Hindawi
Journal of Advanced Transportation
Volume 2022, Article ID 7690309, 11 pages
https://doi.org/10.1155/2022/7690309

Research Article

Examine the Prediction Error of Ride-Hailing Travel Demands with Various Ignored Sparse Demand Effects

Zhiju Chen,[1] **Kai Liu**,[1] **and Tao Feng**[2]

[1]*School of Transportation and Logistics, Dalian University of Technology, Dalian 116024, China*
[2]*Urban and Data Science Lab, Graduate School of Advanced Science and Engineering, Hiroshima University, Higashi-Hiroshima 739-8511, Japan*

Correspondence should be addressed to Kai Liu; liukai@dlut.edu.cn

Received 14 February 2022; Accepted 28 March 2022; Published 12 April 2022

Academic Editor: Hongtai Yang

The accurate short-term travel demand predictions of ride-hailing orders can promote the optimal dispatching of vehicles in space and time, which is the crucial issue to achieve sustainable development of such dynamic demand-responsive service. The sparse demands are always ignored in the previous models, and the uncertainties in the spatiotemporal distribution of the predictions induced by setting subjective thresholds are rarely explored. This paper attempts to fill this gap and examine the spatiotemporal sparsity effect on ride-hailing travel demand prediction by using Didi Chuxing order data recorded in Chengdu, China. To obtain the spatiotemporal characteristics of the travel demand, three hexagon-based deep learning models (H-CNN-LSTM, H-CNN-GRU, and H-ConvLSTM) are compared by setting various threshold values. The results show that the H-ConvLSTM model has better prediction performance than the others due to its ability to simultaneously capture spatiotemporal features, especially in areas with a high proportion of sparse demands. We found that increasing the minimum demand threshold to delete more sparse data improves the overall prediction accuracy to a certain extent, but the spatiotemporal coverage of the data is also significantly reduced. Results of this study could guide traffic operations in providing better travel services for different regions.

1. Introduction

Ride-hailing services achieve shared mobility between travel demand and idle supply via Internet matching. By effectively improving vehicle utilization, this energy-efficient mode of transportation reduces fuel consumption, traffic congestion, and vehicle emissions, which is conducive to urban sustainability [1–3]. Benefiting from the rapid development of information and communication technologies, the service platform has collected a large number of order data, which also provides strong support for the analysis of travel behavior, traffic management, and energy consumption [4, 5]. Compared with traditional taxi services, online booking effectively integrates the time and locational information between passengers and vehicles, which reduces the waiting time and improves the overall efficiency [6]. To meet the diversified and customized travel demand of residents in different urban areas, a variety of ride-hailing services have been constantly updated and evolved, which effectively alleviates the problems in traditional transport services, for example, taxi-hailing difficulty in rush hour and low service levels [7, 8]. However, due to the high heterogeneity of individuals' travel requests in different times and spaces [9–11], the problem of mismatching between drivers and passengers is still common; for example, cruising drivers spend much time finding passengers, and travel requests cannot be responded on time during peak hours. Thus, accurate short-term prediction of ride-hailing travel demand is needed and timely to solve the mismatching problems.

A large number of existing studies have explored the temporal and spatial patterns of travel demand in ride-hailing [12–15]. They usually focus on central urban areas where demand is high and omit or ignore the areas with sparse demands by setting a threshold of the minimum demand. Because the effects of the areas with sparse travel demand are relatively small, the treatment of the small travel requests does not influence the overall prediction performance of the models. Although private car ownership is

abundant in cities [16], it possibly leads to a social equity issue in practice as passengers living in rural areas or less dense areas will have to face a lower level of services anyhow. Moreover, from a methodological point of view, the selection of the minimum demand threshold based on experience lacks a unified standard. More importantly, an improper setting of the minimum demand threshold may unavoidably increase the uncertainty of demand distribution and spatiotemporal sparsity of travel demand, which misleads the demand prediction and vehicle dispatch. A proper method that considers the sparse travel demand is needed. In particular, the effects of incorporating sparse travel requests into the overall demand prediction, which has been overlooked for a long time in the existing literature, need to be further investigated.

Therefore, in this paper, we attempt to assess the effects of the sparsity difference of ride-hailing travel requests in different urban areas on travel demand prediction. We comparatively utilize three deep learning methods that have been proposed in the literature to verify the influence of the choice of minimum demand threshold on the accuracy of short-term travel demand prediction. To overcome the issues of square-based spatial partition, that is, two different kinds of adjacent neighbors between each square, side-connected and corner-connected, hexagon partition, which is considered to have a symmetric and equivalent distribution, is adopted. Specifically, three hybrid models that combine a convolutional neural network and long short-term memory (H-CNN-LSTM), convolutional neural network and gate recurrent unit (H-CNN-GRU), and hexagon partition and convolutional LSTM (H-ConvLSTM) are applied. The results of the models are evaluated using the big data from Didi Chuxing GAIA Initiative, an open data project of Didi company [17].

The rest of this paper is organized as follows. The "Related Works" section briefly reviews the existing literature on traffic demand prediction with an emphasis on the use of deep learning methods. The method adopted in this paper will be presented in the "Methodology" section. The "Experimental Result" section will introduce the data and present the experimental implementation details, followed by the conclusions and outline directions for further research in the "Conclusions" section.

2. Related Works

How to effectively mine the potential spatiotemporal characteristics of mobility patterns and accurately predict travel demand using ride-hailing data has been increasingly concerned in the domains of demand-responsive mobility. In this section, we mainly discuss the related work of travel demand prediction and sparse demand processing.

Early researches on traffic prediction are mainly based on time series models, such as the autoregressive integrated moving average (ARIMIA) model and its integration with other models [18–22]. Pavlyuk [23] compares ARIMA with different vector autoregressive models to discuss the influence of temporal aggregation on the spatiotemporal prediction accuracy of traffic flow. Machine learning models provide researches with more options. Zhu et al. [24] utilize a linear conditional Gaussian (LCG) Bayesian network (BN) model to consider both continuous and discrete variables for short-term traffic flow prediction. The results indicate that the prediction accuracy increases significantly when both spatial data and speed data are included. Lu and Zhou [25] propose a short-term highway traffic state prediction method based on a Kalman filtering model, which highlights the advantage of combining a polynomial trend model and historical patterns. Li et al. [26] combine the complementary advantages of wavelets analysis and least squares support vector machine (LS-SVM) models to predict short-term travel demand. Compared with the other models, the proposed model not only has better prediction performance but also is capable of capturing the nonstationary characteristics of short-term traffic dynamic systems. Phiboonbanakit and Horanont [27] compare the prediction results of standard random forest (RF), decision tree (DT), and gradient-boosted regression tree (GBDT) with actual operational data to reveal the mobility patterns of taxi trajectories. Liu et al. [28] propose a combined model of the random forest model (RFM) and ridge regression model (RRM) and take environmental and meteorological factors into account to predict the taxi demand in hotspots. The results indicate that the prediction effect of the combined model is better than those of RFM and RRM. Li et al. [29] develop a Markov-based time series model (MTSM) framework to predict traffic network conditions by integrating archived and real-time data under various external conditions, including weather, work zones, incidents, and special events. Sharma et al. [30] present an artificial neural network- (ANN-) based short-term traffic volume prediction model for two-lane undivided highways with mixed traffic conditions in India. These models are fit for different scenarios. However, with the explosive growth of the scale of data and the computational complexity between different dimensional features in the big data scenario, these methods will struggle to capture complex temporal and spatial correlations [31].

Deep learning has been widely used in traffic prediction [32, 33] in recent years due to its successful performance in big data processing and computational vision. CNN has a strong ability to capture local trend features and scale-invariant features when the nearby data points typically have a strong relationship with each other [34]. Zhang et al. [35] partition a city into a grid map based on the longitude and latitude and apply CNN to predict the travel flow in real time. Fedorov et al. [36] employ the state-of-the-art Faster R-CNN two-stage detector together with an SORT tracker to address the problem of traffic flow estimation with data from a video surveillance camera. By setting gated recurrent learning units inside the network, LSTM has the ability to capture the time characteristics of long and short states and has been broadly used in time series prediction [37]. Tang et al. [38] propose a Genetic Algorithm with Attention-based LSTM and combine with spatiotemporal correlation analysis to predict urban road traffic volume. Zhao et al. [39] further improve the computational efficiency of the LSTM model by simplifying the gating structure using GRU. To overcome the shortcomings of a single technology, the fusion of

multiple intelligent prediction methods is also the focus of recent research. Deep neural network models combining CNN with LSTM have been widely used in different disciplines, including traffic speed prediction and travel demand prediction [40–43].

In the process of traffic demand forecasting, the road network is often partitioned using squares. In recent years, hexagon partition has become increasingly popular in different domains. Hexagon partition has many advantages over square and other forms of partitioning. For instance, in the case of image processing, hexagon partition has higher computational efficiency, better robustness, and more accurate image alignment [44–46]. Because of the better approximation of the shape to a circular, a hexagon per unit area has a lower perimeter than a square. This improvement in reducing bias produced by edge effects allows hexagon partition to better aggregate travel demands with similar travel characteristics. Ke et al. [14] propose a hexagon-based convolutional neural network to predict the short-term supply-demand gap in ridesharing services. In addition, each square is connected to adjacent neighbors in two ways, by edges and angles, respectively. However, hexagon partition, which is symmetrically equivalent to its six adjacent neighbors [47], can better characterize the connectivity in the hierarchical network topology and become a popular grid division method for optimal scheduling of ride-hailing [48, 49].

Furthermore, as a common data processing method, setting the minimum demand threshold has been applied in many studies where the central urban areas with intensive travel demand are often the research focus. For example, Yao et al. [15] filter out the samples with a demand value of less than 10. Ke et al. [14] investigate the grids with daily requesting orders larger than 100. Huang et al. [50] empirically limit their research to the areas within the fourth Ring Road of Chengdu. Although the deletion of sparse demand is statistically indisputable, theoretically, it destroys the topological reliability of subsequent spatiotemporal correlation analysis. In addition, the selection of these threshold values has brought great spatiotemporal uncertainty to travel demand prediction. These issues, however, have not yet been addressed in the existing literature.

Therefore, the current paper contributes to the existing literature by systematically assessing the effects of the sparsity of ride-hailing travel requests in different urban areas on travel demand prediction. Our ultimate goal is to increase the reliability of travel demand prediction by avoiding the excessive omission of sparse data (e.g., in rural areas). Due to the complex nature underlying big data, various deep learning methods are compared based on hexagon partitions. To the best of our knowledge, this is the first effort to quantitatively identify the effects of sparse data in travel demand forecasting by setting various threshold values. The findings are expected to draw attention in the transportation community to this ignored but important issue. The results of the comparative analysis among different deep learning algorithms will also offer additional insights into the performance of the various cutting-edge methods.

3. Methodology

To predict travel demand, we first divide a city into uniform hexagon partitions based on latitude and longitude coordinates and divide the whole day into uniform time intervals. Based on the partition in space and time, the ride-hailing orders at hexagon partition i and time interval t are aggregated as travel demand y_i^t.

The short-term travel demand prediction problem aims to predict the future time interval travel demand y_i^{t+1} for ride-hailing in a specific partition of the city using multiple historical local spatial data $\mathcal{Y}_i^{t-h+1}, \ldots, \mathcal{Y}_i^t$ collected until time interval t, as shown in Figure 1. $\mathcal{Y}_i^t$ is represented by three-layer local adjacent partitions centralized at hexagon partition i. Since the computational propagation of deep learning is based on matrix transformation, the three-layer local adjacent partitions travel demands $\mathcal{Y}_i^t$ are transformed into the matrix or tensor through parity coordinate transformation following previous studies [14].

3.1. H-CNN-LSTM Model. We define X_i as the transformation matrix, and the position of the hexagon partition in a matrix is indexed by its parity coordinates. The transformation matrix of the three-layer local adjacent partitions is established as follows:

$$X_i^t = \begin{bmatrix} 0 & 0 & y_{i_9}^t & 0 & y_{i_{10}}^t & 0 & y_{i_{11}}^t & 0 & 0 \\ 0 & y_{i_8}^t & 0 & y_{i_2}^t & 0 & y_{i_3}^t & 0 & y_{i_{12}}^t & 0 \\ y_{i_7}^t & 0 & y_{i_1}^t & 0 & y_{i_0}^t & 0 & y_{i_4}^t & 0 & y_{i_{13}}^t \\ 0 & y_{i_{18}}^t & 0 & y_{i_6}^t & 0 & y_{i_5}^t & 0 & y_{i_{14}}^t & 0 \\ 0 & 0 & y_{i_{17}}^t & 0 & y_{i_{16}}^t & 0 & y_{i_{15}}^t & 0 & 0 \end{bmatrix}, \quad (1)$$

where $y_{i_0}^t = y_i^t$ denotes the travel demand for the central hexagon partition of $\mathcal{Y}_i^t$.

Figure 2 shows that the input of the CNN model is the image reflecting the spatial characteristics of travel demand; that is, the input X_i^t for hexagon partition i and time interval t is a 5×9 matrix (as shown in equation (1)) transformed by the three-layer local adjacent partitions centralized at hexagon partition i. The spatial feature of the demand is extracted by convolution layers, and the convolution operation between each layer is transformed as follows:

$$X_{i,k}^t = a\left(X_{i,k-1}^t * W_k^t + b_k^t\right), \quad (2)$$

where a is the ReLU function, which is selected as the activation function in this paper. W_k^t and b_k^t are parameter collections of the k^{th} layer. $*$ denotes the convolution operation. The last convolution layer $X_{i,K}^t$ is transformed into a dense vector that can be written as $\widehat{F}_i^t = \text{flatten}(X_{i,K}^t)$, where flatten denotes the concatenating procedure. Finally, we use a fully connected layer to reduce the dimension of the dense vector $\widehat{F}_i^t$ and learn the essential spatial feature for location i and time interval t. The output of this layer can be described as follows:

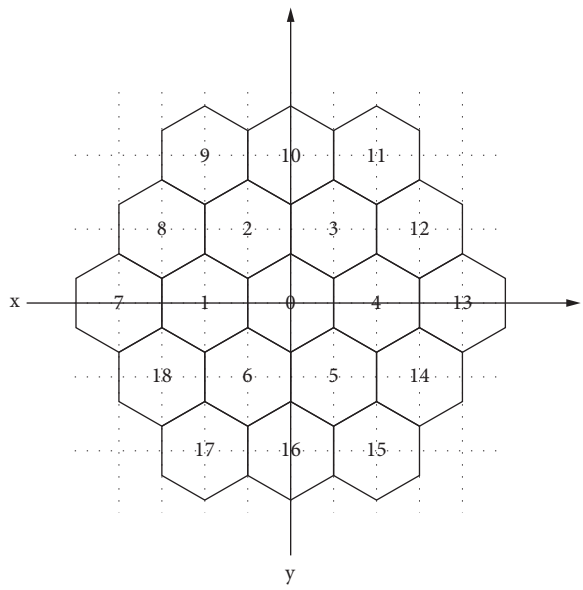


Figure 1: Three-layer local adjacent partitions [14]

$$F_i^t = a\left(W_{\text{fc}}^t \widehat{F}_i^t + b_{\text{fc}}^t\right), \tag{3}$$

where W_{fc}^t and b_{fc}^t are trainable parameters and a is the ReLU function.

LSTM is a special RNN model that is proposed to solve its gradient dispersion problem. In this model, LSTM is adopted to memorize the characteristics of the temporal dimension of travel demand. The key to LSTM is the cell state, which ensures the memory and circulation of information. It has the ability to remove or increase information to the cell state by carefully designing structures called gates, which consist of forget gate f_i^t, input gate i_i^t, and output gate o_i^t. The specific functional relationship of LSTM is shown as follows:

$$\begin{aligned}
f_i^t &= \sigma\left(W_f * \left(h_i^{t-1} + F_i^t\right) + b_f\right),\\
i_i^t &= \sigma\left(W_i * \left(h_i^{t-1} + F_i^t\right) + b_i\right),\\
\widetilde{C}_i^t &= \tanh\left(W_c * \left(h_i^{t-1} + F_i^t\right) + b_c\right),\\
C_i^t &= f_i^t \circ C_i^{t-1} + i_i^t \circ \widetilde{C}_i^t,\\
o_i^t &= \sigma\left(W_o * \left(h_i^{t-1} + F_i^t\right) + b_o\right),\\
h_i^t &= o_i^t \circ \tanh\left(C_i^t\right),
\end{aligned} \tag{4}$$

where $\circ$ denotes the Hadamard product, which calculates the elementwise products of two vectors, matrices, or tensors with the same dimensions, and σ and tanh denote the nonlinear activation function of a sigmoid function and a hyperbolic tangent function, respectively. W_f, W_i, W_c, W_o are trainable parameter matrices, and b_f, b_i, b_c, b_o are corresponding biases. The dense vector h_i^t represents the output of the model given the input F_i^t, h_i^{t-1}, which integrates the spatiotemporal features at location i and time interval t. The prediction demand $\widehat{y}_i^{t+1}$ for location i and time interval $t+1$ can be finally obtained by a fully connected network as follows:

$$\widehat{y}_i^{t+1} = \sigma\left(W_{\text{fu}} h_i^t + b_{\text{fu}}\right), \tag{5}$$

where W_{fu} and b_{fu} are trainable parameters and σ is the sigmoid function, which is selected as the activation function of fully connected layers.

3.2. H-CNN-GRU Model. In this section, GRU is applied instead of LSTM to capture the temporal characteristics. As a variant of LSTM, GRU eliminates the cell state and uses the hidden state for information transmission. Therefore, its simpler structure facilitates faster computation, and it has a better performance on limited training data. GRU combines the forget gate and input gate into a single update gate z_i^t, which controls how much information needs to be forgotten in the hidden layer h_i^{t-1} of the previous moment and how much information needs to be added from the candidate hidden layer $\widetilde{h}_i^t$ of the current moment. The candidate hidden layer $\widetilde{h}_i^t$ is similar to $\widetilde{C}_i^t$ in LSTM, which can be regarded as the new information at the current moment. The reset gate r_i^t is used to control how much of the previous memory needs to be retained. The specific functional relationship of GRU is shown as follows:

$$\begin{aligned}
z_i^t &= \sigma\left(W_z * \left(h_i^{t-1} + F_i^t\right) + b_z\right),\\
r_i^t &= \sigma\left(W_r * \left(h_i^{t-1} + F_i^t\right) + b_r\right),\\
\widetilde{h}_i^t &= \tanh\left(W_{\widetilde{h}} * \left(h_i^{t-1} + F_i^t\right) + b_{\widetilde{h}}\right),\\
h_i^t &= \left(1 - z_i^t\right) \circ h_i^{t-1} + z_i^t \circ \widetilde{h}_i^t,
\end{aligned} \tag{6}$$

where $\circ$ denotes Hadamard product and σ and tanh denote the activation function of a sigmoid function and a hyperbolic tangent function, respectively. W_z, W_r, $W_{\widetilde{h}}$ and b_z, b_r, $b_{\widetilde{h}}$ are trainable parameters. As presented in the previous section, a dense vector h_i^t can be output, and the prediction demand $\widehat{y}_i^{t+1}$ can be obtained by giving the input F_i^t, h_i^{t-1}. Figure 3 is a structural diagram of the H-CNN-GRU model.

3.3. H-ConvLSTM Model. Figure 4 shows the architecture of the H-ConvLSTM model. As an improved form of the LSTM model, ConvLSTM has convolutional structures in both the input-to-state and state-to-state transitions and has good performance in simultaneously capturing temporal and spatial features.

Similar to LSTM, ConvLSTM also consists of inputs $\mathcal{X}_i^{t-h+1}, \ldots, \mathcal{X}_i^t$, cell states $\mathcal{C}_i^{t-h+1}, \ldots, \mathcal{C}_i^t$, hidden states $\mathcal{H}_i^{t-h+1}, \ldots, \mathcal{H}_i^t$, and the other gates. These states are converted to 3D tensors ($X_i^t \longrightarrow \mathcal{X}_i^t$) suitable for the convolution operation, the last two dimensions of which are rows and columns of spatial information. The historical local spatial features of a certain partition can therefore be directly taken and transited in the model. The specific functional relationship of ConvLSTM is shown as follows:

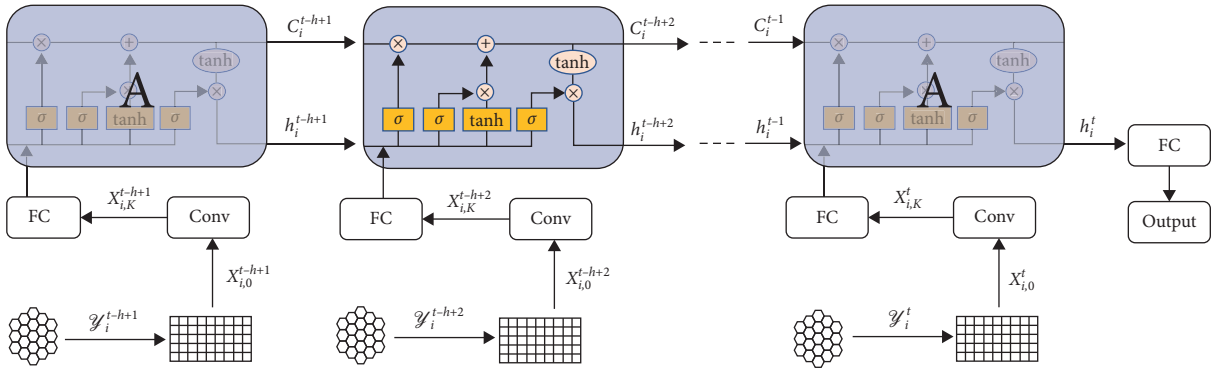


Figure 2: The architecture of H-CNN-LSTM.

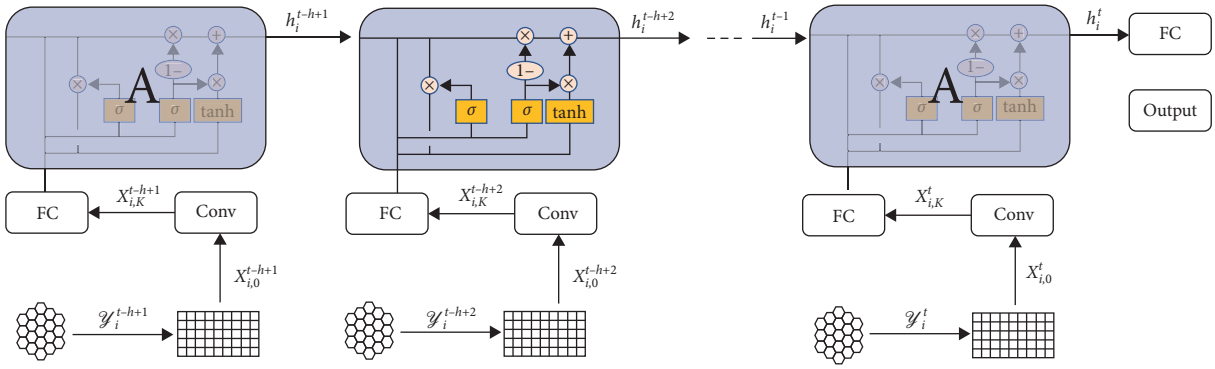


Figure 3: The architecture of the H-CNN-GRU.

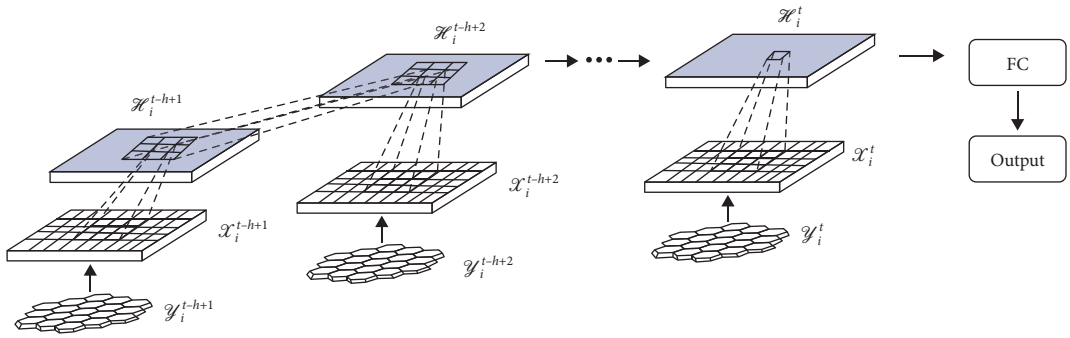


Figure 4: The architecture of H-ConvLSTM.

$$
\begin{aligned}
f_i^t &= \sigma\left(W_f * \left(\mathcal{H}_i^{t-1}, \mathcal{X}_i^t\right) + b_f\right),\\
i_i^t &= \sigma\left(W_i * \left(\mathcal{H}_i^{t-1}, \mathcal{X}_i^t\right) + b_i\right),\\
\tilde{C}_i^t &= \tanh\left(W_c * \left(\mathcal{H}_i^{t-1}, \mathcal{X}_i^t\right) + b_c\right),\\
C_i^t &= f_i^t \circ C_i^{t-1} + i_i^t \circ \tilde{C}_i^t,\\
o_i^t &= \sigma\left(W_o * \left(\mathcal{H}_i^{t-1}, \mathcal{X}_i^t\right) + b_o\right),\\
\mathcal{H}_i^t &= o_i^t \circ \tanh\left(C_i^t\right),
\end{aligned}
\tag{7}
$$

where $*$ denotes the convolution operator and the other parameter settings are consistent with those of LSTM. The output state $\mathcal{H}_i^{t-1}$ is finally taken into a fully connected network to obtain the prediction demand $\hat{y}_i^{t+1}$.

3.4. Loss Function. The symmetric mean absolute percentage error (SMAPE) between the estimated and real demand is sensitive to sparse demand, while the root mean square error (RMSE) can better assess the prediction performance of the central area by amplifying the impact of larger outliers [15]. Therefore, a composite loss function can be created as follows:

$$
L = \frac{1}{n}\sum_{i=1}^{n}\frac{\left|\hat{y}_i^{t+1} - y_i^{t+1}\right|}{\left|\hat{y}_i^{t+1}\right| + \left|y_i^{t+1}\right| + \varepsilon} + \lambda * \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\hat{y}_i^{t+1} - y_i^{t+1}\right)^2}, \tag{8}
$$

where ε is a very small value to avoid having a zero value in the denominator and λ is a hyperparameter.

4. Experimental Result

4.1. Dataset and Preprocessing. The dataset used in this study is the online ride-hailing order data provided by the Didi Gaia Plan platform of Chengdu in November 2016. Firstly, latitude and longitude are allocated into 35×46 hexagon partitions, and the side length of each partition is 800 m. Secondly, a time interval of 30 min is used to label each order data point based on its starting time. The hexagon partitions and time-labeled order data points are then intersected in Quantum Geographic Information System (QGIS) to obtain the spatial labels. Finally, the ride-hailing travel demands can be easily aggregated in different time intervals and spatial partitions.

Here, to evaluate the sparse effects of different models in the prediction of travel demand, two types of minimum demand thresholds are selected to screen samples with different degrees of sparsity. In the training and testing processes of the proposed deep learning models, a travel demand sample y_i^t needs to be expanded into the corresponding sample group $\left\{\mathcal{Y}_i^{t-h+1}, \ldots, \mathcal{Y}_i^t, y_i^{t+1}\right\}$, where $\mathcal{Y}_i^{t-h+1}, \ldots, \mathcal{Y}_i^t$ represents the input of the model and y_i^{t+1} represents the corresponding label. The first type is to set the minimum demand threshold as 1, 2, 4, and 8 for all hexagon partitions. If the demand y_i^t at the center of $\mathcal{Y}_i^t$ is less than the threshold, the sample group $\left\{\mathcal{Y}_i^{t-h+1}, \ldots, \mathcal{Y}_i^t, y_i^{t+1}\right\}$ is removed from the corresponding dataset. In order to ensure that the order coverage of different thresholds is roughly close to that of the first type for further comparative analysis, the second type selects all sample groups of the hexagon partition i with average daily demands above 25, 50, 100, and 200. For all datasets with different threshold settings, the first three weeks of data are used for training and the rest for testing. The statistical characteristics of the ride-hailing order data y_i^t under different types of threshold scales are shown in Table 1. Under the condition of similar order volume coverage (proportion of the order sample to the total demand after removing the sparse demand), the first threshold type covers more partitions in the spatial scale but is discontinuous in the temporal scale, while the second threshold type can predict all time intervals in the temporal scales and only cover limited partitions in the spatial scale.

4.2. Model Setup. The experiment is run on a server with an Intel(R) Xeon(R) Gold-5218 CPU @ 2.30 GHz, 128 GB RAM, and one GPU (NVIDIA Quadro RTX 5000). The proposed model is implemented based on Python 3.6.6 with TensorFlow and Keras. For the spatial view of the CNN of H-CNN-LSTM and H-CNN-GRU, there are 4 convolution layers that use 8, 16, 32, and 32 filters of size 3×3, respectively. In order to better capture the spatial characteristics of the input segment, virtual partitions with zero demand values are added as neighbors of the partitions on the border. The output dimension of the CNN, which is also the input size of LSTM and GRU, is set to 64. For the temporal view of LSTM and GRU, the time step is the previous 8 time intervals, and the hidden dimension is 64. In addition, the proposed H-ConvLSTM includes 4 ConvLSTM layers, which have 8, 16, 32, and 32 hidden states. The kernel size of each layer is 3×3. Batch normalization and dropout are used for training the model. The training epochs are set as 50 with a batch size of 64. Adam is used for optimization. SMAPE and RMSE are used to evaluate the predictive performance. In the loss function, the hyperparameter λ is set as 100 to ensure that RMSE and SMAPE's influence in the feedback calculation is in the same dimension.

4.3. Model Comparison. We compare our proposed models with the following standard models: (1) ARIMA: autoregressive integrated moving average model, which is a classical algorithm with good predictive performance for time series data. The difference order d is set to 1, with an autoregressive coefficient p and a moving average coefficient q for iterating the previous time intervals between 1 and 8; (2) H-ANN: the spatial and historical temporal features of the demands of a hexagon partition are spliced together as the input of the fully connected neural network, and the predicted demand value of the future time is output. The model includes 5 fully connected layers, which have 128, 64, 32, 16, and 8 hidden neurons, respectively; (3) H-CNN: transform the historical spatial features of the demands of a hexagon partition into images and stack them together as input. The previous 8 time intervals are represented by the numbers of channels in the input image. The H-CNN model includes 4 convolution layers, which have 8, 16, 32, and 32 hidden states. The kernel size of each layer is 3×3. Batch normalization and dropout are used to train the model; (4) H-LSTM: the spatial features of the demands of hexagon partitions are flattened into 1-dimensional demand vectors as the input of each historical time interval, and the spatiotemporal demands in the eight historical time intervals are used to predict the results of the next time interval. The hidden state is set to 64.

The prediction results of each model under different minimum demand threshold settings are shown in Tables 2 and 3. Our proposed hybrid deep learning models that combine spatiotemporal prediction with the advantages of the hexagon partition always achieve lower SMAPE and RMSE values, indicating better performance. Due to the simplified structure of the gate unit, H-CNN-GRU has a faster training efficiency than H-CNN-LSTM. For example, when the minimum demand threshold is set to 1, the training time of H-CNN-LSTM and H-CNN-GRU is 1.32 h and 1.21 h, respectively. Instead of the simple splicing of the spatial view model (CNN) and temporal view model (LSTM and GRU), ConvLSTM adopts convolutional structures in both the input-to-state and state-to-state transitions to integrate the advantages of these models, which reduces the loss of spatial topological relations of the data. This deeper fusion minimizes the prediction error of the H-ConvLSTM model at each threshold setting while increasing the difficulty of training (the training time at the corresponding threshold 1 is 5.52 h). However, considering that there is no order of magnitude differences of the training times among the three proposed hybrid deep learning models and that all training is conducted offline, the potential disadvantage of

Table 1: Order data characteristics under different minimum threshold settings.

Threshold type	Threshold scale	Order coverage	Partition number	Partition coverage	Average daily time interval	Average coverage of daily time interval
Type 1	1	1	1208	1	11.86	0.25
	2	0.98	1026	0.85	10.19	0.21
	4	0.95	550	0.46	13.17	0.27
	8	0.90	324	0.27	15.4	0.32
Type 2	25	0.98	345	0.29	48	1
	50	0.97	267	0.22	48	1
	100	0.95	205	0.17	48	1
	200	0.92	152	0.13	48	1

Table 2: Comparison of different models in minimum demand threshold type 1.

Model	Threshold 1		Threshold 2		Threshold 4		Threshold 8	
	SMAPE ($\times 10^{-2}$)	RMSE	SMAPE ($\times 10^{-2}$)	RMSE	SMAPE ($\times 10^{-2}$)	RMSE	SMAPE ($\times 10^{-2}$)	RMSE
ARIMA	22.67	4.12	19.92	5.32	16.23	7.26	13.15	8.79
H-ANN	21.85	4.01	18.76	5.13	15.76	7.22	12.94	8.55
H-CNN	21.4	3.52	18.94	4.69	14.28	6.24	12.2	7.92
H-LSTM	19.45	3.67	15.07	4.54	13.89	6.41	10.7	7.27
H-CNN-LSTM	17.38	3.4	14.52	4.14	12.57	5.78	10.06	6.9
H-CNN-GRU	17.29	3.34	14.49	4.24	12.41	5.58	10.13	6.88
H-ConvLSTM	16.34	3.03	13.75	3.97	11.35	5.34	9.612	6.66

Table 3: Comparison of different models in minimum demand threshold type 2.

Model	Threshold 25		Threshold 50		Threshold 100		Threshold 200	
	SMAPE ($\times 10^{-2}$)	RMSE	SMAPE ($\times 10^{-2}$)	RMSE	SMAPE ($\times 10^{-2}$)	RMSE	SMAPE ($\times 10^{-2}$)	RMSE
ARIMA	46.64	4.11	37.83	5.24	31.91	6.17	25.73	7.34
H-ANN	45.89	3.82	37.91	4.26	29.47	5.13	24.45	6.71
H-CNN	44.94	3.73	36.64	4.11	30.15	4.79	23.82	6.2
H-LSTM	41.61	3.55	37.29	4.07	27.08	4.95	22.75	5.94
H-CNN-LSTM	41.02	3.53	32.65	3.96	25.94	4.64	20.45	5.36
H-CNN-GRU	40.24	3.18	32.46	3.74	26.13	4.66	21.46	5.69
H-ConvLSTM	37.52	3.18	32.09	3.71	23.23	4.02	19.54	5.26

this increased training time for achieving superior demand prediction results seems to be tolerable.

The continuous reduction of the minimum demand threshold gradually enhances the sparsity of the sample data, and the corresponding SMAPE also significantly increases. In the case of similar order coverage, the SMAPE of setting the minimum demand threshold of type 1 is only approximately half of that of type 2. Even the prediction performance of type 1 with a minimum demand threshold of 1 (16.34) is better than either of type 2, which indicates that there are many 0-value demand prediction problems in type 2 that are difficult to predict accurately.

Figure 5 presents the spatial distribution of SMAPE of the H-ConvLSTM model under different thresholds. It can be found that, with the increase in the distance from the central areas, the travel demand sparsity of the suburbs gradually increases, and SMAPE in type 2 also gradually increases. However, the corresponding SMAPE prediction results in type 1 show a trend of increasing first and decreasing afterwards. The prediction results of each hexagon partition in type 1 are divided into four interval types according to the size of SMAPE, and the demand composition of each type is statistically analyzed, as shown in Figure 6. It can be found that the sparse demand ratio between 2 and 8 has a positive correlation with the SMAPE value. In addition, it can be seen from the 1-value demand prediction results that when the sample size proportion is high enough, the prediction results of sparse demand can also be improved. Therefore, the circular region with the highest SMAPE value in type 1 may be due to the large fluctuation of demand in these regions and the relatively high proportion of sparse demand, and a large number of 1-value predictions makes the outermost region SMAPE value decrease to a certain extent.

Figure 7 presents the spatial distribution of the SMAPE difference between H-ConvLSTM and the other two hybrid deep learning models. This shows that when the difference value in the figure is negative, indicating a smaller SMAPE value of H-ConvLSTM, the predicted results achieve better accuracy. In most cases, the prediction accuracy of H-ConvLSTM is better than that of the other two models, especially in areas where the sparse demand distribution is relatively dense.

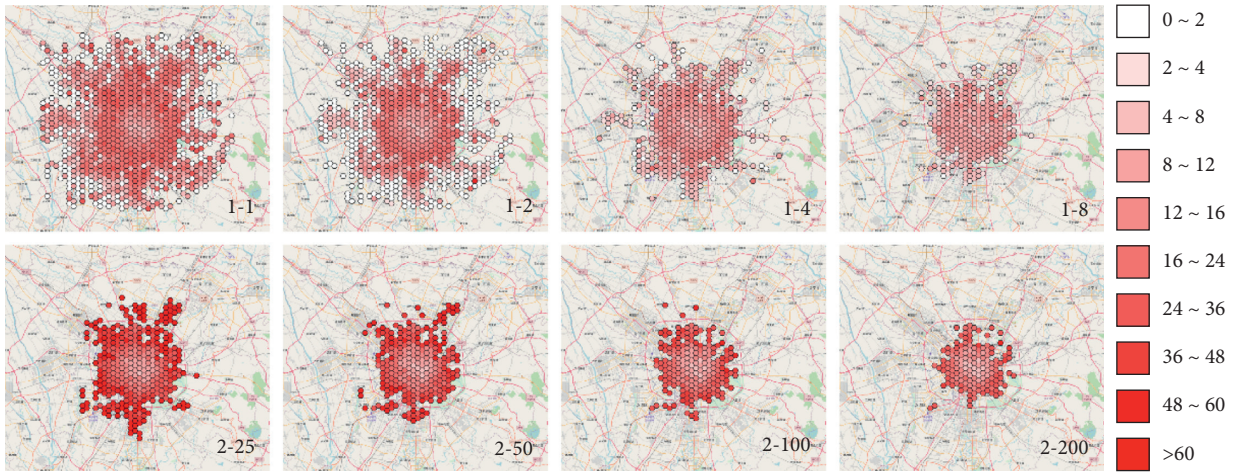


Figure 5: Spatial distribution of SMAPE of the H-ConvLSTM model under different thresholds.

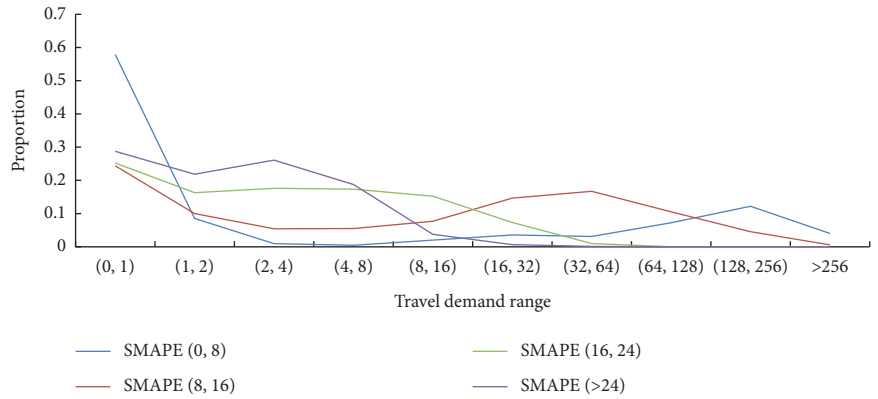
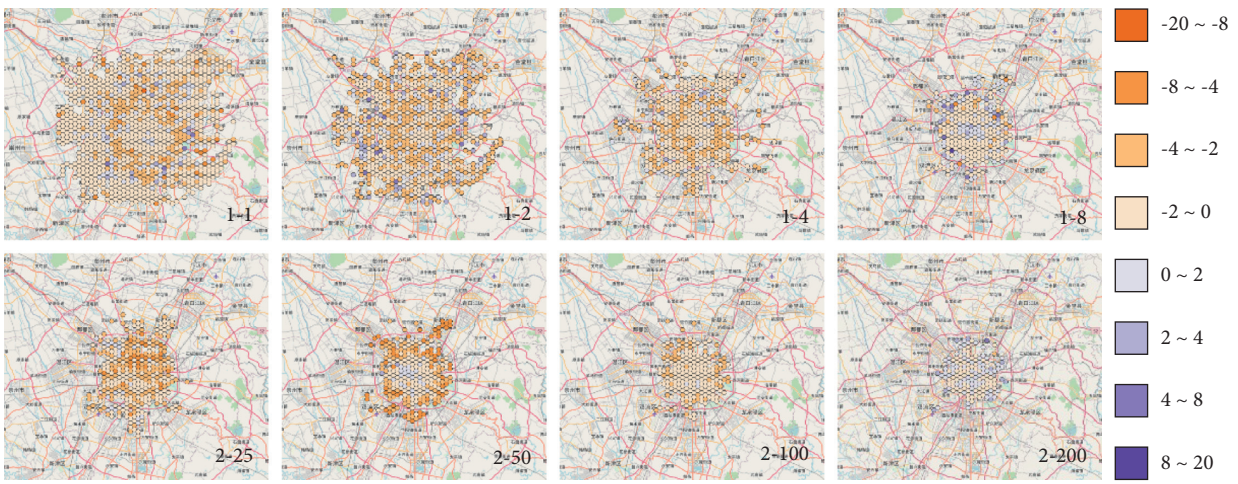


Figure 6: Composition of travel demand in different SMAPE hexagon partitions.



(a)

Figure 7: Continued.

(b)

Figure 7: Spatial distribution of SMAPE difference between H-ConvLSTM and other models: (a) H-CNN-LSTM; (b) H-CNN-GRU.

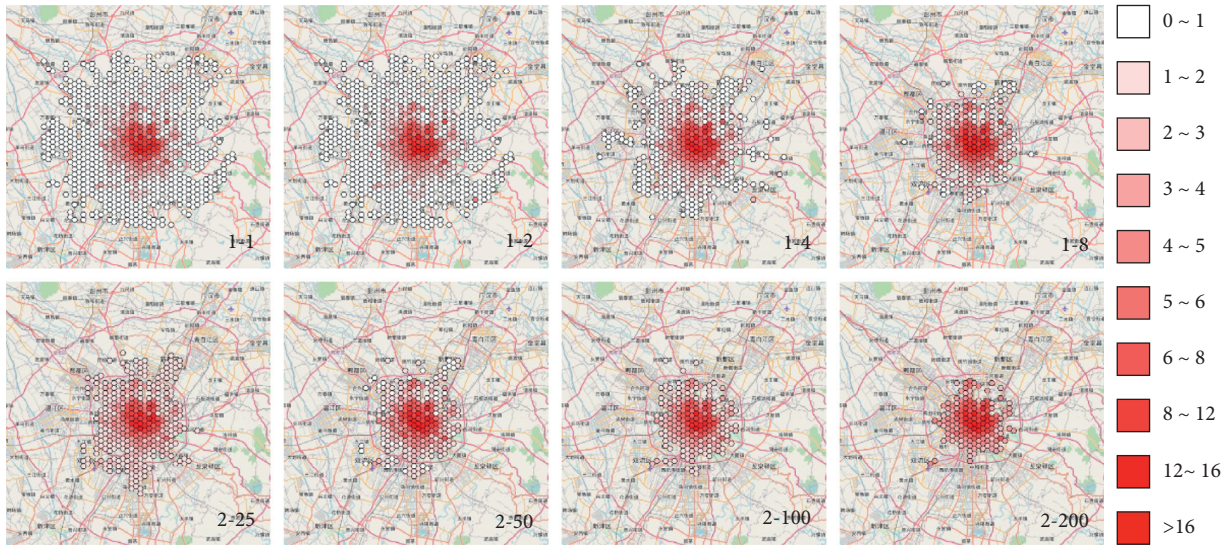


Figure 8: Spatial distribution of RMSE of the H-ConvLSTM model under different thresholds.

The continuous decrease in the minimum demand threshold makes the overall sample sparser and makes the overall RMSE smaller. However, since RMSE is greatly affected by the large demand samples, the RMSE value in the central area presents a stable trend, as shown in Figure 8.

5. Conclusions

To reduce the uncertainty in the mismatch between the supply and demand of online ride-hailing services, in this paper, we propose three hybrid deep learning methods based on hexagon partitions to analyze the effects of sparse data in travel demand prediction. The results are helpful for making policy recommendations to improve the operational efficiency and quality of ride-hailing services. The comparative results of the empirical study highlight that the H-ConvLSTM model has a better prediction performance than H-CNN-LSTM and H-CNN-GRU due to the excellent feature in capturing the temporal and spatial characteristics simultaneously, especially in the area with a high proportion of sparse demand. Because the model adopts convolutional structures in both the input-to-state and state-to-state transitions, it reduces the loss of spatial topological relations of the data.

The setting of different minimum demand thresholds changes the sparsity of the whole sample data, which has a significant impact on the prediction results of the models. Since the denominator of SMAPE in the sparse demand is generally small, the proportion of sparse demand between thresholds 2 and 8 has a significant positive correlation with the SMAPE value. In addition, it can be seen from the 1-value demand prediction results that when the sample size is high enough, the prediction results of sparse demand can also be improved. Therefore, with increasing distance from

the central area, the SMAPE in the case of the type 1 threshold values shows a trend of first increasing and then decreasing. The SMAPE in the type 1 threshold value is much lower than that of type 2, which indicates that there are many 0-value demand prediction problems in type 2 that are difficult to predict accurately. We also found that RMSE is largely affected by large demand values. Although the overall sparsity of the sample keeps decreasing, the central region shows a stable trend.

The current study explored the effectiveness of several advanced deep learning methods to address the effects of sparse data on travel demand prediction. However, more in-depth analysis of the spatiotemporal scale of Internet-based ride-hailing demand is needed; for example, more threshold values may be tested. The proposed methods may also be applied and validated in different cities, especially smaller and megacities where large differences in travel orders exist between central and rural areas and between different times of day. Furthermore, to improve the impact of sparse differences on travel demand prediction, multiscale spatiotemporal partitioning may be considered in the future because the resolution of partitioning may determine the resolution of sparse data. Nevertheless, we leave these considerations as our future work.

Data Availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

The authors would like to acknowledge the Gaia open data from DiDi Chuxing. This research was funded by the National Natural Science Foundation of China (Grants nos. 51378091 and 71871043).

References

[1] F. Ferrero, G. Perboli, M. Rosano, and A. Vesco, "Car-sharing services: an annotated review," *Sustainable Cities and Society*, vol. 37, pp. 501–518, 2018.

[2] M. Lin, C. Huang, and Z. Xu, "MULTIMOORA based MCDM model for site selection of car sharing station under picture fuzzy environment," *Sustainable Cities and Society*, vol. 53, Article ID 101873, 2020.

[3] S. Shaheen and A. Cohen, "Shared ride services in North America: definitions, impacts, and the future of pooling," *Transport Reviews*, vol. 39, pp. 427–442, 2019.

[4] F. Chen, Z. Yin, Y. Ye, and D. Sun, "Taxi hailing choice behavior and economic benefit analysis of emission reduction based on multi-mode travel big data," *Transport Policy*, vol. 97, pp. 73–84, 2020.

[5] D. Sun, K. Zhang, and S. Shen, "Analyzing spatiotemporal traffic line source emissions based on massive didi online car-hailing service data," *Transportation Research Part D: Transport and Environment*, vol. 62, pp. 699–714, 2018.

[6] H. Yang, G. Zhai, L. Yang, and K. Xie, "How does the suspension of ride-sourcing affect the transportation system and environment?" *Transportation Research Part D: Transport and Environment*, vol. 102, Article ID 103131, 2022.

[7] M. Stiglic, N. Agatz, M. Savelsbergh, and M. Gradisar, "The benefits of meeting points in ride-sharing systems," *Transportation Research Part B: Methodological*, vol. 82, pp. 36–53, 2015.

[8] H. Wang and H. Yang, "Ridesourcing systems: a framework and review," *Transportation Research Part B: Methodological*, vol. 129, pp. 122–155, 2019.

[9] J. Tang, F. Gao, C. Han, X. Cen, and Z. Li, "Uncovering the spatially heterogeneous effects of shared mobility on public transit and taxi," *Journal of Transport Geography*, vol. 95, Article ID 103134, 2021.

[10] J. Huo, H. Yang, C. Li, R. Zheng, L. Yang, and Y. Wen, "Influence of the built environment on E-scooter sharing ridership: a tale of five cities," *Journal of Transport Geography*, vol. 93, Article ID 103084, 2021.

[11] D. Sun and X. Ding, "Spatiotemporal evolution of ride-sourcing markets under the new restriction policy: A case study in Shanghai," *Transportation Research Part A: Policy and Practice*, vol. 130, pp. 227–239, 2019.

[12] Y. Cheng, X. Chen, X. Ding, and L. Zeng, "Optimizing location of car-sharing stations based on potential travel demand and present operation characteristics: the case of Chengdu," *Journal of Advanced Transportation*, vol. 2019, Article ID 7546303, 13 pages, 2019.

[13] Z. Wu and G. Lian, "A novel dynamically adjusted regressor chain for taxi demand prediction," in *Proceedings of the International Joint Conference on Neural Networks*, Glasgow, UK, July 2020.

[14] J. Ke, H. Yang, H. Zheng et al., "Hexagon-based convolutional neural network for supply-demand forecasting of ride-sourcing services," *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, pp. 4160–4173, 2019.

[15] H. Yao, F. Wu, J. Ke et al., "Deep multi-view spatial-temporal network for taxi demand prediction," in *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence*, Orleans Louisiana, USA, February 2018.

[16] H. Yang, G. Zhai, X. Liu, L. Yang, Y. Liu, and Q. Yuan, "Determinants of city-level private car ownership: effect of vehicle regulation policies and the relative price," *Transport Policy*, vol. 115, pp. 40–48, 2022.

[17] W. Li, Z. Pu, Y. Li, Jeff, and X. Ban, "Characterization of ridesplitting based on observed data: a case study of Chengdu, China," *Transportation Research Part C: Emerging Technologies*, vol. 100, pp. 330–353, 2019.

[18] M. S. Ahmed and A. R. Cook, "Analysis of freeway traffic time-series data by using box-jenkins techniques," *Transportation Research Record*, vol. 1–9, 1979.

[19] W. Min and L. Wynter, "Real-time road traffic prediction with spatio-temporal correlations," *Transportation Research Part C: Emerging Technologies*, vol. 19, pp. 606–616, 2011.

[20] X. Li, G. Pan, Z. Wu, G. Qi, S. Li, and D. Zhang, "Prediction of urban human mobility using large-scale taxi traces and its applications," *Frontiers of Computer Science in China*, vol. 6, pp. 111–121, 2012.

[21] A. Kaltenbrunner, R. Meza, J. Grivolla, J. Codina, and R. Banchs, "Urban cycles and mobility patterns: exploring and predicting trends in a bicycle-based public transport system," *Pervasive and Mobile Computing*, vol. 6, pp. 455–466, 2010.

[22] D. Billings and J. S. Yang, "Application of the ARIMA models to urban roadway travel time prediction-a case study," in

Proceedings of IEEE International Conference on Systems, Man and Cybernetics, pp. 2529–2534, Taipei, Taiwan, October 2006.

[23] D. Pavlyuk, "Temporal aggregation effects in spatiotemporal traffic modelling," *Sensors*, vol. 20, p. 6931, 2020.

[24] Z. Zhu, B. Peng, C. Xiong, and L. Zhang, "Short-term traffic flow prediction with linear conditional Gaussian Bayesian network," *Journal of Advanced Transportation*, vol. 50, pp. 1111–1123, 2016.

[25] C. C. Lu and X. Zhou, "Short-term highway traffic state prediction using structural state space models," *Journal of Intelligent Transportation Systems: Technology, Planning, and Operations*, vol. 18, pp. 309–322, 2014.

[26] Y. Li, J. Lu, L. Zhang, and Y. Zhao, "Taxi booking mobile app order demand prediction based on short-term traffic forecasting," *Transportation Research Record*, vol. 2634, pp. 57–68, 2017.

[27] T. Phiboonbanakit and T. Horanont, "Analyzing Bangkok city taxi ride: reforming fares for profit sustainability using big data driven model," *Journal of Big Data*, vol. 8, 2021.

[28] Z. Liu, H. Chen, Y. Li, and Q. Zhang, "Taxi demand prediction based on a combination forecasting model in hotspots," *Journal of Advanced Transportation*, vol. 2020, Article ID 1302586, 13 pages, 2020.

[29] T. Li, J. Ma, and C. Lee, "Markov-based time series modeling framework for traffic-network state prediction under various external conditions," *Journal of Transportation Engineering Part A System*, vol. 146, pp. 1–14, 2020.

[30] B. Sharma, S. Kumar, P. Tiwari, P. Yadav, and M. I. Nezhurina, "ANN based short-term traffic flow forecasting in undivided two lane highway," *Journal of Big Data*, vol. 5, 2018.

[31] H. Nguyen, L. M. Kieu, T. Wen, and C. Cai, "Deep learning methods in transportation domain: a review," *IET Intelligent Transport Systems*, vol. 12, pp. 998–1004, 2018.

[32] D. Jo, B. Yu, H. Jeon, and K. Sohn, "Image-to-image learning to predict traffic speeds by considering area-wide spatio-temporal dependencies," *IEEE Transactions on Vehicular Technology*, vol. 68, pp. 1188–1197, 2019.

[33] C. Yuan, X. Yu, D. Li, and Y. Xi, "Overall traffic mode prediction by VOMM approach and AR mining algorithm with large-scale data," *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, pp. 1508–1516, 2019.

[34] L. Alzubaidi, J. Zhang, A. J. Humaidi et al., "Review of deep learning: concepts, CNN architectures, challenges, applications, future directions," *Journal of Big Data*, vol. 8, 2021.

[35] J. Zhang, Y. Zheng, D. Qi, R. Li, and X. Yi, "DNN-based prediction model for spatio-temporal data," in *Proceedings of the 24th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, pp. 1–4, California, USA, October 2016.

[36] A. Fedorov, K. Nikolskaia, S. Ivanov, V. Shepelev, and A. Minbaleev, "Traffic flow estimation with data from a video surveillance camera," *Journal of Big Data*, vol. 6, 2019.

[37] R. L. Abduljabbar, H. Dia, and P. W. Tsai, "Unidirectional and bidirectional LSTM models for short-term traffic prediction," *Journal of Advanced Transportation*, vol. 2021, Article ID 5589075, 16 pages, 2021.

[38] J. Tang, J. Zeng, Y. Wang, H. Yuan, F. Liu, and H. Huang, "Traffic flow prediction on urban road network based on License Plate Recognition data: combining attention-LSTM with Genetic Algorithm," *Transp A Transportation Science*, vol. 17, pp. 1217–1243, 2021.

[39] J. Zhao, Y. Gao, Y. Qu, H. Yin, Y. Liu, and H. Sun, "Travel time prediction: based on gated recurrent unit method and data fusion," *IEEE Access*, vol. 6, pp. 70463–70472, 2018.

[40] 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," *Transportation Research Part C: Emerging Technologies*, vol. 85, pp. 591–608, 2017.

[41] G. Yang, Y. Wang, H. Yu, Y. Ren, and J. Xie, "Short-term traffic state prediction based on the spatiotemporal features of critical road sections," *Sensors*, vol. 18, Article ID 2287, 2018.

[42] H. Yu, Z. Wu, S. Wang, Y. Wang, and X. Ma, "Spatiotemporal recurrent convolutional networks for traffic prediction in transportation networks," *Sensors*, vol. 17, pp. 1–16, 2017.

[43] D. Wang, Y. Yang, and S. Ning, "DeepSTCL: a deep spatio-temporal ConvLSTM for travel demand prediction," in *Proceedings of the 2018 International Joint Conference on Neural Networks*, pp. 1–8, Rio de Janeiro, Brazil, July 2018.

[44] T. Shima, S. Sugimoto, and M. Okutomi, "Comparison of image alignment on hexagonal and square lattices," in *Proceedings of the International Conference on Image Processing*, Hong Kong, China, September 2010.

[45] D. Burdescu, M. Brezovan, E. Ganea, and L. Stanescu, "New algorithm for segmentation of images represented as hypergraph hexagonal-grid," in *Proceedings of the 5th Iberian Conference on Pattern Recognition and Image Analysis*, Lecture Notes in Computer Science, Las Palmas, Spain, June 2011.

[46] R. C. Staunton and N. Storey, "A Comparison between square and hexagonal sampling methods for pipeline image processing," *Optics, Illumination, and Image Sensing for Machine Vision IV*, vol. 1194, p. 142, 1990.

[47] C. P. D. Birch, N. Vuichard, and B. R. Werkman, "Modelling the effects of patch size on vegetation dynamics: bracken [Pteridium aquilitnum (L.) Kuhn] under grazing," *Annals of Botany*, vol. 85, pp. 63–76, 2000.

[48] X. Tang, Z. Qin, F. Zhang, Z. Wang, Z. Xu, and Y. Ma, "A deep value-network based approach for multi-driver order dispatching," in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, New York, NY, USA, July 2019.

[49] Y. Jiao, X. Tang, Z. Qin, S. Li, F. Zhang, and H. Zhu, "Real-world ride-hailing vehicle repositioning using deep reinforcement learning," *Transportation Research Part C: Emerging Technologies*, vol. 130, 2021.

[50] Z. Huang, G. Huang, Z. Chen, C. Wu, X. Ma, and H. Wang, "Multi-regional online car-hailing order quantity forecasting based on the convolutional neural network," *OR Informatie*, vol. 10, 2019.