-based fusion method proposed in [18], the prediction output was finally calculated through fully-connected layer. The fusion formula is as follows:

$$\hat{\boldsymbol{Y}}_t = \boldsymbol{X}_t^1 \boldsymbol{W}_{b1} + \boldsymbol{X}_t^2 \boldsymbol{W}_{b2} + \boldsymbol{X}_t^3 \boldsymbol{W}_{b3} + \boldsymbol{b}_f \quad (9)$$

where $\boldsymbol{X}_t^\cdot$ is the output by each block at t time step, $\boldsymbol{W}_{b\cdot}$ denotes the weight matrix of each block, and $\boldsymbol{b}_f$ indicates the bias of fusion layer.

Since the proposed model is end-to-end, once we obtain the prediction output, we can optimize model through calculating loss function to minimize the mean squared error between real value and predicted value in each grid cell $l_{m\times n}$. Furthermore, we note that the output $\hat{\boldsymbol{Y}}_t$ includes both supply forecast $\hat{y}_t^s$ and demand forecast $\hat{y}_t^d$. The loss function is given as:

$$\boldsymbol{J}(\gamma) = (1-\gamma)(y_t^s - \hat{y}_t^s)^2 + \gamma(y_t^d - \hat{y}_t^d)^2 \quad (10)$$

TABLE 3. The specific information of datasets.

Datasets	Description	
Location	Central Districts of Shanghai	
Time span	8/26/2018 - 9/8/2018	
Time interval	15min, 30 min, 45min, 60min	
Base map	TAZ	274
	AOI	6 grades
Meteorology	Humidity (%)	[36,100]
	Wind speed (km/h)	[0,29]
	Temperature (℃)	[22,36]
	Weather conditions	4 types

where γ is a parameter to balance the effect of supply and demand. Both real value y_t^s, y_t^d and predicted value $\hat{y}_t^s, \hat{y}_t^d$ are calculated based on a whole region $\boldsymbol{L}$ of $M \times N$.

V. EXPERIMENTS

A. SETTINGS

1) DATASETS

In this study, we used the multi-source data processed in Section III for experimental analysis. Further, we named the three datasets as "Dataset-15", "Dataset-30", "Dataset-45", and "Dataset-60". The details of databsets are shown in Table 3.

For time series data, we divided the first 12 days' data into training set and validation set according to the ratio of 8:2, and took the remaining 2 days' data as the test set to verify the validity of our model. To improve the generalization ability of the model, we trained the parameters for several times.

2) BASELINES AND EVALUATION METRIC

In order to verify prediction performance of our proposed model, we compared MBH model with the single time series prediction models, the aggregated models, and two variant models of MBH, which are described as follows.

- **RNN:** A deep learning model which only considers the feature of spatial dimension. Following the practice in [15], we selected Adam optimization algorithm with learning rate $lr = 0.01$ and hyperparameters $\beta_1 = 0.9, \beta_2 = 0.999, \epsilon = 10e^{-8}$ to train model. The optimized model contained one input layer, three hidden layers with 64 hidden units in each layer, and one output layer.
- **LSTM:** It is an extension of RNN by introducing three "gates" to control the flow of data. The hidden layer is one LSTM layer with memory blocks and the other settings are the same as RNN.
- **GRU:** Compared with LSTM, GRU removes cell state and uses hidden state for information transmission. It only contains update gate and reset gate, which GRU can better capture the dependence of step distance in time series with a simpler model structure than LSTM. But the main parameters are the same as LSTM.
- **XGBoost [43]:** XGBoost is a powerful boosting tree based method, which performs well in various competitions. The following settings were used in the experiment: the number of trees is 50, the maximum depth is 4, and the learning rate is 0.002.
- **ConvLSTM [32]:** By extending the fully connected LSTM to have convolutional structures in both the input-to-state and state-to-state transitions, the convolutional LSTM is used to build an end-to-end trainable model for prediction. We set input-to-state and state-to-state kernel size to 5×5, and set 64 hidden states in each of the three hidden layers.
- **ConvGRU:** The structure of ConvGRU is similar to that of ConvLSTM, whose input are the output calculated by convolution operator.
- **ST-ResNet [26]:** An end-to-end structure named deep spatial-temporal residual networks can forecast crowd inflow and outflow in every region of a city.
- **DMVST-Net [17]:** Deep Multi-View Spatial-Temporal Network is proposed to predict taxi demand. We compared it with our model by feeding the datasets of this study.
- **MBH-1 (without block 1):** It is a variant of our proposed MBH model, in which we chose to ignore the influence of land use patterns.
- **MBH-2 (without block 2):** It is another variant of our proposed MBH model, in which we did not incorporate meteorological factors into the model.

The experimental results of all models were compared and analyzed by three classical evaluation metrics: mean absolute error (MAE), average percentage error (MAPE), and root mean square error (RMSE). The specific calculations are as follows:

$$MAE = \frac{1}{m}\sum_{i=1}^{m}\left|y_i - \hat{y}_i\right| \tag{11}$$

$$MAPE = \frac{1}{m}\sum_{i=1}^{m}\frac{\left|y_i - \hat{y}_i\right|}{\hat{y}_i} \tag{12}$$

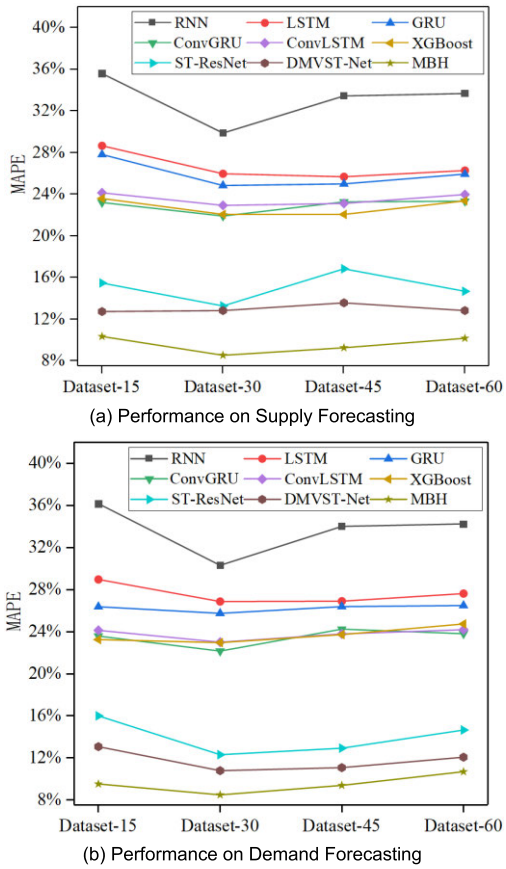
$$RMSE = \sqrt{\frac{1}{m}\sum_{i=1}^{m}\left(y_i - \hat{y}_i\right)^2} \tag{13}$$

where m is the number of test samples, y_i and $\hat{y}_i$ are the real and predicted value, respectively.

3) EXPERIMENTAL BASE

We note that the modeling process was under an open source Python distribution – Anoconda1.9.7, in which we used Sklearn to train XGBoost, RNN, LSTM and GRU by feature engineering and applied Keras based on the back end of Tensorflow to train other complex models by representation learning [44], [45]. All the training and testing processes were performed on a sever with CPU (Intel(R) Core(TM) i9-9900KF CPU @3.60GHz), 32-GB RAM, and GPU (NVIDIA Quadro P4000 with 8G memory).

Specifically, to make a fair comparison, the input data of different types were normalized in the range [0, 1] through



(a) Performance on Supply Forecasting

(b) Performance on Demand Forecasting

FIGURE 10. Comparison of prediction results of benchmark models on test samples.

Min-Max standardization. Whereafter, we denormalized the forecasting results for evaluation. The parameters of each model were set by referring to its early work in relevant literature. Further, the learning rate of three blocks in MBH was also set to 0.002, and all the training process could stop on training process to meet proper epoch.

B. MODEL COMPARISON

In this section, we compared the proposed MBH model with eight benchmark algorithms, including three traditional time series prediction models (i.e., RNN, LSTM, and GRU), a ensemble learning model based on regression tree (i.e., XGBoost), two widely used spatial-temporal models (i.e., ConvLSTM and ConvGRU), and two recent aggregated models (i.e., ST-ResNet and DMVST-Net). All the eight benchmark models were trained by using the same training samples to MHB training.

1) EFFECT OF TIME INTERVAL

Fig. 10 shows the supply and demand forecast results of MAPE on different models respectively. Comparing the forecast accuracy for four time intervals, we can see that the prediction effect of dataset divided by 30 min is significantly higher than that divided by 15 min, 45 min, and 60 min. A previous study [30] on bike-sharing prediction has also found that the best prediction was made at 30 min, but it indicated that the prediction performance would increase with time interval.

As shown in Fig. 10, we extend time interval to one hour and find that the prediction accuracy also declines when time interval is too long. The reason is that too short sampling period will cause more data noise and too long sampling period will ignore part of the fluctuation characteristics. Therefore, we believe that the sampling frequency of 30 min is the best, and too large or too small frequency will fail to reflect the rule of bike-sharing trips. The subsequent comparative experiments were all analyzed with ''Dataset-30'' as input.

2) FORECASTING PERFORMANCE COMPARISON

Table 4 and 5 show the supply and demand forecasting results of benchmark models on test samples. For illustrative purposes, we divided benchmark models into two groups and took one of the prediction results of supply and demand as examples. Fig. 11 further compared the prediction performance of eight benchmark models on test data samples. The following results can be concluded from comparisons.

a. It is obvious that the forecasting results of supply and demand are similar under the four datasets, and the MBH model produces higher prediction accuracy than other methods. Especially the highest accuracy of MBH reaches 91.51%. We calculated the mean value of supply and

demand MAPE, which demonstrate that MBH is relatively 16.78% better than GRU, 14.00% better than XGBoost, 13.52% better than ConvGRU, and 3.30% better than DMVST-Net.

b. Combined with CNN to capture spatial features, ConvGRU and ConvLSTM yield better performance than traditional time series prediction models. Additionally, due to the simpler structure, GRU algorithm has an advantage over LSTM in handling smaller data volume similar to this study.

c. Compared with aggregated models, the performance of

XGBoost is poor when data mutates. The reason why model failed to capture the abrupt fluctuations is that exogenous variables are fed into regression tree model through feature engineering as inputs and are not better explained in terms of spatial-temporal dependencies.

d. Although both ST-ResNet and DMVST-Net use spatial-temporal information, they are worse than MBH because they did not analyze and model exogenous variables individually by different categories, which also demonstrates the importance of spatial and temporal blocks in our model.

C. MODEL INTERPRETATION

To further pursuit the effect of spatial and temporal blocks modeling, we compare the proposed MBH model with two variants: MBH-1 (without block 1) and MBH-2 (without block 2), which are defined above.

Table 6 shows the prediction results of two variants on Dataset-30. MBH-2 has lower prediction performance than

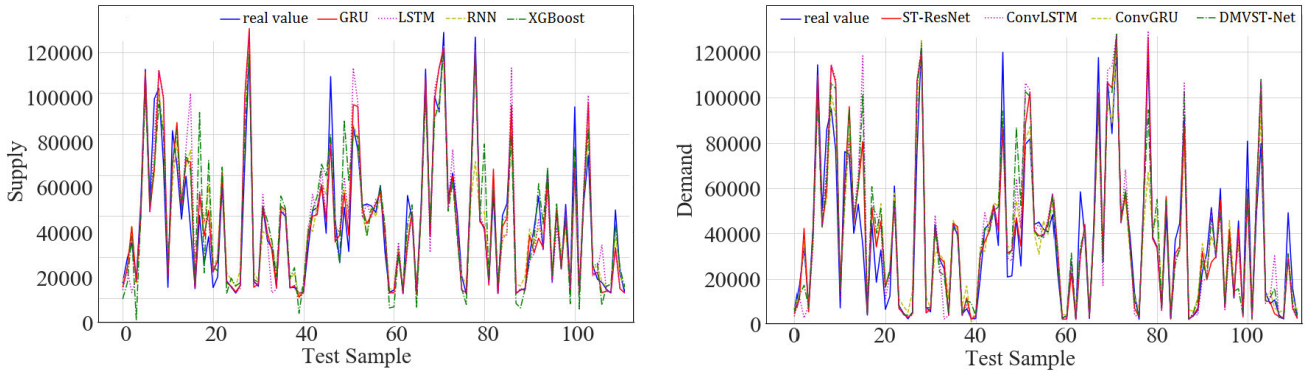


FIGURE 11. Forecasting visualization of benchmark models. We took the results of "Dataset-30" as example.

TABLE 4. Comparisons of related works.

Time Interval	Metric	RNN	LSTM	GRU	ConvLSTM	ConvGRU	XGBoost	ST-ResNet	DMVST-Net	**MBH**
Dataset-15	MAPE	35.59%	28.65%	27.82%	24.13%	23.2%	23.59%	15.47%	12.73%	**10.34%**
	RMSE	0.3973	0.3208	0.2965	0.2732	0.2561	0.2719	0.2108	0.1265	**0.1084**
	MAE	0.4168	0.3619	0.3601	0.3385	0.2952	0.3402	0.3011	0.2719	**0.2147**
Dataset-30	MAPE	29.87%	25.96%	24.84%	22.93%	21.9%	22.06%	13.27%	12.82%	**8.53%**
	RMSE	0.3125	0.2695	0.2622	0.2407	0.2287	0.2302	0.1910	0.0907	**0.0700**
	MAE	0.3510	0.3124	0.3091	0.2708	0.2259	0.2760	0.2347	0.2117	**0.1172**
Dataset-45	MAPE	33.44%	25.68%	24.99%	23.11%	23.27%	22.06%	16.84%	13.57%	**9.26%**
	RMSE	0.2953	0.2573	0.2412	0.2291	0.1992	0.2302	0.1985	0.1280	**0.0977**
	MAE	0.3692	0.3216	0.2999	0.2808	0.2293	0.2782	0.2562	0.2210	**0.1320**
Dataset-60	MAPE	33.67%	26.28%	25.93%	23.97%	23.33%	23.36%	14.68%	12.82%	**10.17%**
	RMSE	0.3819	0.3112	0.3097	0.2642	0.2386	0.2651	0.2025	0.1122	**0.1007**
	MAE	0.3701	0.3305	0.3004	0.2841	0.2329	0.2739	0.2562	0.2136	**0.1443**

TABLE 5. Performance comparison on demand forecasting of BSS.

Time Interval	Metric	RNN	LSTM	GRU	ConvLSTM	ConvGRU	XGBoost	ST-ResNet	DMVST-Net	**MBH**
Dataset-15	MAPE	36.18%	28.97%	26.38%	24.14%	23.59%	23.26%	16.00%	13.07%	**9.53%**
	RMSE	0.3899	0.3225	0.2891	0.2779	0.2607	0.2693	0.2095	0.1238	**0.1057**
	MAE	0.4227	0.3779	0.3569	0.3421	0.3001	0.3317	0.3105	0.2686	**0.2383**
Dataset-30	MAPE	30.32%	26.87%	25.75%	23.01%	22.16%	22.97%	12.31%	10.79%	**8.49%**
	RMSE	0.3209	0.2743	0.2722	0.2377	0.2201	0.2313	0.1909	0.0912	**0.0758**
	MAE	0.3499	0.3207	0.3003	0.2696	0.2304	0.2700	0.2411	0.2107	**0.1254**
Dtaset-45	MAPE	34.02%	26.91%	26.39%	23.79%	24.25%	23.73%	12.93%	11.08%	**9.39%**
	RMSE	0.2897	0.2602	0.2407	0.2312	0.2004	0.2350	0.1962	0.1325	**0.0906**
	MAE	0.3697	0.3226	0.3018	0.2912	0.2341	0.2802	0.2633	0.2306	**0.1295**
Dtaset-60	MAPE	34.26%	27.63%	26.49%	24.20%	23.81%	24.75%	14.65%	12.07%	**10.69%**
	RMSE	0.3800	0.3160	0.3047	0.2608	0.2421	0.2663	0.2010	0.1083	**0.1071**
	MAE	0.3723	0.3277	0.2989	0.2902	0.2327	0.2782	0.2604	0.2117	**0.1587**

TABLE 6. Performance comparison of MBH variants on Dataset-30.

	Supply			Demand		
	MAPE	RMSE	MAE	MAPE	RMSE	MAE
MBH-2	10.28%	0.0863	0.1947	11.10%	0.0900	0.2033
MBH-1	9.77%	0.0801	0.1393	10.83%	0.0816	0.1405

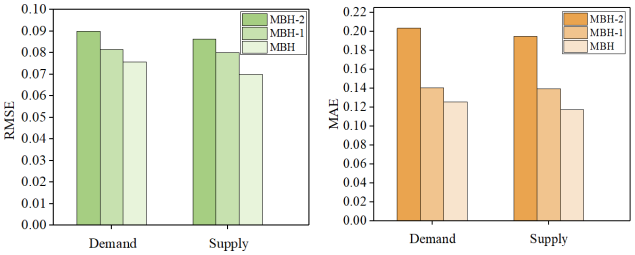


FIGURE 12. Prediction results of MBH and its variant models.

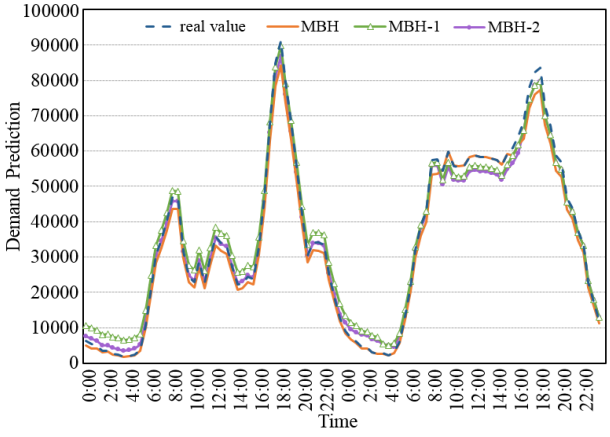


FIGURE 13. Forecasting visualization of MBH and its variant models.

MBH-1 in terms of all evaluation metrics. Obviously, meteorological factors have a greater impact on prediction accuracy than land use patterns. This is because the input of spatial-temporal block also implies the influence of some spatial factors. Fig. 12 indicates that the prediction performance of two variants are lower than the original model, which further proves the necessity of incorporating external factors into the prediction model.

Moreover, we visualize the forecasting curves to better evaluate the proposed model. As shown in Fig. 13, MBH generates accurate predictions at both peak and trough hours of the day. In addition, the variant models have poor performance when the data fluctuated greatly or at early morning from 0:00 to 4:00. This essentially means that exogenous variables can increase the robustness of MBH.

VI. APPLICATION OF PREDICTION RESULTS

The ultimate goal of our prediction model we have established is to obtain supply-demand gap of urban bike-sharing in time for distribution and rebalancing. Therefore, the

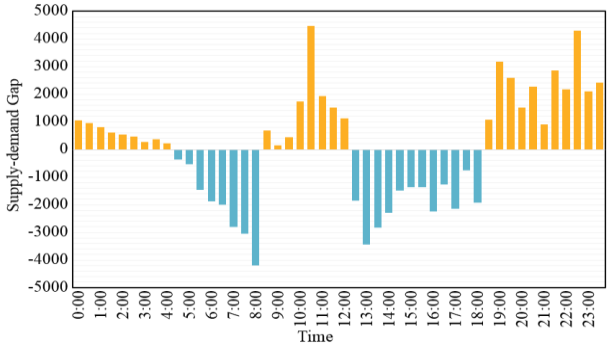


FIGURE 14. The predicted supply-demand gap.

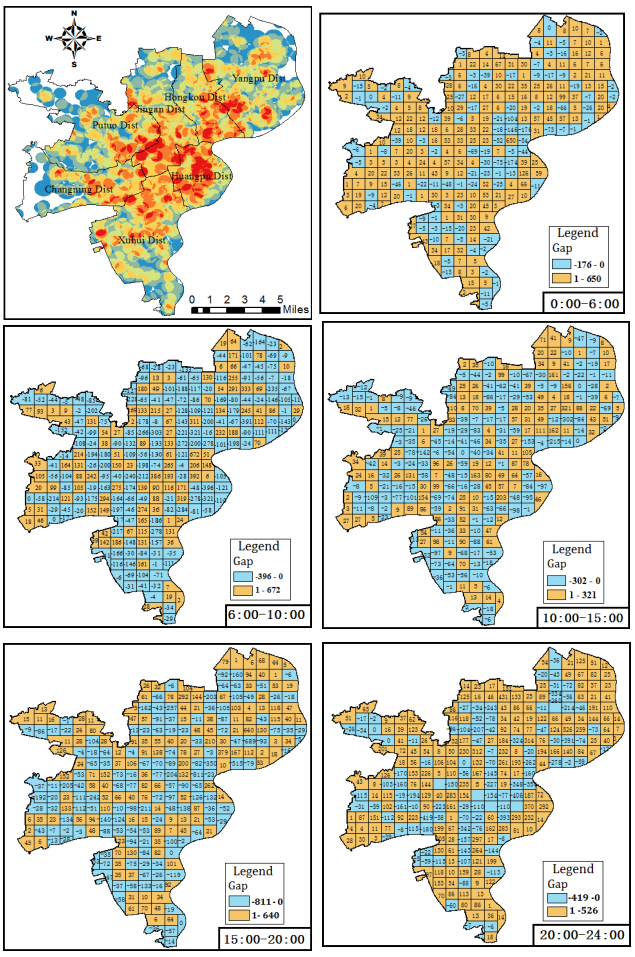


FIGURE 15. The spatial-temporal distributions of supply-demand gap over TAZs in central districts of Shanghai. TAZ displayed in blue is demand mode and yellow is supply mode.

predictive outputs of MBH model for the future day were calculated by (3) in Definition 3.

The results showed in Fig. 14 illustrate that 4:30-8:00 and 12:30-18:30 are the stages of demand, especially at 4:30-8:00 when demand continues to grow. A large amount of supply, however, is provided at 8:30-12:00 and after 19:00. To some extent, it reflects the tidal phenomenon of Shanghai

residents' trip and imbalance between supply and demand of bike-sharing. The developed forecast model can be utilized to analyze the spatial-temporal fluctuation demand of urban bike-sharing in advance and improve the operation efficiency of the system. A better practical application is to provide useful information for rebalancing.

For illustrative purposes, we extracted the AOI grade heat map showed in the first map on the upper left in Fig. 13 and TAZ distribution map of central area respectively from Fig. 1. The edge merging was performed and TAZs were labeled with a total number of 274 separately. The supply-demand gap of each TAZ was predicted by MBH model and visualized by Arcgis 10.2. The spatial-temporal distributions of supply-demand gap over 274 TAZs are shown below.

According to residents' commuting habits, a day was divided into 5 time periods as shown in Fig. 15. Referring to AOI grade heat map, demand mode appears in the higher grade areas at 0:00-6:00 and 20:00-24:00 while other areas present supply mode, but it is exactly opposite at 6:00-10:00 and 10:00-15:00. Especially at 15:00-20:00, supply-demand gap reached the highest value in the day. The bike-sharing distribution is uneven and the supply-demand shows obvious imbalance. Based on this, we can implement replenishment or transfer of bike-sharing at different time and space according to prediction results, and further realize the scheduling and rebalancing of different regions in advance.

VII. CONCLUSION AND FUTURE WORK

In this study, we rendered the supply-demand forecasting of BSS into a spatial-temporal prediction problem, and proposed a multi-block hybrid model that captured spatial-temporal characteristics of multi-source data. Specifically, we integrated CNN and GRU-Net into the structure to express the effect of exogenous variables on spatial and temporal respectively, and aggregated ConvGRU-Net to learn the spatial-temporal dependencies on the usage of bike-sharing. To evaluate the effectiveness of MBH model, we compared the proposed model with ten baselines including RNN, LSTM, GRU, ConvLSTM, ConvGRU, XGBoost, ST-ResNet, DMVST-Net and two variants, based on four datasets divided by 15, 30, 45 and 60 min. The comparison results show that 30 min is the best time interval to realize the supply-demand forecast of bike-sharing. The proposed model can achieve more efficient and accurate prediction than other benchmark models in terms of MAPE, RMSE and MAE. We further investigated the effect of spatial and temporal blocks modeling. The evidence indicates that the modeling of exogenous variables is essential and meteorological factors have a greater impact on bike-sharing usage forecast. Moreover, the application of prediction results demonstrates that MBH model can be used to forecast supply-demand gap, which provides useful information for rebalancing in BSS.

In the future, we will focus on the applicability of the proposed model to other spatiotemporal traffic prediction tasks. In addition, more data source such as mobile phone signaling data and social network data could be added into our model for better depicting residents' trip patterns and achieving higher prediction accuracy.

ACKNOWLEDGMENT

The authors would like to thank all reviewers for their valuable comments and helpful suggestions.

REFERENCES

[1] Z. Cheng, M.-Y. Chow, D. Jung, and J. Jeon, "A big data based deep learning approach for vehicle speed prediction," in *Proc. IEEE 26th Int. Symp. Ind. Electron. (ISIE)*, Jun. 2017, pp. 389–394.

[2] Y. Hou and P. Edara, "Network scale travel time prediction using deep learning," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 2672, no. 45, pp. 115–123, Dec. 2018, doi: 10.1177/0361198118776139.

[3] Z. Ma, H. Yu, W. Chen, and J. Guo, "Short utterance based speech language identification in intelligent vehicles with time-scale modifications and deep bottleneck features," *IEEE Trans. Veh. Technol.*, vol. 68, no. 1, pp. 121–128, Jan. 2019.

[4] Y. Chen, L. Shu, and L. Wang, "Poster abstract: Traffic flow prediction with big data: A deep learning based time series model," in *Proc. IEEE Conf. Comput. Commun. Workshops*, May 2017, pp. 1010–1011, doi: 10.1109/INFCOMW.2017.8116535.

[5] L. Chen and J. Jakubowicz, "Inferring bike trip patterns from bike sharing system open data," in *Proc. IEEE Int. Conf. Big Data*, Oct. 2015, pp. 2898–2900, doi: 10.1109/BigData.2015.7364115.

[6] Y. Zheng, L. Capra, O. Wolfson, and H. Yang, "Urban computing: Concepts, methodologies, and applications," *ACM Trans. Intell. Syst. Technol.*, vol. 5, no. 3, pp. 1–55, Oct. 2014, doi: 10.1145/2629592.

[7] C. L. Philip Chen and C.-Y. Zhang, "Data-intensive applications, challenges, techniques and technologies: A survey on big data," *Inf. Sci.*, vol. 275, pp. 314–347, Aug. 2014.

[8] K. Gebhart and R. B. Noland, "The impact of weather conditions on bikeshare trips in washington, DC," *Transportation*, vol. 41, no. 6, pp. 1205–1225, Nov. 2014, doi: 10.1007/s11116-014-9540-7.

[9] J. Bachand-Marleau, B. H. Y. Lee, and A. M. El-Geneidy, "Better understanding of factors influencing likelihood of using shared bicycle systems and frequency of use," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 2314, no. 1, pp. 66–71, Jan. 2012.

[10] H. I. Ashqar, M. Elhenawy, M. H. Almannaa, A. Ghanem, H. A. Rakha, and L. House, "Modeling bike availability in a bike-sharing system using machine learning," in *Proc. 5th IEEE Int. Conf. Models Technol. Intell. Transp. Syst. (MT-ITS)*, Jun. 2017, pp. 374–378, doi: 10.1109/MTITS.2017.8005700.

[11] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015, doi: 10.1038/nature14539.

[12] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 6, pp. 84–90, May 2017, doi: 10.1145/3065386.

[13] L. Lin, Z. He, and S. Peeta, "Predicting station-level hourly demand in a large-scale bike-sharing network: A graph convolutional neural network approach," *Transp. Res. C, Emerg. Technol.*, vol. 97, pp. 258–276, Dec. 2018.

[14] I. Sutskever, O. Vinyals, and Q. V. Le, "Sequence to sequence learning with neural networks," *Proc. Adv. Neural Inf. Process. Syst.*, Sep. 2014, pp. 3104–3112.

[15] X. Ma, Z. Tao, Y. Wang, H. Yu, and Y. Wang, "Long short-term memory neural network for traffic speed prediction using remote microwave sensor data," *Transp. Res. C, Emerg. Technol.*, vol. 54, pp. 187–197, May 2015.

[16] J. Ke, H. Zheng, H. Yang, and X. Chen, "Short-term forecasting of passenger demand under on-demand ride services: A spatio-temporal deep learning approach," *Transp. Res. C, Emerg. Technol.*, vol. 85, pp. 591–608, Dec. 2017.

[17] H. X. Yao, "Deep multi-view spatial-temporal network for taxi demand prediction," in *Proc. 32nd AAAI Conf. Artif. Intell.*, 2018, pp. 2588–2595.

[18] D. Deng, C. Shahabi, U. Demiryurek, L. Zhu, R. Yu, and Y. Liu, "Latent space model for road networks to predict time-varying traffic," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Aug. 2016, pp. 1525–1534, doi: 10.1145/2939672.2939860.

[19] Y. Tong, Y. Chen, Z. Zhou, L. Chen, J. Wang, Q. Yang, J. Ye, and W. Lv, "The simpler the better: A unified approach to predicting original taxi demands based on large-scale online platforms," in *Proc. 23rd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2017, pp. 1653–1662, doi: 10.1145/3097983.3098018.
[20] H. I. Ashqar, M. Elhenawy, and H. A. Rakha, "Modeling bike counts in a bike-sharing system considering the effect of weather conditions," *Case Stud. Transp. Policy*, vol. 7, no. 2, pp. 261–268, Jun. 2019.
[21] F. Wu, H. Wang, and Z. Li, "Interpreting traffic dynamics using ubiquitous urban data," in *Proc. 24th ACM SIGSPATIAL Int. Conf. Adv. Geographic Inf. Syst.*, 2016, pp. 1–4.
[22] B. Pan, U. Demiryurek, and C. Shahabi, "Utilizing real-world transportation data for accurate traffic prediction," in *Proc. IEEE 12th Int. Conf. Data Mining*, Dec. 2012, pp. 595–604.
[23] R. Beecham and J. Wood, "Exploring gendered cycling behaviours within a large-scale behavioural data-set," *Transp. Planning Technol.*, vol. 37, no. 1, pp. 83–97, Jan. 2014, doi: 10.1080/03081060.2013.844903.
[24] D. Scherer, A. Müller, and S. Behnke, "Evaluation of pooling operations in convolutional architectures for object recognition," in *Proc. 20th Int. Conf. Artif. Neural Netw. (ICANN)*, pp. 92–101, 2010.
[25] X. Ma, Z. Dai, Z. He, J. Ma, Y. Wang, and Y. Wang, "Learning traffic as images: A deep convolutional neural network for large-scale transportation network speed prediction," *Sensors*, vol. 17, no. 4, p. 818, Apr. 2017, doi: 10.3390/s17040818.
[26] J. Zhang, Y. Zheng, D. Qi, R. Li, X. Yi, and T. Li, "Predicting citywide crowd flows using deep spatio-temporal residual networks," *Artif. Intell.*, vol. 259, pp. 147–166, Jun. 2018.
[27] K. Zhang, Z. Liu, and L. Zheng, "Short-term prediction of passenger demand in multi-zone level: Temporal convolutional neural network with multi-task learning," *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 4, pp. 1480–1490, Apr. 2020, doi: 10.1109/TITS.2019.2909571.
[28] K. Cho, B. van Merrienboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, "Learning phrase representations using RNN encoder–decoder for statistical machine translation," in *Proc. Conf. Empirical Methods Natural Lang. Process. (EMNLP)*, 2014, pp. 1724–1734, doi: 10.3115/v1/D14-1179.
[29] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997, doi: 10.1162/neco.1997.9.8.1735.
[30] C. Xu, J. Ji, and P. Liu, "The station-free sharing bike demand forecasting with a deep learning approach and large-scale datasets," *Transp. Res. C, Emerg. Technol.*, vol. 95, pp. 47–60, Oct. 2018.
[31] Z. Yang, J. Chen, J. Hu, Y. Shu, and P. Cheng, "Mobility modeling and data-driven closed-loop prediction in bike-sharing systems," *IEEE Trans. Intell. Transp. Syst.*, vol. 20, no. 12, pp. 4488–4499, Dec. 2019, doi: 10.1109/TITS.2018.2886456.
[32] X. Shi, Z. Gao, L. Lausen, H. Wang, D.-Y. Yeung, W.-k. Wong, and W.-c. Woo, "Deep learning for precipitation nowcasting: A benchmark and a new model," 2017, *arXiv:1706.03458*. [Online]. Available: http://arxiv.org/abs/1706.03458
[33] K. P. Zhang, L. Zheng, Z. J. Liu, and J. Ning, "A deep learning based multitask model for network-wide traffic speed predication," *Neurocomputing* Apr. 2019, doi: 10.1016/j.neucom.2018.10.09.
[34] D. Wang, W. Cao, J. Li, and J. Ye, "DeepSD: Supply-demand prediction for online car-hailing services using deep neural networks," in *Proc. IEEE 33rd Int. Conf. Data Eng. (ICDE)*, Apr. 2017, pp. 243–254, doi: 10.1109/ICDE.2017.83.
[35] K. Niu, H. Zhang, T. Zhou, C. Cheng, and C. Wang, "A novel spatio-temporal model for city-scale traffic speed prediction," *IEEE Access*, vol. 7, pp. 30050–30057, 2019, doi: 10.1109/ACCESS.2019.2902185.
[36] S. Zhang, Z. Kang, Z. Zhang, C. Lin, C. Wang, and J. Li, "A hybrid model for forecasting traffic flow: Using layerwise structure and Markov transition matrix," *IEEE Access*, vol. 7, pp. 26002–26012, 2019, doi: 10.1109/ACCESS.2019.2901118.
[37] J. Bao, P. Liu, and S. V. Ukkusuri, "A spatiotemporal deep learning approach for citywide short-term crash risk prediction with multi-source data," *Accident Anal. Prevention*, vol. 122, pp. 239–254, Jan. 2019.
[38] B. M. Williams and L. A. Hoel, "Modeling and forecasting vehicular traffic flow as a seasonal ARIMA process: Theoretical basis and empirical results," *J. Transp. Eng.*, vol. 129, no. 6, pp. 664–672, Nov. 2003, doi: 10.1061/(asce)0733-947x(2003)129:6(664).
[39] Y. Liu and H. Wu, "Prediction of road traffic congestion based on random forest," in *Proc. 10th Int. Symp. Comput. Intell. Design (ISCID)*, Dec. 2017, pp. 361–364, doi: 10.1109/ISCID.2017.216.
[40] C.-H. Wu, J.-M. Ho, and D. T. Lee, "Travel-time prediction with support vector regression," *IEEE Trans. Intell. Transp. Syst.*, vol. 5, no. 4, pp. 276–281, Dec. 2004, doi: 10.1109/TITS.2004.837813.
[41] G. N. Oliveira, J. L. Sotomayor, R. P. Torchelsen, C. T. Silva, and J. L. D. Comba, "Visual analysis of bike-sharing systems," *Comput. Graph.*, vol. 60, pp. 119–129, Nov. 2016.
[42] A. Zhang, M. Li, Z. C. Lipton, and A. J. Smola. (2019). *Hans-On Learning Deep Learning*. [Online]. Available: https://zh.d2l.ai
[43] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2016, pp. 785–794, doi: 10.1145/2939672.2939785.
[44] F. Chollet. (2015). *Keras*. [Online]. Available: https://github.com/fchollet/keras
[45] M. Abadi *et al.*, "TensorFlow: Large-scale machine learning on heterogeneous distributed systems," 2016, *arXiv:1603.04467*. [Online]. Available: http://arxiv.org/abs/1603.04467

MIAO XU received the B.E. degree from the Shandong University of Technology, Zibo, China, in 2017. She is currently pursuing the Ph.D. degree with the School of Transportation, Jilin University. Her research interests include intelligent transportation systems, big data visualization analysis, and deep learning.

HONGFEI LIU received the B.E. degree in machinery design and manufacture from the Nanjing University of Science and Technology, Jiangsu, China, in 1996, and the M.A.Eng. and Ph.D. degrees in vehicle operating engineering from Jilin University, in 2002 and 2005, respectively. From 1996 to 1999, he was an Engineer in aeronautical facility research and development. From 2008 to 2009, he was a Visiting Scholar with the Chiba Institute of Technology, Japan. Since 2006, he has been a Teacher with Jilin University, where he is currently a Professor. He has authored four books, more than 30 articles, and more than ten inventions. He received the First Prize in the area of scientific and technological progress in Changchun. His research interests include traffic safety research and application, vehicle-road cooperative control, and vehicle reliability.

HONGBO YANG received the B.E. degree in traffic engineering from Shanghai Maritime University, Shanghai, China, in 2016. He is currently pursuing the master's degree with the School of Transportation, Jilin University. His research interests include machine learning, traffic signal control, and vehicle-road cooperative control in intelligent transportation systems.