Transportation Research Part C 140 (2022) 103709

Contents lists available at ScienceDirect

Transportation Research Part C

journal homepage: www.elsevier.com/locate/trc

H-ConvLSTM-based bagging learning approach for ride-hailing demand prediction considering imbalance problems and sparse uncertainty

Zhiju Chen [a], Kai Liu [a,*], Jiangbo Wang [a], Toshiyuki Yamamoto [b]

[a] School of Transportation and Logistics, Dalian University of Technology, Dalian 116024, China
[b] Institute of Materials and Systems for Sustainability, Nagoya University, Nagoya 464-8603, Japan

ARTICLE INFO

Keywords:
Ride-hailing demand prediction
Sparse uncertainty
Hexagonal convolutional long short-term memory (H-ConvLSTM)
Bagging learning

ABSTRACT

The problem of learning from imbalanced ride-hailing demand data with spatiotemporal heterogeneity and highly skewed demand distributions is a relatively new challenge. Current prediction methods usually filter out some spatiotemporal partitions with sparse demands by setting a minimum ride-hailing demand threshold, where the dataset is always assumed to be well balanced in terms of its spatiotemporal partitions, with equal misprediction costs. However, this widely used assumption results in large prediction biases. To achieve better prediction performance, we propose a bagging learning approach based on hexagonal convolutional long short-term memory (H-ConvLSTM), which combines three components. 1) By setting multiple minimum ride-hailing demand thresholds, several subdatasets with different majority ride-hailing demand prediction ranges are obtained. The H-ConvLSTM regression model is applied to each undersampled dataset to train multiple submodels with their respective biased ride-hailing demand prediction ranges. 2) The H-ConvLSTM classification model is trained on the total ride-hailing demand dataset to predict the potential demand range for a certain partition at a future time. 3) The submodel with the best performance with respect to the potential demand range is selected to predict the future demand for this partition. Experiments conducted on order data obtained from Didi Chuxing in Chengdu, China, are conducted. The results show that the proposed approach achieves significantly improved prediction performance relative to that of other models.

1. Introduction

Internet-based ride-hailing services, which connect drivers and passengers in real time, have attracted much interest as travel options for residents in recent years (Vazifeh et al., 2018, Xu et al., 2020). Compared to a traditional taxi service, with a ride-hailing service, passengers can book orders online in advance through a mobile app instead of standing on the side of the road and spending time waiting for a taxi to arrive; this improves the mobility of vehicles and the service level of travel (Alonso-Mora et al., 2017). With the collection and analysis of large amounts of user order data and vehicle trajectory data, ride-hailing services are constantly updating and evolving (Alisoltani et al., 2021), thereby becoming a disruptive force to the traditional transportation industry (Wang and Yang,

* Corresponding author.
E-mail addresses: chenzhiju@mail.dlut.edu.cn (Z. Chen), liukai@dlut.edu.cn (K. Liu), Jiangbo_Wang@dlut.edu.cn (J. Wang), yamamoto@civil.nagoya-u.ac.jp (T. Yamamoto).

https://doi.org/10.1016/j.trc.2022.103709
Received 1 November 2021; Received in revised form 5 April 2022; Accepted 2 May 2022
Available online 10 May 2022
0968-090X/

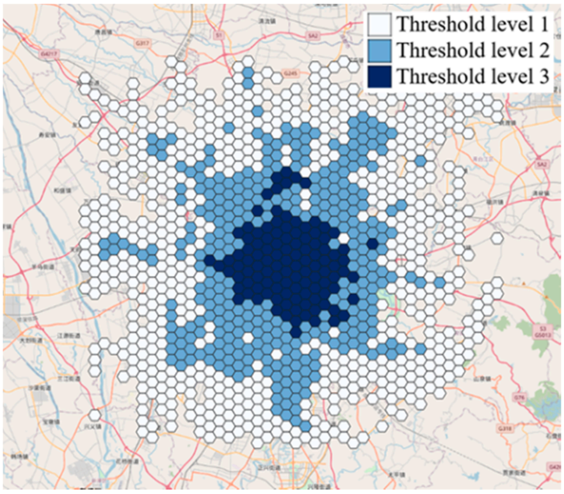
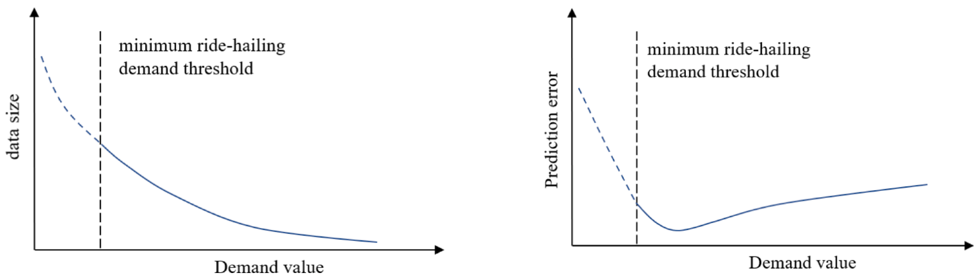


Fig. 1. Spatial coverages under different minimum ride-hailing demand thresholds.



(a) Data size distribution under different demand levels (b) Prediction errors induced with different demand levels

Fig. 2. The influence of the minimum threshold on ride-hailing demand prediction.

2019). Accurate short-term passenger demand prediction is the basis for improving the operating efficiency of internet-based ride-hailing platforms, which plays a crucial role in formulating regulation strategies and improving the balance between supply and demand.

Human travel behavior has a high degree of temporal and spatial regularity (González et al., 2008), and many contributions have by focusing on capturing the temporal and spatial correlations of ride-hailing demand (Ke et al., 2019). Such methods usually grid an urban space and predict the demands of future time intervals through historical spatiotemporal demands (Wu et al., 2018; Yu et al., 2017). In general, ride-hailing demand also exhibits complex spatiotemporal heterogeneity (Shen et al., 2020). In the temporal aspect, ride-hailing demand during rush hour is higher than that during the normal peak period, and demand during the day is higher than that at night. Spatially, with increasing distance from the city center, ride-hailing demand gradually becomes sparse. In a high demand prediction case focusing on a central urban area, sparse demand has little influence on the training of the utilized model, and it is therefore difficult to achieve good prediction performance. As the improvement of transport infrastructure often lags behind urban development and expansion, the sparse demand for suburban long-distance commuter travel also deserves fair attention. Therefore, a data imbalance problem is present in the observed ride-hailing demand. Different spatiotemporal scale divisions determined based on experience further aggravate the uncertainties of such a highly skewed ride-hailing demand distribution that are aggregated in various spatiotemporal granularities. To address the supply-demand imbalance issue, ride-sourcing platforms attempt to provide relocation guidance for idling drivers (Chen et al., 2020; Zhu et al., 2021). However, near-future spatiotemporal supply gap area prediction remains an unanswered question (Daganzo et al., 2020; Guo et al. 2021).

A gap remains in terms of predicting the highly uncertain demands of sparse areas, where the supply-demand imbalance is grievous and requires prescient dispatching in advance. Previous studies usually deleted spatiotemporal partitions with sparse demand from all recorded data by setting a minimum ride-hailing demand threshold, which helped to alleviate the problem of data imbalance. An increase in the level of the minimum ride-hailing demand threshold significantly reduces the spatial coverage of research and changes the sparsity of data, as shown in Fig. 1. However, the imbalance of the reduced data leads to worse demand prediction, as shown in Fig. 2. With the increase in demand per partition, the corresponding dataset size also decreases. For a given small threshold, adjacent ranges that are larger than the threshold tend to yield better prediction due to their higher data size distributions. The challenge is how to improve the prediction results for demand-sparse partitions with ranges smaller than the minimum threshold.

This paper takes a step toward closing this gap. Due to the use of different minimum ride-hailing demand threshold settings, the corresponding datasets have different right-adjacent optimal prediction ranges. A hexagonal convolutional long short-term memory (H-ConvLSTM)-based bagging learning approach is proposed to integrate the bias preferences of H-ConvLSTM models at different data sparsity levels. The results are helpful for providing suggestions regarding the optimal deployment of ride-hailing services, reducing driver operating costs, and improving the travel quality of residents. The main contributions of this paper are summarized as follows.

- We propose an H-ConvLSTM regression model to compare and analyze the ride-hailing demand prediction performances achieved under different minimum ride-hailing demand threshold settings.
- An H-ConvLSTM-based bagging learning approach is further proposed to integrate the bias prediction preferences of each H-ConvLSTM regression model trained at different data sparsity levels.
- An experimental analysis conducted on the order data obtained from Didi Chuxing in Chengdu city over one month shows that the proposed approach can achieve improved prediction performance on the total dataset.

The rest of this paper is organized as follows. Section 2 is a literature review of ride-hailing demand prediction and the data imbalance problem. Section 3 describes the main structural framework of the developed prediction models. Section 4 presents the experimental results, followed by the conclusions in Section 5.

2. Literature review

The use of historical travel records to predict future ride-hailing demand is helpful for assisting online ride-hailing platforms in carrying out dynamic operation strategies and optimizing the balance between supply and demand. In this section, we discuss traditional and existing travel prediction approaches, the advantages of hexagonal partitioning, and the related work that deals with sparse demand data.

2.1. *Travel demand prediction approaches*

The most common travel prediction method is a time series model, such as an autoregressive integrated moving average (ARIMA) model and its various improved versions (Kaltenbrunner et al., 2010; Min and Wynter, 2011). Machine learning models and statistical models such as neural network models (Zheng et al., 2006), Bayesian network models, Kalman filtering models, and least absolute shrinkage and selection operator (LASSO) models have also been proposed to solve various prediction problems related to travel demand. Jiang et al. (2014) integrated ensemble empirical mode decomposition (EEMD) and a gray support vector machine (GSVM) into a mixed-demand prediction model for high-speed railways. Ma et al. (2014) proposed an interactive multiple model-based pattern hybrid (IMMPH) approach to predict short-term passenger demand, and this approach maximizes the effective information by assembling the knowledge obtained from pattern models. Davis et al. (2016) proposed a multilayer clustering technique that utilizes the correlation between adjacent geographic hashes to reduce prediction errors. Zhu et al. (2019) integrated the joint probability distribution of traffic flows at nearby locations into a time series traffic speed prediction model. Although these models have achieved improved prediction performance through continuous improvement, they still struggle to capture complex temporal and spatial correlations.

The great advantages of deep learning in terms of computing power and characterizing big data enable its wide application to travel prediction (Jo et al., 2019; Yuan et al., 2019). By approximating the grid of an urban space into image pixels, a convolutional neural network (CNN) can effectively identify the spatial correlations among the grid data. Zhang et al. (2016) applied a CNN to a deep spatiotemporal prediction model to predict travel flows in real time. Both LSTM and gated recurrent units (GRUs) have good performance with respect to capturing complex time-sequential interactions. Therefore, combinations of these models seem to have better performance in dealing with complex temporal and spatial correlations (Chen et al., 2022). Yu et al. (2017) combined a CNN and LSTM to obtain spatial and temporal features for the prediction of traffic speed. Shi et al. (2015) applied ConvLSTM to address precipitation nowcasting. As an improved form of an LSTM model, ConvLSTM employs convolutional structures in both the input-to-state and state-to-state transitions to reduce the loss of spatiotemporal topology data. In the field of transportation, ConvLSTM has also been applied to solve prediction problems such as travel speed and ride-hailing demand and has achieved good prediction performance (Ke et al., 2017; Wang et al., 2018; Yang et al., 2018). However, these models, which are based on square partitions, are often difficult to directly apply to hexagonal networks.

2.2. *The advantages of hexagonal partitioning*

Compared with a square, a hexagon is closer to a circle, and its distribution is symmetric and equivalent (Birch et al., 2000). Therefore, travel demands with similar spatiotemporal characteristics are more easily aggregated, and the flows of vehicles between partitions are more accurately characterized. In addition, in a square partition space, the partition distance transformed from the same actual distance is much larger in the oblique direction than in the vertical and horizontal directions. The better isotropy of a hexagon partition enables it to better express the spatial proximity between partitions during the calculation process. Based on these advantages, hexagonal partitioning has been widely used in regional and urban science research. Shoman et al. (2019) performed a comparative analysis between hexagonal partitions, triangles, and squares and found that hexagonal partitions can better reduce the area errors of urban fabric. Csiszár et al. (2019) applied the hexagonal partition method to an evaluation of charging station

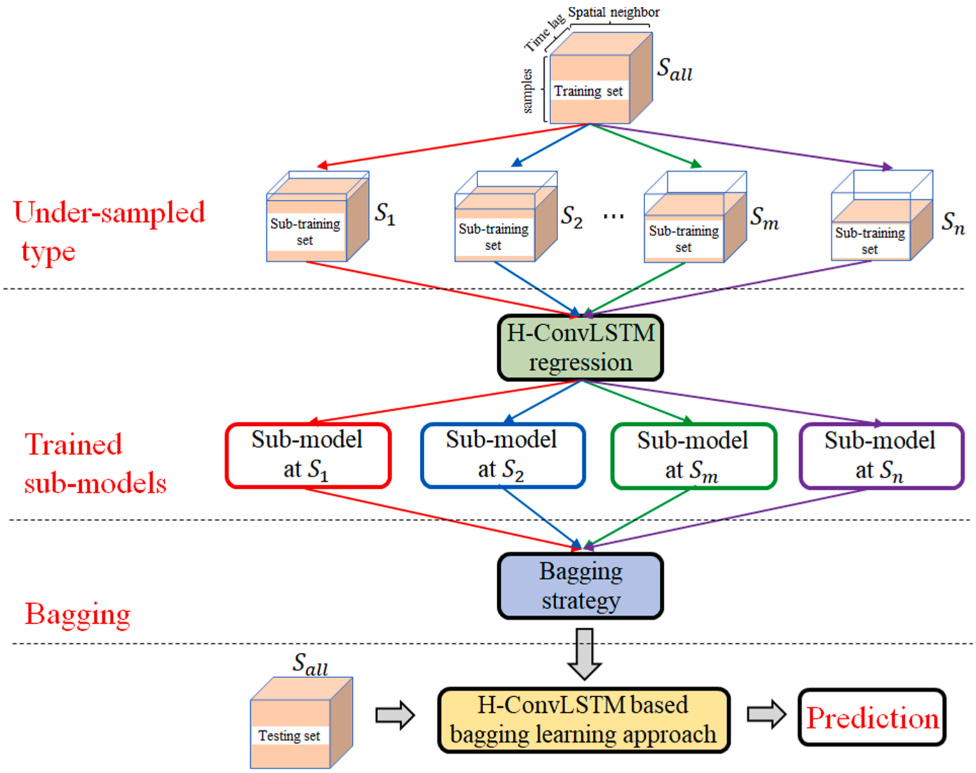


Fig. 3. The architecture of the H-ConvLSTM-based bagging learning approach.

configurations in urban areas to further optimize the distribution of charging stations. To the best of our knowledge, Ke et al. (2019) were the first to propose a successful hexagon-based deep learning model for travel demand prediction; they also discussed the advantages of the hexagonal partition approach mentioned above in detail. However, hexagonal data must be mapped to a matrix before executing feedforward propagation calculations, which destroys the spatial position relationships between the hexagonal partitions. The HexagDLy framework proposed by Steppa and Holch (2019) subtly solved this problem; however, it has difficulty grasping the complex time correlations in time series data.

2.3. Addressing sparse demand data

The highly skewed spatial and temporal distributions of ride-hailing demand lead to severe demand imbalances among spatiotemporal partitions. As a result, the demand information in minority spatiotemporal partitions is overwhelmed by that in majority spatiotemporal partitions. The different settings of a minimum ride-hailing demand threshold make the corresponding datasets have certain sparse distribution characteristics. The levels for these sparse demands are often difficult to accurately predict. The most common approaches for solving this problem include data-level methods, algorithm-level methods, and hybrid methods that combine the advantages of the other two types of techniques (Krawczyk, 2016). Data-level methods aim to change the input training set to fit a standard learning algorithm. To achieve a balanced data distribution, previous studies usually increased the number of minority ranges (the number of classes in a classification task or the target values in a regression task that have the lowest data sizes in the dataset) by oversampling (Chawla et al., 2002; Vluymans, 2019) or decreased the number of majority ranges (the number of classes in a classification task or the target values in a regression task that have the highest data sizes in the dataset) by undersampling (Lin et al., 2017). Moniz et al. (2017) combined resampling methods with standard regression models (such as SVMs) to achieve improved prediction accuracy for imbalanced time series. Zhang et al. (2021) proposed a clustering decision tree-based multimodel prediction method to solve the data imbalance problem in building energy load prediction. Cheng et al. (2020) developed a dynamic spatiotemporal k-nearest neighbor (D-STKNN) model to identify heterogeneous travel patterns in different temporal and spatial units, which were further considered for conducting short-term travel speed prediction to improve the prediction accuracy of the model. However, little effort has been directed toward solving the data imbalance problem while capturing the complex spatiotemporal correlations of ride-hailing demands with sparse uncertainties.

Although the mathematical structures of prediction models exhibit significant difference, the training objective of both statistical models and machine learning models is always the same: minimizing their total/mean prediction errors (loss function) on the observed or training dataset. The utilized evaluation indices (such as the symmetric mean absolute percentage error (SMAPE) and root mean

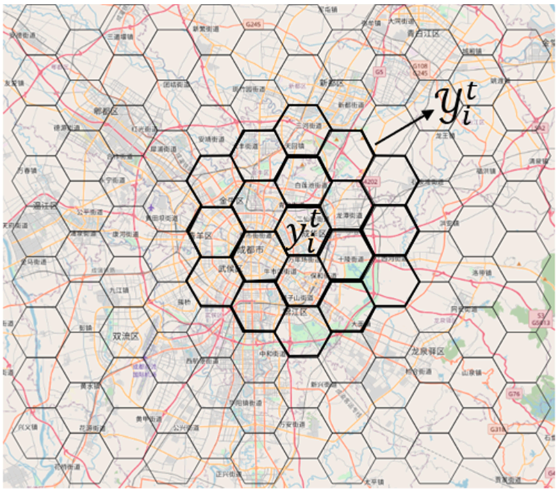


Fig. 4. The two-layer local adjacent map of y_i^t.

square error (RMSE)), guided by global prediction performance, are often biased toward the majority ranges of ride-hailing demand (Japkowicz and Stephen, 2002). The minority ranges of partitioned ride-hailing demand induce high costs when the demand is not well-predicted. Previous studies related to ride-hailing demand prediction usually filtered out large amounts of spatiotemporal units with sparse demand by setting a minimum ride-hailing demand threshold (Ke et al., 2017). Then, the dataset was always assumed have well-balanced spatiotemporal partitions with equal numbers of mispredictions. However, this assumption results in great bias in the prediction results due to the spatiotemporal data imbalance problem. Therefore, more attention should be given to designing appropriate prediction algorithms for imbalanced ride-hailing demand data and to ensuring good prediction performance in different spatial and temporal locations.

In this paper, we integrate the bias preferences of a standard prediction model with multiple majority ranges of ride-hailing demand to improve the total prediction accuracy. A hexagon is chosen as the basic spatiotemporal partition to facilitate the aggregation of ride-hailing demands with similar characteristics. Previous studies (Ke et al., 2019; Huang et al., 2019) usually focused their research area on limited ranges by setting minimum ride-hailing demand thresholds, as this is a common data processing method. Different minimum ride-hailing demand threshold settings cause the corresponding datasets to have their own majority ride-hailing demand ranges, leading to an imbalanced data problem with uncertain sparsities in ride-hailing demand prediction. Therefore, H-ConvLSTM is proposed as a submodel to compare the prediction performances achieved with different threshold settings, in which hexagonal convolution kernels are applied to directly conduct convolution calculations on hexagonal partitions. In addition, an H-ConvLSTM-based bagging learning approach is further proposed to integrate the optimal prediction ranges of the submodel at different data sampling degrees.

3. Methodology

Fig. 3 shows the architecture of the proposed H-ConvLSTM-based bagging learning approach for ride-hailing demand prediction. The architecture is composed of three parts. First, several undersampled datasets $S_1, \cdots S_n$ are established for all ride-hailing order data by setting a minimum ride-hailing demand threshold. Second, an H-ConvLSTM regression model is established, and the corresponding predictive submodels are trained on each subtraining dataset. Finally, a bagging strategy is developed to integrate the bias preferences of each submodel.

3.1. Preliminary

In this section, a city is divided into uniform hexagonal partitions, and a day is divided into uniform time intervals to aggregate the ride-hailing orders of different areas. Therefore, the ride-hailing demand y_i^t can be defined as the number of ride-hailing orders issued in hexagon partition i during time interval t.

Due to the presence of significant spatiotemporal correlations, h historical ride-hailing demand features of two-layer local adjacent maps $\mathscr{Y}_i^{t-h+1}, \cdots, \mathscr{Y}_i^t$ centralized at hexagon i, as shown in Fig. 4, are selected to jointly predict the ride-hailing demands y_i^{t+1} of target partition i for future time intervals.

3.2. H-ConvLSTM regression model

As an improved form of the LSTM model, ConvLSTM has convolutional structures in both the input-to-state and state-to-state

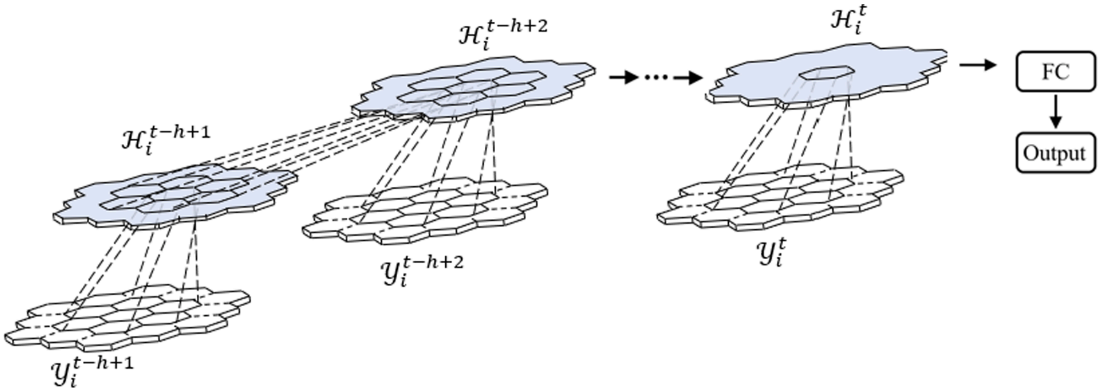


Fig. 5. The architecture of the H-ConvLSTM regression model.

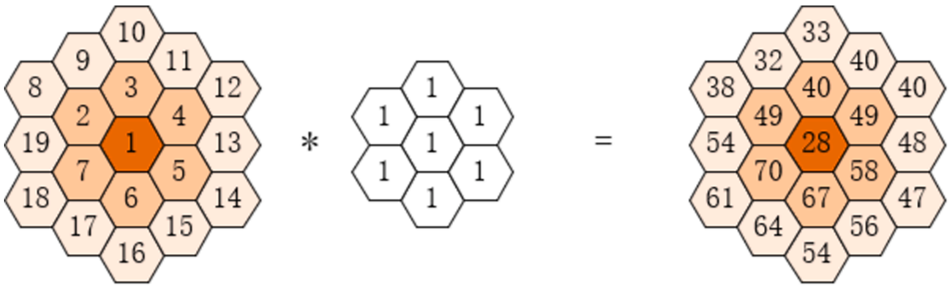


Fig. 6. Hexagonal convolution operation with a kernel size of 1.

transitions and has good performance in terms of simultaneously capturing temporal and spatial features. The key to ConvLSTM involves the cell states $\mathscr{C}_i^t$, which memorize and cycle information through gate structures that consist of forget gates f_i^t, input gates i_i^t and output gates $\mathscr{O}_i^t$. To capture spatial dependencies, the historical cell states $\mathscr{C}_i^{t-h+1},\cdots,\mathscr{C}_i^t$, input states $\mathscr{Y}_i^{t-h+1},\cdots,\mathscr{Y}_i^t$, hidden states $\mathscr{H}_i^{t-h+1},\cdots,\mathscr{H}_i^t$ and other gates of ConvLSTM are 3D tensors whose last two dimensions are rows and columns of spatial information. The forget gate layer f_i^t determines what information we discard from cell state $\mathscr{C}_i^t$. The input gate layer i_i^t determines what information to input and updates the old cell state C_i^{t-1} to C_i^t through a tanh layer. Then, parts of the cell state C_i^t determined by the output gate layer $\mathscr{O}_i^t$ are exported as the memorized hidden state $\mathscr{H}_i^t$.

To incorporate the advantages of hexagonal partitioning, we propose an H-ConvLSTM regression model to capture the spatiotemporal characteristics of ride-hailing demand, as shown in Fig. 5. H-ConvLSTM directly adopts hexagonal convolution calculations during feedforward propagation. Following previous research (Steppa and Holch, 2019), we apply a hexagonal convolution kernel to extract the spatial and temporal features of the two-layer local adjacency map, as shown in Fig. 6. The specific functional relationships of H-ConvLSTM are as follows:

$$f_i^t = \sigma\left(W_{hf}*\mathscr{H}_i^{t-1} + W_{yf}*\mathscr{Y}_i^t + b_f\right) \tag{1}$$

$$i_i^t = \sigma\left(W_{hi}*\mathscr{H}_i^{t-1} + W_{yi}*\mathscr{Y}_i^t + b_i\right) \tag{2}$$

$$\widetilde{C}_i^t = tanh\left(W_{hc}*\mathscr{H}_i^{t-1} + W_{yc}*\mathscr{Y}_i^t + b_c\right) \tag{3}$$

$$C_i^t = f_i^t \circ C_i^{t-1} + i_i^t \circ \widetilde{C}_i^t \tag{4}$$

$$\mathscr{O}_i^t = \sigma\left(W_{ho}*\mathscr{H}_i^{t-1} + W_{yo}*\mathscr{Y}_i^t + b_o\right) \tag{5}$$

$$\mathscr{H}_i^t = \mathscr{O}_i^t \circ tanh\left(C_i^t\right) \tag{6}$$

where * denotes the hexagonal convolution operator and ◦ denotes the Hadamard operator. W_{hf}, W_{hi}, W_{hc}, W_{ho}, W_{yf}, W_{yi}, W_{yc}, W_{yo}, b_f, b_i, b_c, b_o denote the trainable parameters. σ and $tanh$ denote the sigmoid and hyperbolic tangent activation functions, respectively.

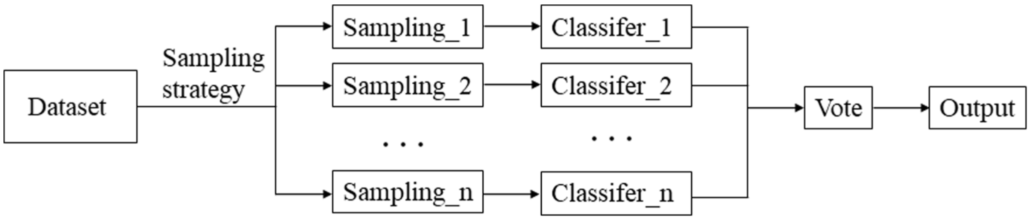


Fig. 7. Bagging structure.

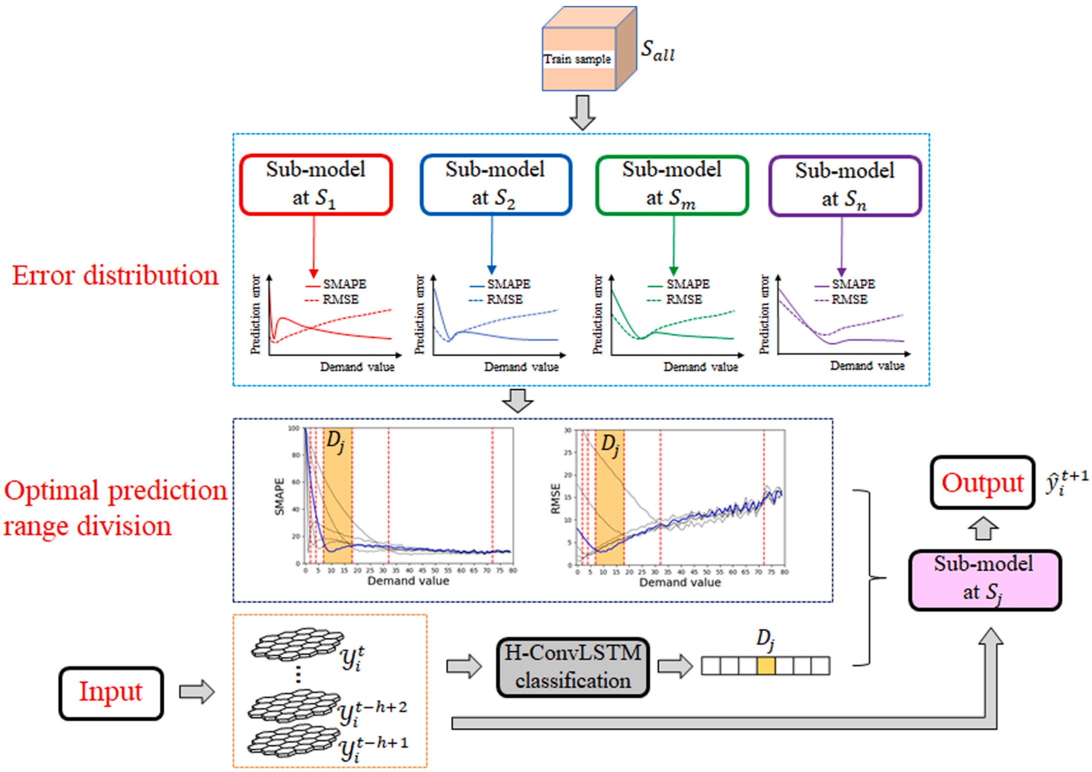


Fig. 8. Bagging strategy of the H-ConvLSTM-based bagging learning approach.

Following a series of fully connected layers, the ride-hailing demand $\hat{y}_i^{t+1}$ for location i and time interval $t+1$ can be predicted.

3.3. Bagging strategy

Bootstrap aggregation, known as bagging, is one of the earliest ensemble algorithms (Breiman, 1996). The bagging structure is shown in Fig. 7. The original dataset is sampled n times according to a certain sampling strategy, and n subdatasets are obtained. N weak classification models are trained on these subdatasets, and the final classification result is obtained by voting on the prediction results of each model. This algorithm effectively improves the classification performance of weak classifiers, especially when dealing with data imbalance problems.

The bagging strategy of the H-ConvLSTM-based bagging learning approach is shown in Fig. 8, and it contains three parts. First, the trained submodels are used to predict the total training set S_{all}, and the prediction error distribution of each trained submodel is counted. Then, the optimal prediction range of each submodel in terms of the demand value distribution is identified and labeled as a category. Finally, instead of utilizing the traditional voting method, the H-ConvLSTM classification model is trained on the total training set S_{all} to predict the potential range of the demand level for a certain location at a future time. The submodel with the best performance regarding the range of potential demand levels is selected to predict the future demand at this location.

The SMAPE and RMSE are selected as the prediction error evaluation indices, and they are formulated as follows:

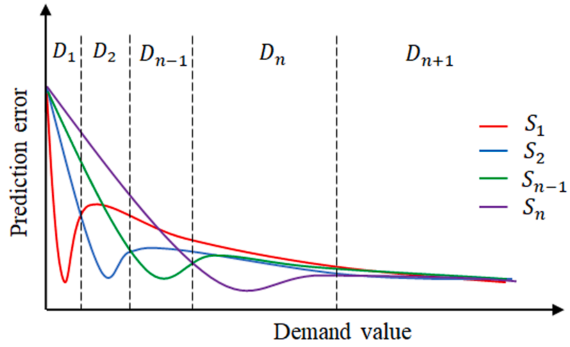


Fig. 9. Prediction error distributions of the submodel trained on datasets $S_1, \cdots S_n$.

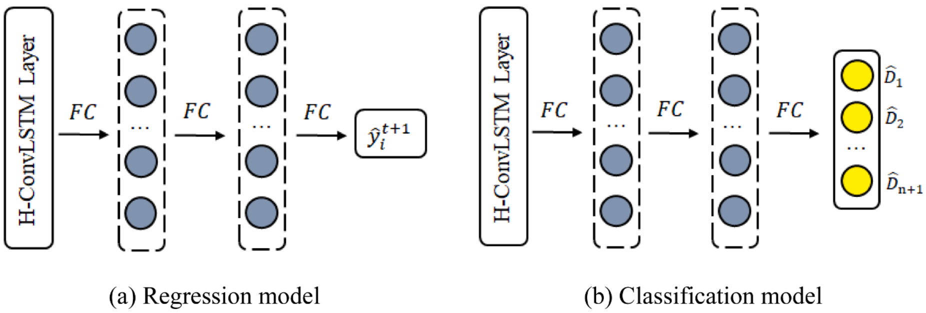


Fig. 10. Structures of the fully connected layers in the prediction models.

$$SMAPE = \frac{1}{n}\sum_{i=1}^{n}\frac{\left|\hat{y}_i^{t+1} - y_i^{t+1}\right|}{\left|\hat{y}_i^{t+1}\right| + \left|y_i^{t+1}\right| + \varepsilon} \tag{7}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\hat{y}_i^{t+1} - y_i^{t+1}\right)^2} \tag{8}$$

where $\hat{y}_i^{t+1}$ and y_i^{t+1} are the predicted ride-hailing demands and true ride-hailing demands, respectively, and ε is a very small value that prevents the denominator from being 0.

The undersampled datasets $S_1, \cdots S_n$ have different ride-hailing demand distribution structures. Different majority ranges of ride-hailing demand make the corresponding H-ConvLSTM regression submodels have their own prediction bias preferences in S_{all}, as shown in Fig. 9. Therefore, we divide the demand values into continuous $n+1$ sections $D_1, \cdots D_{n+1}$ according to size, in which the first n sections $D_1, \cdots D_n$ represent the optimal prediction ranges of the n submodels. Due to slight prediction performance differences regarding the demand distribution, we can generate two sets of boundary points $D^s_{1,2}, \cdots, D^s_{n,n+1}$ and $D^r_{1,2}, \cdots, D^r_{n,n+1}$, corresponding to the SMAPE and RMSE, respectively. The final optimal prediction range boundary points $D_{1,2}, \cdots D_{n,n+1}$ can be obtained by taking the average values of the two sets of data.

Different from the corresponding regression model, the H-ConvLSTM classification model identifies the potential range category $\hat{D}_j$ of the ride-hailing demand y_i^{t+1} for a future time interval based on the historical spatiotemporal ride-hailing demand features $\mathscr{Y}_i^{t-h+1}$, $\cdots$, $\mathscr{Y}_i^{t}$, as shown in Fig. 10. One-hot encoding is used to convert the categories $D_1, \cdots D_{n+1}$ to binary vectors of length $n+1$. The H-ConvLSTM submodel at S_j corresponding to the predicted range category $\hat{D}_j$ is selected to obtain the predicted ride-hailing demand $\hat{y}_i^{t+1}$ at a future time.

4. Experimental results

4.1. Dataset and model setup

The dataset, including all the online ride-hailing order data for Chengdu in November 2016, is provided by the Didi Gaia Plan platform. To achieve better prediction performance, the selection of the spatiotemporal granularities in this case follows the research

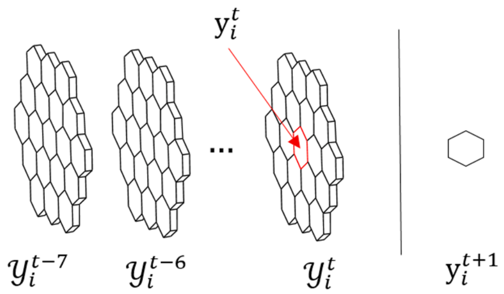


Fig. 11. The contents of a sample group.

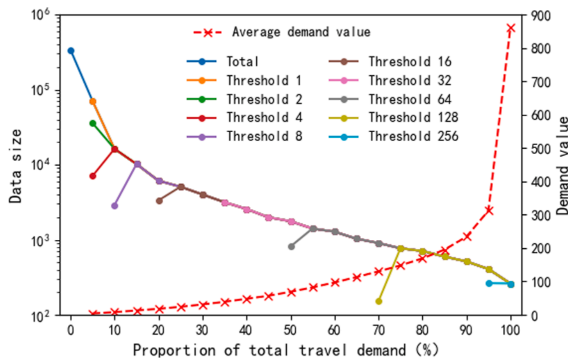


Fig. 12. The data size distributions obtained under different minimum ride-hailing demand thresholds and the average demand values over the total ride-hailing demand.

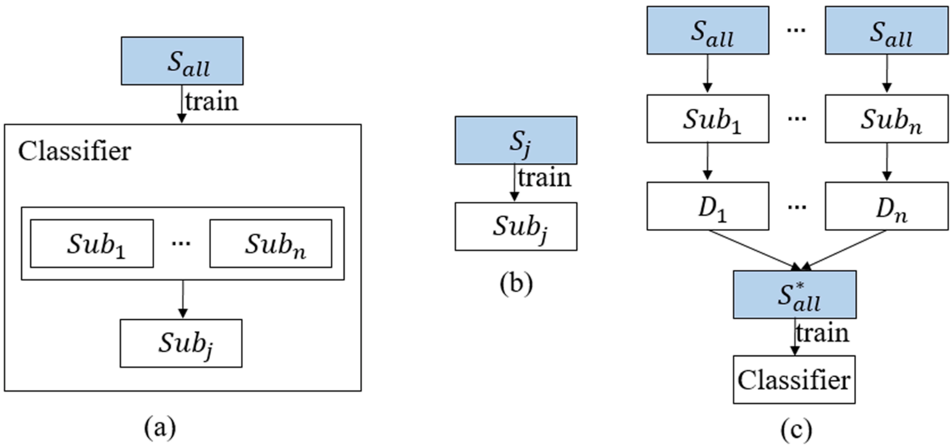


Fig. 13. Training process of the H-ConvLSTM-based bagging learning approach.

of Liu et al. (2022). Each day is decentralized by setting 30 min as the time interval, and a time partition label is added for each order data point based on its starting time. Then, hexagonal partitions are added to the urban space based on the Quantum Geographic Information System (QGIS), and the intersection operation is performed with the order data and their added time partition labels. The city is divided into 35 × 46 hexagonal partitions with a side length of 800 m, and each order data point is further labeled with a hexagonal partition ID. Based on the time interval labels and the hexagonal partition IDs, we can easily aggregate the ride-hailing demand into different spatiotemporal partitions. Two-layer local adjacent maps centralized at the target partition in the previous 8 time intervals are used to predict the ride-hailing demand in the next time interval. Therefore, during the training and testing processes of the proposed deep learning model, a travel demand sample y_i^t needs to be expanded into the corresponding sample group $G_i^t =$

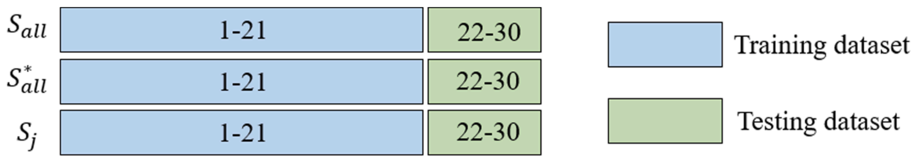
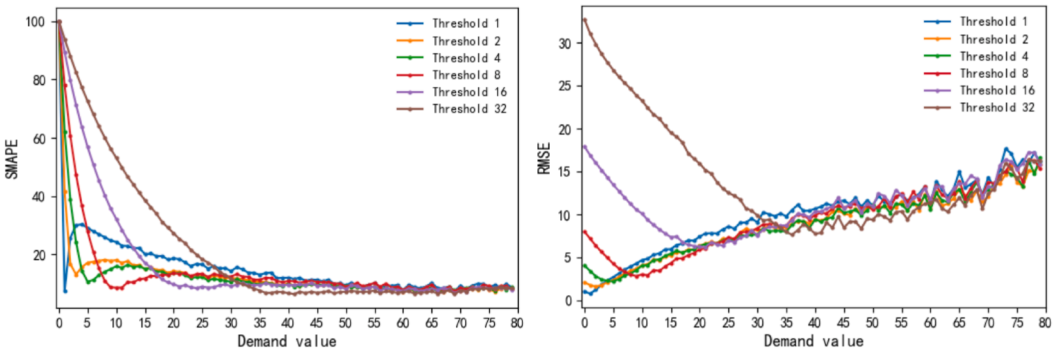


Fig. 14. Division of the training dataset and testing dataset.



(a) SMAPE (b) RMSE

Fig. 15. The distributions of the demand value prediction errors obtained under different minimum ride-hailing demand thresholds.

$\{\mathscr{Y}_i^{t-7},\cdots,\mathscr{Y}_i^{t},y_i^{t+1}\}$, where $\mathscr{Y}_i^{t-7},\cdots,\mathscr{Y}_i^{t}$ represents the input of the model and y_i^{t+1} represents the corresponding label, as shown in Fig. 11.

Minimum ride-hailing demand thresholds are set for all spatiotemporal partitions (from 1 to 256, doubling each time) to create multiple datasets with different ride-hailing demand coverages. The ride-hailing demands that are less than the corresponding threshold in each dataset are excluded. In other words, if the travel demand sample y_i^t at the center of $\mathscr{Y}_i^t$ is less than the threshold, the sample group G_i^t is removed from the corresponding dataset.

The ride-hailing demands are arranged in order from small to large and divided into 20 equal parts according to their proportions of the total ride-hailing demand. The data size distributions obtained under different minimum ride-hailing demand thresholds and the average demand values over the total ride-hailing demand are shown in Fig. 12. The left axis represents the data size of the subdataset corresponding to the minimum ride-hailing demand threshold (1 to 256) in each demand range, and the right axis represents the average demand value of S_{all} in each demand range. As the minimum ride-hailing demand threshold increases, the corresponding majority ride-hailing demand ranges continuously increase.

The training process of the proposed H-ConvLSTM-based bagging learning approach is shown in Fig. 13(a). It consists of multiple regression submodels (H-ConvLSTM regression models) and a classifier (an H-ConvLSTM classification model), which are trained as shown in Fig. 13(b) and Fig. 13(c), respectively. Each submodel Sub_j is trained on the corresponding subdataset S_j, which is an undersampling of the total dataset S_{all}. By evaluating the prediction performance of each submodel Sub_j on S_{all}, the individual optimal prediction ranges D_j can be identified and labeled as separate classes. Then, the dataset S^*_{all} is obtained on the basis of S_{all} by replacing the labels of the sample data with the range categories to which they belong. The classifier is trained on S^*_{all} to identify the potential range of the predicted travel demand.

The division of the training dataset and testing dataset is shown in Fig. 14. The data of S_{all}, S_j and S^*_{all} in the first 21 days are used for training, and the data from the last 9 days are used for testing. The testing process of the H-ConvLSTM-based bagging learning approach is similar to that shown in Fig. 13(a). First, the trained classifier is used to select an appropriate submodel for S_{all}'s input demand, and then this submodel is used to predict the corresponding future demand.

The experimental platform is a server with an Intel(R) Xeon(R) Gold-5218 CPU @ 2.30 GHz, 128 GB of RAM, and one GPU (NVIDIA Quadro RTX 5000). The proposed model is implemented in Python 3.6.6 with PyTorch, TensorFlow and Keras. The proposed H-ConvLSTM regression and classification models both consist of 4 ConvLSTM layers, which have 8, 16, 32, and 32 hidden states. The hexagonal kernel size of each layer is 1. To ensure that the input and output of the hexagonal convolution operation have the same dimensionality, similar to the same padding approach used in traditional CNN models, virtual hexagons with zero demand values are padded as neighbors of the hexagons on the border. Batch normalization and dropout are used for training the model. The number of training epochs is set to 50 with a batch size of 128. Adam is used for optimization with a learning rate of 0.0001. The weighted sum of the SMAPE and RMSE is used as the loss function of the regression model, while the classification cross entropy is used as the loss function of the classification model. The SMAPE and RMSE are used to evaluate the prediction performance of the demand value distribution yielded by the regression model.

Table 1
Statistical results of the boundary points.

Type	Boundary point demand values					
	$D_{1,2}$	$D_{2,4}$	$D_{4,8}$	$D_{8,16}$	$D_{16,32}$	$D_{32,r}$
SMAPE	1	3	7	17	32	67
RMSE	2	4	7	22	34	59
Average value	1.5	3.5	7	19.5	33	63

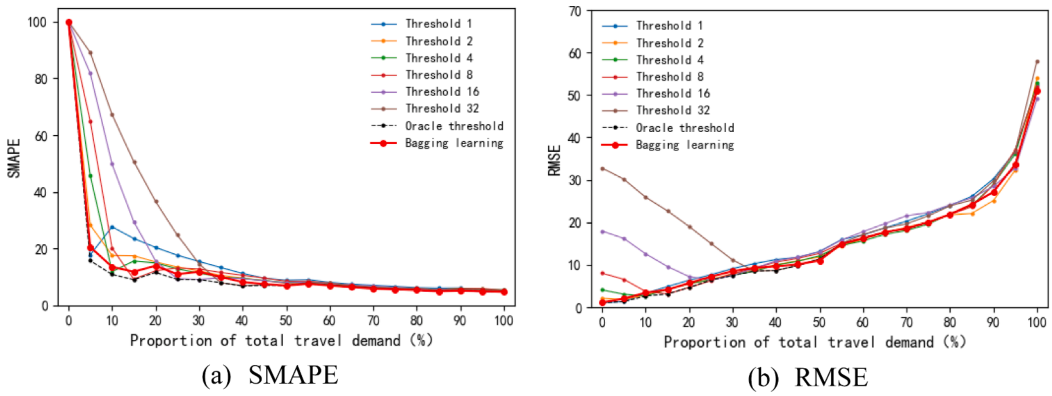


Fig. 16. Prediction errors of the bagging learning approach based on H-ConvLSTM.

4.2. Optimal prediction range division results

The H-ConvLSTM regression submodel is trained on each undersampled dataset from S_1 to S_{256}, and the corresponding prediction performance is calculated. Since only a threshold setting between 1 and 32 can produce a relatively obvious optimal prediction distribution range, we only select the corresponding submodels with this characteristic as the research objects, and the prediction results are shown in Fig. 15. Each prediction distribution curve first exhibits a decreasing trend and then increases near the threshold point. As a percentage error that is sensitive to sparse demand, the SMAPE is mainly used to reflect the influence of different thresholds on the resulting prediction performances. The submodel corresponding to each threshold has an obvious optimal prediction range, and the ride-hailing demand values can be divided into 7 segments according to size. The RMSE is an absolute error and is sensitive to large outliers. Although the RMSEs of the submodels also perform best when the demand values are slightly larger than the threshold, the prediction performance corresponding to these demand values is difficult to make as obvious as that obtained with the SMAPE because their distribution is located in a smaller demand range. Therefore, the prediction result distribution of the RMSE is mainly a supplementary validation of the SMAPE.

The boundary points between each segment are determined as shown in Table 1. The intersection points of adjacent optimal ranges are selected as the first five boundary points. The last boundary point is the closest intersection between the prediction distribution curve of threshold 32 and the other distribution curves on the right. The classification numbers of the demand values distributed in the final 7 segments are set to 1, 2, 3, 4, 5, 6, and 7 and further transformed into corresponding binary vectors through one-hot encoding. Then, the dataset S_{all}^{*} is obtained on the basis of S_{all} by replacing the label y_i^{t+1} of the sample group G_i^t with the classification number to which it belongs.

4.3. Results of the H-ConvLSTM-based bagging learning approach

The H-ConvLSTM classification model is trained on S_{all}^{*}. Similar to the H-ConvLSTM regression submodel, 8 historical ride-hailing demand features of two-layer local adjacent maps $\mathscr{Y}_i^{t-7}, \cdots, \mathscr{Y}_i^{t}$ centralized at hexagon i, are selected to jointly predict the segment category of ride-hailing demands y_i^{t+1} of target partition i for future time interval $t+1$. An accuracy of 85.76% is achieved on the testing dataset (88.94% on the training dataset). The boundaries of segment categories depend on the prediction distribution of each submodel in the training set of S_{all}, and it is assumed that the optimal prediction range of each submodel in the training set and testing set is roughly similar. For the historical data in the testing set of S_{all}, whose prediction categories are the first 6 segments $D_1, \cdots D_6$, the corresponding regression submodel trained on the training set of S_j is used to predict the ride-hailing demand at a future time. The data size of the ride-hailing demand distributed in the last segment D_7 is relatively small as shown in Fig. 12, and no submodel shows significantly better prediction performance in this segment. Therefore, the average value of the prediction results of the 6 submodels is used as the predicted value of ride-hailing demand of segment D_7 at a future time. The prediction error distribution of the H-

Table 2
Model performance comparison.

Model	wSMAPE ($\times 10^{-2}$)	wRMSE	Training time (h)	Testing time (min)
ARIMA	14.53	23.51	0.01	0.01
H-ANN	14.11	22.75	0.31	0.04
H-CNN	13.21	22.36	4.43	0.54
H-CNN-LSTM	12.35	21.12	6.05	0.96
H-CNN-GRU	12.29	21.53	5.71	0.88
H-ConvLSTM + Threshold 1	11.61	20.97	7.16	1.02
H-ConvLSTM + Threshold 2	10.02	19.58	5.36	0.83
H-ConvLSTM + Threshold 4	10.54	20.04	3.59	0.57
H-ConvLSTM + Threshold 8	11.81	20.21	2.33	0.36
H-ConvLSTM + Threshold 16	14.62	20.76	1.02	0.15
H-ConvLSTM + Threshold 32	18.76	24.18	0.75	0.11
H-ConvLSTM + bagging	**9.42**	**18.63**	**25.84**	**1.62**

ConvLSTM-based bagging learning approach is shown in Fig. 16.

Compared with that of the H-ConvLSTM regression submodels trained under different minimum ride-hailing demand threshold settings, the prediction performance of the H-ConvLSTM-based bagging learning approach is improved by different degrees and is closer to the optimal performance limit that can be achieved by this method (i.e., an oracle submodel classifier that always selects the model that performs best, as shown by the dotted black line).

To verify the validity of the proposed model, several basic models are selected for comparison, as follows.

1) ARIMA: This is the autoregressive integrated moving average model that is widely used for time series prediction. 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) Hexagonal artificial neural network (H-ANN): The spatial feature and historical temporal feature of the demands of a hexagonal partition are spliced together as the input for a fully connected neural network, and the predicted demand value of a future time is output. The model includes 5 fully connected layers, which have 128, 64, 32, 16, and 8 hidden neurons.

3) H-CNN: The previous 8 time intervals are represented by the numbers of channels in the input image. A hexagonal convolution operation is applied between each pair of layers. The H-CNN model includes 4 convolution layers, which have 8, 16, 32, and 32 hidden states. The hexagonal kernel size of each layer is 1. Batch normalization and dropout are used to train the model.

4) H-CNN-LSTM: An H-CNN model with one channel is selected to extract the spatial ride-hailing demand characteristics of the previous 8 time intervals. The settings for the convolution layer and the convolution kernel remain the same. The outputs of the H-CNN for the previous 8 time intervals are expanded into vectors and used as the inputs for the LSTM to extract the temporal characteristics of ride-hailing demand. The hidden state of the LSTM is set to 128.

5) H-CNN-GRU: The output of the H-CNN is taken as the input of a GRU, and the other settings are consistent with those of H-CNN-LSTM.

The data sizes of different ride-hailing demands are greatly different, which causes the overall prediction error to be significantly affected by the sparse ride-hailing demand prediction results with large data sizes. The sparse ride-hailing demands, which account for approximately 70% of the total data size (i.e., 70% of the spatiotemporal partitions are sparse demands), contain less than 10% of the total ride-hailing demand quantity. To better evaluate the prediction performance of each model, we propose to utilize the weighted SMAPE (wSMAPE) and weighted RMSE (wRMSE) to comprehensively consider the prediction results corresponding to different ride-hailing demand size distributions as follows:

$$wSMAPE = \frac{1}{\sum_{i=1}^{20} w_i n_i} \sum_{i=1}^{20} \left(w_i \sum_{j=1}^{n_i} \left| \frac{\widehat{y}_{ij} - y_{ij}}{\widehat{y}_{ij} + y_{ij} + \varepsilon} \right| \right) \tag{9}$$

$$wRMSE = \sqrt{\frac{1}{\sum_{i=1}^{20} w_i n_i} \sum_{i=1}^{20} \left(w_i \sum_{j=1}^{n_i} \left(\widehat{y}_{ij} - y_{ij} \right)^2 \right)} \tag{10}$$

The ride-hailing demands are arranged in order from small to large and divided into 20 equal parts according to their proportions of the total demand value. The subdata size of each 5% ride-hailing demand segment is denoted as $n_1, n_2 \cdots n_{20}$. $w_i = \frac{n_1}{n_i}$ denotes the weight of segment i. $\widehat{y}_{ij}$ denotes the j th predicted value of segment i, and y_{ij} is the corresponding true value. ε is a very small value that prevents the denominator from being 0.

The overall prediction performance achieved by each model on the testing dataset is shown in Table 2. With the enhancement in the ability of the model to capture the temporal and spatial characteristics of ride-hailing demand, both the wSMAPE and wRMSE of the H-

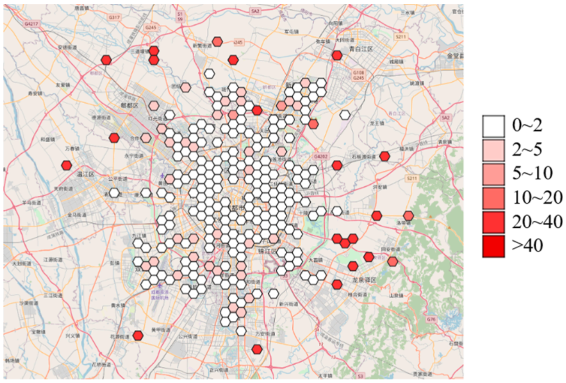


Fig. 17. Spatial distribution of the wSMAPE difference between the H-ConvLSTM-based bagging learning approach and H-ConvLSTM-Oracle.

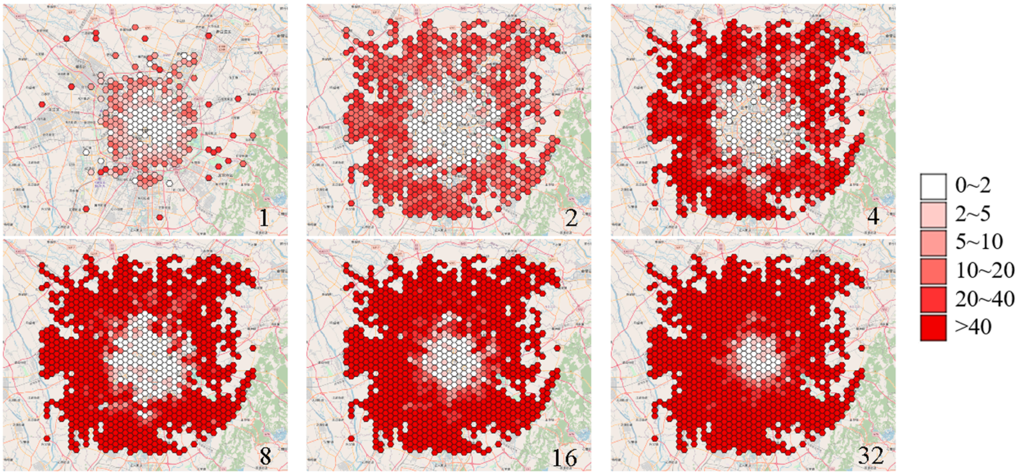


Fig. 18. Spatial distributions of the wSMAPE differences between the submodels and H-ConvLSTM-Oracle.

ConvLSTM regression model are lower values. By integrating the bias prediction preferences of each submodel in different segments, the prediction performance of our proposed bagging learning approach based on H-ConvLSTM improves by 5.99% and 4.85% over the values obtained with the optimal threshold setting in terms of the wMAPE and wRMSE, respectively. Due to the inclusion of multiple regression submodels and an additional classification model, the proposed H-ConvLSTM-based bagging learning approach requires more training time.

Assume that H-ConvLSTM-Oracle has an oracle submodel classifier that always selects the version that performs best. This model represents the upper bound performance of our H-ConvLSTM-based bagging learning approach. The spatial distributions of the wSMAPE differences between the H-ConvLSTM-based bagging learning approach and each of the other 6 submodels against H-ConvLSTM-Oracle are shown in Fig. 17 and Fig. 18, respectively. A smaller difference value means that the wSMAPE value is close to that of H-ConvLSTM-Oracle, and the corresponding prediction results are better. The proposed H-ConvLSTM-based bagging learning approach effectively selects the optimal submodel in the whole spatial distribution, and the prediction results are less different from those of H-ConvLSTM-Oracle. When compared with the results obtained under a minimum ride-hailing demand threshold of 1, the prediction results of H-ConvLSTM-Oracle are mainly improved in the central urban area, especially in the transition areas between urban and suburban areas. For other minimum ride-hailing demand threshold settings, the prediction results of H-ConvLSTM-Oracle are more significantly improved in the outer suburbs. With the continuous increase in the threshold value, the improvement effect and coverage area of the corresponding H-ConvLSTM-Oracle prediction results continuously increase.

5. Conclusion

In this paper, we propose an H-ConvLSTM regression model to compare and analyze the ride-hailing demand prediction performances achieved under different data distribution characteristics. Minimum ride-hailing demand thresholds are set for all spatio-temporal partitions to create multiple datasets with different data sparsities. The H-ConvLSTM regression models trained using different datasets have their own optimal prediction ranges on the testing set, and each prediction distribution curve first exhibits a decreasing trend and then increases near the threshold point.

An H-ConvLSTM-based bagging learning approach is further proposed to integrate the bias prediction preferences of each H-ConvLSTM regression model trained on datasets with different data sparsities. An experimental analysis conducted on the order data obtained from Didi Chuxing in Chengdu city over one month shows that the proposed H-ConvLSTM-based bagging learning approach can achieve significantly improved prediction performance.

In future work, we are committed to performing more in-depth qualitative and quantitative analyses of the spatiotemporal scale of internet-based ride-hailing demand. The influence of an imbalanced ride-hailing demand distribution (caused by the division of different spatial and temporal scales) on the prediction performance will be discussed. Policy recommendations will also be made to improve the operational efficiency and quality of ride-hailing.

CRediT authorship contribution statement

Zhiju Chen: Conceptualization, Methodology, Software, Visualization, Investigation, Writing – original draft, Writing – review & editing. **Kai Liu:** Conceptualization, Methodology, Software, Visualization, Investigation, Writing – original draft, Writing – review & editing. **Jiangbo Wang:** Writing – review & editing, Investigation. **Toshiyuki Yamamoto:** Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The authors do not have permission to share data.

Acknowledgments

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

Appendix A. Comparison of predicted results between H-ConvLSTM and ConvLSTM

Traditional ConvLSTM based on matrix convolution operation is compared with our H-ConvLSTM model to verify the advantages of hexagonal convolution operation. The model parameter configuration can refer to Liu et al. (2022) for a detailed explanation and instruction. The predicted results are shown in Table A1. Compared with a square, a hexagon is closer to a circle, and its distribution is symmetric and equivalent. Therefore, travel demands with hexagon partition are more accurately predicted. Since the hexagonal convolution operation solves the problem of topological loss of spatial relations caused by matrix transformation in traditional ConvLSTM, the proposed H-ConvLSTM model shows stable optimal prediction performance.

Table A1
Comparison of different demand prediction models.

Model	Partition shape	RMSE			MAPE ($\times 10^{-2}$)		
		Testing set	Avg.	Sd.	Testing set	Avg.	Sd.
ConvLSTM	Square	9.37	9.38	0.12	17.18	17.55	0.46
	Hexagon	9.03	9.12	0.07	17.02	17.27	0.55
H-ConvLSTM	Hexagon	8.82	8.80	0.05	16.71	16.76	0.36

References

Alisoltani, N., Leclercq, L., Zargayouna, M., 2021. Can dynamic ride-sharing reduce traffic congestion? Transp. Res. Part B Methodol. 145, 212–246.
Alonso-Mora, J., Samaranayake, S., Wallar, A., Frazzoli, E., Rus, D., 2017. On-demand high-capacity ride-sharing via dynamic trip-vehicle assignment. Proc. Natl. Acad. Sci. U. S. A. 114 (3), 462–467.
Birch, C.P.D., Vuichard, N., Werkman, B.R., 2000. Modelling the effects of patch size on vegetation dynamics: Bracken [Pteridium aquilitnum (L.) Kuhn] under grazing. Ann. Bot. 85, 63–76.
Breiman, L., 1996. Bagging predictors. Mach. Learn. 24 (2), 123–140.
Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P., 2002. SMOTE: Synthetic Minority Over-sampling Technique. J. Artif. Intell. Res. 16, 321–357.
Chen, Zhiju, Liu, Kai, Feng, Tao, 2022. Examine the Prediction Error of Ride-Hailing Travel Demands with Various Ignored Sparse Demand Effects. J. Adv. Transp. 2022, 7690309 https://doi.org/10.1155/2022/7690309.
Chen, X., Zheng, H., Ke, J., Yang, H., 2020. Dynamic optimization strategies for on-demand ride services platform: Surge pricing, commission rate, and incentives. Transp. Res. Part B Methodol. 138, 23–45.
Cheng, S., Lu, F., Peng, P., 2020. Short-Term Traffic Forecasting by Mining the Non-Stationarity of Spatiotemporal Patterns. IEEE Trans. Intell. Transp. Syst. 22 (10), 6365–6383.
Csiszár, C., Csonka, B., Földes, D., Wirth, E., Lovas, T., 2019. Urban public charging station locating method for electric vehicles based on land use approach. J. Transp. Geogr. 74, 173–180.
Daganzo, C.F., Ouyang, Y., Yang, H., 2020. Analysis of ride-sharing with service time and detour guarantees. Transp. Res. Part B Methodol. 140, 130–150.
Davis, N., Raina, G., Jagannathan, K., 2016. A multi-level clustering approach for forecasting taxi ride-hailing demand. In: IEEE Conf. Intell. Transp. Syst. Proceedings, ITSC, pp. 223–228.
González, M.C., Hidalgo, C.A., Barabási, A.L., 2008. Understanding individual human mobility patterns. Nature 453, 779–782.
Guo, X.T., Caros, N.S., Zhao, J.H., 2021. Robust matching-integrated vehicle rebalancing in ride-hailing system with uncertain demand. Transp. Res. Part B Methodol. 150, 161–189.
Huang, Z., Huang, G., Chen, Z., Wu, C., Ma, X., Wang, H., 2019. Multi-regional online car-hailing order quantity forecasting based on the convolutional neural network. Inf. 10 (6), 193.
Japkowicz, N., Stephen, S., 2002. The class imbalance problem: A systematic study. Intell. Data Anal. 6, 429–449.
Jo, D., Yu, B., Jeon, H., Sohn, K., 2019. Image-to-image learning to predict traffic speeds by considering area-wide spatio-temporal dependencies. IEEE Trans. Veh. Technol. 68, 1188–1197.
Jiang, X., Zhang, L., Chen, M.X., 2014. Short-term forecasting of high-speed rail demand: A hybrid approach combining ensemble empirical mode decomposition and gray support vector machine with real-world applications in China. Transp. Res. Part C Emerg. Technol. 44, 110–127.
Kaltenbrunner, A., Meza, R., Grivolla, J., Codina, J., Banchs, R., 2010. Urban cycles and mobility patterns: Exploring and predicting trends in a bicycle-based public transport system. Pervasive Mob. Comput. 6, 455–466.
Ke, J., Yang, H., Zheng, H., Chen, X., Jia, Y., Gong, P., Ye, J., 2019. Hexagon-Based Convolutional Neural Network for Supply-Demand Forecasting of Ride-Sourcing Services. IEEE Trans. Intell. Transp. Syst. 20, 4160–4173.
Ke, J., Zheng, H., Yang, H., Chen, X. (Michael), 2017. Short-term forecasting of passenger demand under on-demand ride services: A spatio-temporal deep learning approach. Transp. Res. Part C Emerg. Technol. 85, 591–608.
Krawczyk, B., 2016. Learning from imbalanced data: open challenges and future directions. Prog. Artif. Intell. 5, 221–232.
Lin, W.C., Tsai, C.F., Hu, Y.H., Jhang, J.S., 2017. Clustering-based undersampling in class-imbalanced data. Inf. Sci. (Ny) 409–410, 17–26.
Liu, K., Chen, Z., Yamamoto, T., Tuo, L., 2022. Exploring the impact of spatiotemporal granularity on the demand prediction of dynamic ride-hailing. preprint arXiv: 2203.10301.
Ma, Z., Xing, J., Mesbah, M., Ferreira, L., 2014. Predicting short-term bus passenger demand using a pattern hybrid approach. Transp. Res. Part C Emerg. Technol. 39, 148–163.
Min, W., Wynter, L., 2011. Real-time road traffic prediction with spatio-temporal correlations. Transp. Res. Part C Emerg. Technol. 19, 606–616.
Moniz, N., Branco, P., Torgo, L., 2017. Resampling strategies for imbalanced time series forecasting. Int. J. Data Sci. Anal. 3, 161–181.
Shen, X., Zhou, Y., Jin, S., Wang, D., 2020. Spatiotemporal influence of land use and household properties on automobile ride-hailing demand. Transp. Res. Part D Transp. Environ. 84, 102359.
Shi, X., Chen, Z., Wang, H., Yeung, D.-Y., Wong, W., Woo, W., 2015. Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting. Adv. Neural Inf. Process. Syst. 2015-Janua, 68-80.
Shoman, W., Alganci, U., Demirel, H., 2019. A comparative analysis of gridding systems for point-based land cover/use analysis. Geocarto Int. 34, 867–886.
Steppa, C., Holch, T.L., 2019. HexagDLy-Processing hexagonally sampled data with CNNs in PyTorch. SoftwareX 9, 193–198.
Vazifeh, M.M., Santi, P., Resta, G., Strogatz, S.H., Ratti, C., 2018. Addressing the minimum fleet problem in on-demand urban mobility. Nature 557, 534–538.
Vluymans, S., 2019. Learning from imbalanced data. Stud. Comput. Intell. 807, 81–110.
Wang, D., Yang, Y., Ning, S., 2018. DeepSTCL: A Deep Spatio-temporal ConvLSTM for Ride-hailing demand Prediction. In: 2018 International Joint Conference on Neural Networks (IJCNN). IEEE, pp. 1–8.
Wang, H., Yang, H., 2019. Ridesourcing systems: A framework and review. Transp. Res. Part B Methodol. 129, 122–155.
Wu, X., Guo, J., Xian, K., Zhou, X., 2018. Hierarchical ride-hailing demand estimation using multiple data sources: A forward and backward propagation algorithmic framework on a layered computational graph. Transp. Res. Part C Emerg. Technol. 96, 321–346.
Xu, Z., Yin, Y., Ye, J., 2020. On the supply curve of ride-hailing systems. Transp. Res. Part B Methodol. 132, 29–43.
Yang, G., Wang, Y., Yu, H., Ren, Y., Xie, J., 2018. Short-Term Traffic State Prediction Based on the Spatiotemporal Features of Critical Road Sections. Sensors 18, 2287.
Yu, H., Wu, Z., Wang, S., Wang, Y., Ma, X., 2017. Spatiotemporal recurrent convolutional networks for traffic prediction in transportation networks. Sensors (Switzerland) 17, 1–16.
Yuan, C., Yu, X., Li, D., Xi, Y., 2019. Overall Traffic Mode Prediction by VOMM Approach and AR Mining Algorithm with Large-Scale Data. IEEE Trans. Intell. Transp. Syst. 20, 1508–1516.
Zhang, C., Li, J., Zhao, Y., Li, T., Chen, Q., Zhang, X., Qiu, W., 2021. Problem of data imbalance in building energy load prediction: Concept, influence, and solution. Appl. Energy 297, 117139.
Zhang, J., Zheng, Y., Qi, D., Li, R., Yi, X., 2016. DNN-based prediction model for spatio-temporal data. In: Proceedings of the 24th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems - GIS '16. ACM Press, New York, New York, USA, pp. 1–4.
Zheng, W., Lee, D.H., Shi, Q., 2006. Short-term freeway traffic flow prediction: Bayesian combined neural network approach. J. Transp. Eng. 132, 114–121.
Zhu, Z., Tang, L., Xiong, C., Chen, X., Zhang, L., 2019. The conditional probability of travel speed and its application to short-term prediction. Transp. B 7 (1), 684–706.
Zhu, Z., Ke, J., Wang, H., 2021. A mean-field Markov decision process model for spatial-temporal subsidies in ride-sourcing markets. Transp. Res. Part B Methodol. 150, 540–565.