SSA-CONV-LSTM-BASED TIME-SPACE DEMAND FORECASTING METHOD FOR ONLINE CAR RESERVATION

Xue XING[1]*, Yaqi ZHAI[2]

In order to address the disparity between residents' travel demand and online car rental supply under certain space-time conditions and to improve the travel efficiency of passengers, a method for predicting online car rental demand is proposed herein. Accordingly, the popular point of interest (POI) coordinates was selected as the node to divide the geographical map according to the Tyson polygon, the origin-destination (OD) demand data from the drip track data of the local area of the second ring road in Chengdu was extracted, and the OD data was matched with the demand of the divided area to obtain the OD flow matrix. The extracted OD flow matrix was then used to study the law of online car demand under the condition of space-time change, and it was concluded that the change of online car demand exhibits a clear temporal and spatial correlation. Subsequently, the sparrow optimization algorithm-based research on the online car demand prediction method based on the convolution-long and short memory neural network (SSA-Conv-LSTM) was conducted. The results indicate that the SSA-Conv-LSTM network demand forecasting model for car rental was 0.8 higher than the Conv-LSTM model and had greater accuracy and reliability than alternative models. In addition, the SSA-Conv-LSTM prediction model further improves the accuracy of the model for the demand prediction of online car reservations and provides data support to the relevant departments to formulate a reasonable and feasible regional scheduling plan for online car reservations.

Keywords: Online car demand forecast; Sparrow optimization algorithm; Conv-LSTM network; Tyson polygon

1. Introduction

Numerous nations are currently committed to the digitalization, networking, and intelligence of transportation. The ability to accurately predict traffic flow has become the determining factor in the development of intelligent transportation systems. The accurate demand forecast for online car reservations facilitates the formulation of a reasonable and effective regional scheduling plan for online car

[1]* Prof., College of Information and Control Engineering, Jilin Institute of Chemical Technology, Jilin, China, corresponding author, e-mail: xingx@jlict.edu.cn
[2] MD, College of Information and Control Engineering, Jilin Institute of Chemical Technology, Jilin, China, e-mail: zyq15843292545@163.com

reservations and the reduction of residents' travel waiting time by the relevant departments.

There are three primary research methods for online car and taxi demand forecasting: parametric forecasting mode, non-parametric forecasting model, and intelligent combination forecasting model. Rodrigues et al., [1] utilized the ARIMA model to predict taxi demand and analyzed each seasonal cycle, added external factors to the ARIMA model in the form of Fourier terms, and displayed the effect of the model on the data sets of four different clusters, demonstrating good results.

Numerous scholars have also investigated non-parametric prediction models. Yan et al., [2] highlighted a classification and regression tree-k nearest neighbor (CART-KNN) combination forecasting model for predicting short-term taxi demand. Additionally, Xu [3] designed a KNN-based real-time taxi demand forecasting system for spatiotemporal data. Moreover, New York City taxi demand was forecast by Poongodi et al., [4] using a combination of XGBoost (eXtreme Gradient Boosting) and multi-layer perceptron models. Gupta et al. [5] used XGBoost to infer the distributions of travel time, taxi demand, distribution, congestion, and formation length. Moreover, Chang et al., used K-means and other clustering algorithms to predict the demand distribution based on taxi location, time, and weather. In addition, Liu et al., [6] integrated a stochastic forest model with a ridge regression model to forecast taxi demand in high-demand regions.

The technique of deep learning has also been widely used for online car rental demand forecasting. These techniques consist of artificial neural network [7], back propagation (BP) neural networks with strong nonlinear mapping ability [8,9], cyclic (recurrent) neural network (RNN) models with memory ability [10,11], convolutional neural networks with convolution calculation and depth structure [12], and graph neural networks [13], among others. In addition, researchers have begun combining different models to predict the fluctuating online car demand. Accordingly, Yao et al., employed a model that integrates convolutional neural networks (CNN) and long short-term memory (LSTM) to capture the temporal and spatial correlations of taxi data, as well as the similarities between regions with similar time patterns. They conducted experiments on taxi datasets to evaluate the model's performance. Taking into account the taxi demand law under varying time conditions, Peng et al., [14] proposed a taxi demand forecasting model based on the extraction of time series from multiple angles. Furthermore, to improve the accuracy of the online car demand prediction model, Lv et al., [15] proposed a prediction model based on the attention mechanism that utilizes a multi-scale convolutional neural network. They transformed the taxi trajectory into a two-dimensional image and used it as the model's input. Additionally, Liu et al., [16] used feed-forward neural network and con-volution neural network to extract the temporal and spatial characteristics of taxi demand data. They then used two attention mechanisms to capture spatiotemporal information and performed an example verification using the taxi appointment data from Didi Travel Network. Moreover, Dong et al., [17] proposed a model, GAT-Conv-GRU, which combined an attention mechanism-based graph neural network with a door mechanism-based convolutional neural network to predict the demand for network car reservations. The model effectively captured the spatial dependence between regions and demonstrated superior performance on real datasets.

In the past, the area division method in the study of online car rental demand was primarily the traditional grid division, which did not account for the uneven density of online car rental demand in the central area and surrounding areas. In addition, the traditional Conv-LSTM model uses back-propagation algorithm in the optimization of network parameters, which has a high level of complexity and is prone to converge on a local optimal solution. Thus, to address the aforementioned issues, it is necessary to first improve the region division method, use the region division method based on Ty-son polygon, and introduce the sparrow algorithm (SSA) to optimize the Conv-LSTM network car appointment prediction model while avoiding the issue of poor global convergence and slow convergence speed of the algorithm to improve the prediction accuracy.

2. Materials and Methods

2.1. Urban area division based on Tyson polygon

The matrix formed by the statistics of the inflow and outflow of contracted traffic flow in all areas over a given time period is known as the OD flow matrix. When matching the demand data of online car rental, it is necessary to extract the OD (start to end) points from the track data generated during the entire order receiving process, match the extracted OD data with the divided area, and count the number of OD points falling in each area. Thus, the foundation of constructing the OD flow matrix is deciding whether to clean the trajectory data and determining whether the OD point is in the region, followed by counting. Furthermore, the edges of the region must be selected to form a vector using vector cross multiplication, and the points must be tracked in a counterclockwise direction to evaluate whether the sign of the result changes. Herein, the cross-product of the data points in the region and the edge of the region is always positive. The construction process of OD data is shown in Fig. 1.

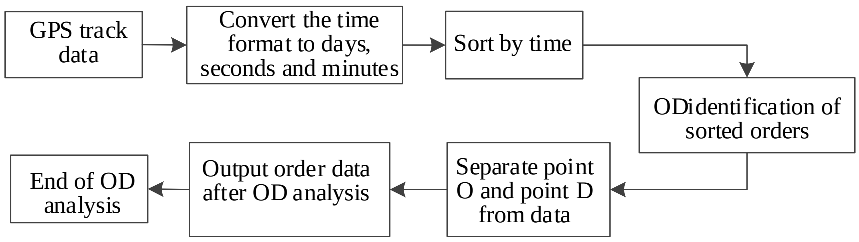


Fig. 1. Flow chart for the origin-destination (OD) data acquisition.

In the past, the majority of area division techniques relied on grid division, which ignored the correlation between the actual road network, resulting in a significant disparity between the data density in the densely populated area and the remote area. Based on an analysis of the city's most visited point of interest (POI), the regional division method of a Voronoi polygon with metro transit stations as nodes is selected. First, based on the actual situation of the city and the distribution of crawled POI interest points, the appropriate regional subway transit stations is selected to prepare for the subsequent construction of the Tyson polygon. On the selected discrete POIs, Delaunay triangulation is then performed. For the extracted Delaunay triangle, the vertical bisector is made on its side, and the intersection of the vertical bisector is connected to create a Tyson polygon. The algorithm for generating Tyson polygons by Delaunay triangulation is as follows:

Step1: Create a list of all selected POI coordinate pairs;

Step2: Traverse the list of POI coordinate points, convert all discrete points to triangles, perform Delaunay triangulation, record all triangles adjacent to each POI point, and save them to a new .txt file;

Step3: Sort each triangle in the Delaunay triangulation counterclockwise. The basis for sorting is that the next triangle must have adjacent edges to the previous triangle;

Step4: Based on the order of the triangles in the Delaunay triangulation, select the discrete point located in the center of the triangulation, connect the center of the circumscribed circle of the triangle adjacent to the discrete point, and form the Tyson polygon. For the discrete points at the edge of the triangulation, select the intersection of the vertical line and the edge of the region to form the Tyson polygon.

2.2. Analysis of spatiotemporal characteristics of online car reservation

The time-space characteristics of online car rental track data are analyzed in order to examine the factors that affect the regional demand distribution of online car rental under different space-time conditions. The following is an example of the November 2016 data analysis of the roundabout track of the second ring road network in Chengdu.

Fig. 2 depicts, over the course of a week, the law governing the variation of residents' demand for online car rental across 24 time periods, one for each hour of the day. Whether during the week or on the weekend, demand is lowest between 0–6:00 h, with the lowest value occurring between 16:00–17:00 h. During the workday, the demand trend of residents for online car reservations is essentially consistent with the morning and evening peak of urban traffic, reaching the morning peak at 8:00–10:00 h and the evening peak at 17:00–18:00 h. However, the demand for online car reservations before and after the peak is still affected by the peak period, demonstrating no abrupt change.

For spatial characteristics analysis, the online car demand with the same week at-tribute (Monday and Saturday for two consecutive weeks) is selected, as depicted in Fig. 3. The region with the highest demand for online car rental is a densely populated, mixed-use commercial and residential area with easy access to public transportation. Consequently, the morning and evening peaks in this type of area have a high proportion of online car rentals, and the spatial distribution characteristics of non-working days and working days are comparable, but the demand for taxi rental on non-working days is obviously lower than that on working days.

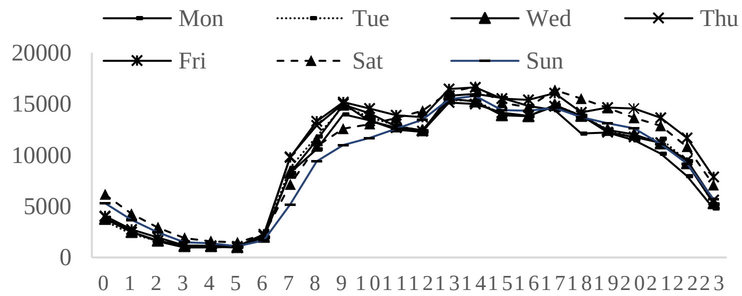


Fig. 2. Line chart of Intranet car-hailing demand over time in a week.

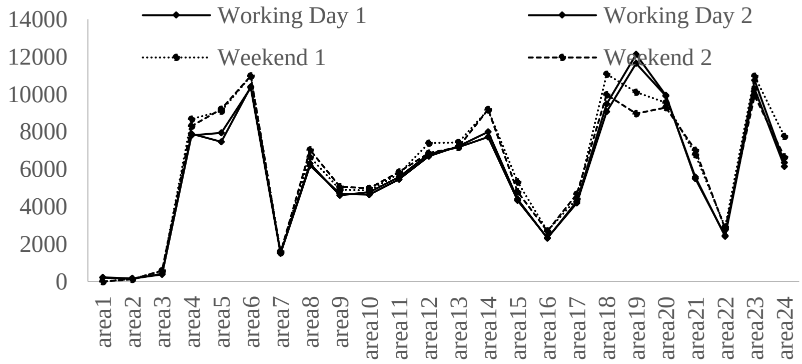


Fig. 3. Line chart of changes in demand for ride-hailing between regions.

2.3. Analysis of influencing factors

According to the analysis of the spatiotemporal distribution characteristics of the demand for online car reservations, it is possible to conclude that the demand for online car reservations exhibits a degree of temporal regularity and will be influenced by the demand in neighboring regions. Spearman correlation [18] coefficient is, therefore, used to analyze the relationship between variables based on the distribution characteristics of online automobile demand. The Spearman correlation coefficient has numerous applications and includes the distribution form and capacity of variables. It is primarily to solve the problem by sorting the position of the source data, thereby analyzing the correlation between the ranks of the two variables. The Spearman correlation coefficient was, therefore, used to analyze the correlation between the online car demand at time t and the subsequent n times, the travel demand at the same historical point in time, and the online car demand in neighboring areas. Accordingly, when a correlation of >0.8 was observed between two variables, a strong relationship between them is inferred. The formula is as follows:

$$\rho_s = 1 - \frac{6\sum_{i=1}^{N} d^2}{N(N-1)} \tag{1}$$

where, d is the difference of rank; N is the number of data in the variable.

Subsequently, the study analyzed the correlation between the demand for online car reservations at time t and the demand for online car reservations in the n preceding time intervals. Accordingly, $M(k,t)$ was set as the demand for online car reservation at the current time slice, $M(k,t-1)$ was set as the demand for online car reservation at time t-1 (the first 15 minutes), and $M(k,t-2)$ was set as the demand for online car reservation at time t-2 (the first 30 minutes). The correlation analysis is shown in Table 1. According to the correlation analysis, the correlation coefficient between the demand for online car rental and the demand for the first six time segments exceeds 0.8, indicating a strong correlation. Therefore, the online car rental demand data for the first six close times were used as the input variable for online car rental demand forecasting.

Table 1

Correlation analysis of the demand for car reservations in the near time network

Variable	$M(k,t)$	$M(k,t-1)$	$M(k,t-2)$	$M(k,t-3)$	$M(k,t-4)$
$M(k,t)$	1	0.9	0.89	0.87	0.86

Subsequently, the data of the same k time segment with the same attributes as the date of the k time segment was selected for correlation analysis. In other words, the correlation between $M(k,t)$ and $M(k-7n,t)(n=1,2,3,4)$ was evaluated. The results of the analysis are shown in Table 2. From the correlation analysis in the table, it can be seen that the demand for online car rental in k time segment and the demand for the same time segment in the first 7 days and the first

14 days exhibit strong county correlation characteristics; therefore, the online car rental demand of the same date attribute and the same segment in the first two weeks was selected as the input of the online car rental demand forecast.

Table 2

Correlation analysis of demand for car reservations in the same period attribute network

Variable	$M(k,t)$	$M(k-7,t)$	$M(k-14,t)$	$M(k-21,t)$
$M(k,t)$	1	0.87	0.84	0.76

With regards to the correlation analysis of the demand for online car reservation in the adjacent regions, 8 adjacent regions to the central region were selected, the demand for online car reservation was counted for every 15 minutes under the same time state for the central region and the eight adjacent regions, and the resultant correlation analysis was conducted. The results of the analysis are shown in Table 3. As evident from the results, the correlation between the central region and the adjacent region is greater than 0.7, indicating a strong correlation. Thus, it can be concluded that the demand for online car reservations and the demand for online car reservations in the adjacent region have an interactive effect.

Table 3

Correlation analysis of demand for car reservations in adjacent regional networks

Variable	$M_1(k,t)$	$M_2(k,t)$	$M_3(k,t)$	$M_4(k,t)$	$M_5(k,t)$	$M_6(k,t)$	$M_7(k,t)$
$M_0(k,t)$	0.84	0.92	0.73	0.80	0.88	0.77	0.95

After analyzing the spatiotemporal characteristics of online car rental demand and identifying influential factors, the study determined the input features for the demand prediction model. These features include the demand for online car rental in the four preceding time intervals, the demand for online car rental during the two preceding weeks with the same date attribute, and the impact of the demand for online car rental in the adjacent time intervals.

3. Forecast model of online car demand

3.1. Conv-LSTM model

Convolution-LSTM neural network (Conv-LSTM) is a recursive neural network that receives as input a second-order matrix sorted by time, recurses in the order direction of time, and links all cycle units in a chain. In addition, the Conv-LSTM model captures the spatiotemporal dependence in the input data. Thus, the fundamental concept of the model is to first extract the spatial features of the data, and then use the extracted spatial features as input to extract the temporal features.

The internal structure of Conv-LSTM is shown in Fig. 4. It uses x_t as input, uses input gate I_t and forgetting gate F_t to update memory unit C_t , uses output gate O_t to update the hidden state H_t, uses sigmoid activation function, and uses δ express.

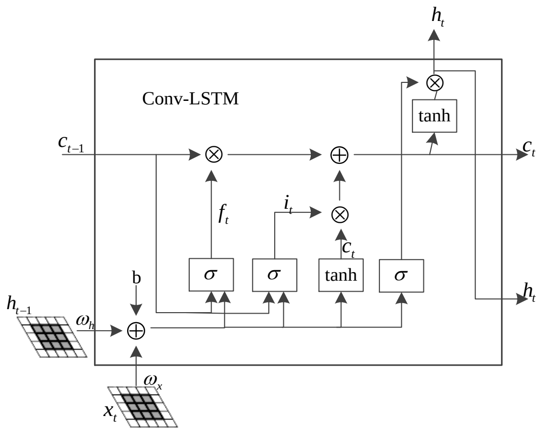


Fig. 4. Schematic diagram of the Conv-LSTM network.

Conv-LSTM first utilizes the memory unit of the previous moment C_{t-1}, the hidden state of the previous moment H_{t-1}, and the input data x_t as inputs through the forgetting gate F_t. The expression of the forgetting door is as follows:

$$F_t = \delta(\omega_{xf} \cdot x_t + \omega_{hf} \cdot H_{t-1} + \omega_{cf} C_{t-1} + b_f) \tag{2}$$

When data flows into the input gate I_t, the input gate also filters the three inputs. The calculation formula is as follows:

$$I_t = \delta(\omega_{xi} \cdot x_t + \omega_{hi} \cdot H_{t-1} + \omega_{ci} C_{t-1} + b_i) \tag{3}$$

The memory unit needs to process the hidden state and the input data flow diagram to iterate the state information. The expression for updating the memory cell $\tilde{C}_t$ is as follows:

$$\tilde{C}_t = \tanh(\omega_{xc} \cdot x_t + \omega_{hc} \cdot H_{t-1} + b_c) \tag{4}$$

The expression of the instantly updated memory unit C_t is as follows:

$$C_t = F_t C_{t-1} + I_t \tilde{C}_t \tag{5}$$

The final output gate O_t output is obtained using the following expression:

$$O_t = \delta(\omega_{xo} \cdot x_t + \omega_{ho} \cdot H_{t-1} + \omega_{co} C_{t-1} + b_o) \tag{6}$$

In addition, the hidden state H_t can be expressed as follows:

$$H_t = O_t\, tanh(C_t) \tag{7}$$

In expressions (2)–(7) for the Conv-LSTM network, the parameters of Conv-LSTM $\omega_{\alpha\beta}(\alpha \in \{x, h, c\}, \beta \in \{f, i, c, o\})$ are the final output of the hidden state.δ is the sigmoid function, and · is the Hadamard product.

3.2. Sparrow optimization algorithm

Derived from the foraging and evading behavior of sparrows, the sparrow search algorithm (SSA) is a swarm intelligence optimization algorithm that boasts high search accuracy, fast convergence speed, and strong stability. According to their ability to find food, sparrows are, therefore, classified as foragers, followers, and guards within the sparrow population. Typically, there is no change in the proportion of foragers and followers in the population. The position of a sparrow population and an individual sparrow is evaluated based on the fitness of the current position, and the optimal individual position and population position are obtained via iterative updating to complete the optimization.

The formula for updating the position of foragers in the population is as follows:

$$X_{i,j}^{t+1} = \begin{cases} X_{i,j} * exp(\dfrac{-i}{\alpha * iter_{max}}), R_2 < ST \\ X_{i,j} + Q * L, R_2 \geq ST \end{cases} \tag{8}$$

In formula (8), t represents the number of iterations, i represents the first sparrow, and j represents the dimension, It is a random number between [0,1], an early warning value between [0,1], and ST is a safe value between [0.5,1]. Q represents a random number subject to a normal distribution, and L represents a total 1 matrix of 1 ow and 1 column.

Follower's position updates the above formula as follows:

$$X_{i,j}^{t+1} = \begin{cases} Q * exp\left(\dfrac{X_{worst}^{t} - X_{i,j}^{t}}{i^2}\right), i > \dfrac{n}{2} \\ X_{p}^{t+1} + \left|X_{i,j}^{t} - X_{p}^{t+1}\right| * A^{+} * L, otherwise \end{cases} \tag{9}$$

In formula (9), $X_{i,j}^{t}$ is the position with the most food, X_{worst}^{t} is the position with the least food, and A^{+} is the matrix with all elements, *n* is the population size.

When the guards are aware of the danger, they carry out anti-predatory behavior, as expressed in the formula (10) below:

$$X_{i,j}^{t+1} = \begin{cases} X_{best}^{t} + \beta * \left|X_{i,j}^{t} - X_{best}^{t}\right|, f_i > f_g \\ X_{i,j}^{t} + K * (\dfrac{X_{i,j}^{t} - X_{best}^{t}}{(f_i - f_w) + \varepsilon}), f_i = f_g \end{cases} \tag{10}$$

where, X_{best}^{t} represents the safest position, β represents the learning rate control parameter, $K \in [-1,1]$ is the sparrow transfer direction parameter, f_i is the sparrow fitness value, f_g is the best fitness value, and f_w is worst fitness value.

3.3. SSA-Conv-LSTM demand forecasting model

Fig. 5 depicts the flow of the SSA-Conv-LSTM-based spatial-temporal demand prediction model for online car rental. The input data first learns the traffic flow characteristics through several residual units before selecting the tensor data under different time steps and conditions to learn the traffic flow space-time characteristics under adjacent conditions and the space-time characteristics under periodic conditions via the Conv-LSTM network. The final prediction results are generated by dynamically fusing the features acquired under two distinct space-time conditions. To initialize the particle dimension in the SSA algorithm, the batch size in the Conv-LSTM demand prediction model, the size of the convolution kernel in the Conv layer, and the number of residual cells in the extracted tensor data are selected as the object of the algorithm optimization. The procedure for model optimization is as follows:

The model operation process is as follows:

Step1: Process GPS data: Obtain the demand flow matrix, which corresponds to the network traffic volume of the t time segment on the d day. The two channels correspond to the inflow and outflow channels, respectively. The input data is first normalized, then passed through residual cells with 16 channels and two convolution layers (convolution core size of 5) to extract traffic characteristics. The extracted traffic characteristics are expressed as $D_d^t \in R^{16\times\hbar\times w}$;

Step2: Process the obtained traffic characteristics: Extract the traffic characteristics of online car rental at close time intervals $E = \left\{D_d^{t-k} \middle| k = n-1, n-2, n \dots, 0\right\}$, and the traffic characteristics of online car rental with periodic characteristics $P = \{D_{d-k}^{t+1} | k = m, m-1, \dots, 1\}$, where m and n are extracted data sequences. The two feature extraction modules in Fig. 5 take E and P as inputs to learn the regular changes of online car rental demand;

Step3: Through the learning of two Conv-LSTM units, E and P are transmitted to the Conv-LSTM unit as inputs, and after the last iteration, the final hidden state $H_E \in R^{16\times\hbar\times w}$ and $H_P \in R^{16\times\hbar\times w}$. The spatiotemporal dependency is extracted as two separate modules;

Step4: Dynamic fusion: Through a full connection layer, H_P learns the weight of the dynamic fusion of the two spatiotemporal dependencies in order to extract the spatiotemporal dependency H_E. Subsequently, using the dynamic fusion of formula (10), the final comprehensive spatiotemporal characteristics of the demand for online car appointment that integrates the proximity characteristics and cycle characteristics $H_{EP}$is obtained;

$$H_{EP} = r \times H_E + (1-r) \times H_P \tag{11}$$

Step5: H_{EP} uses a convolution layer with a convolution core of 1 to obtain the final predicted demand for online car reservations, and finally obtains the predicted value M through the inverse normalization $\hat{M}_d^t$.

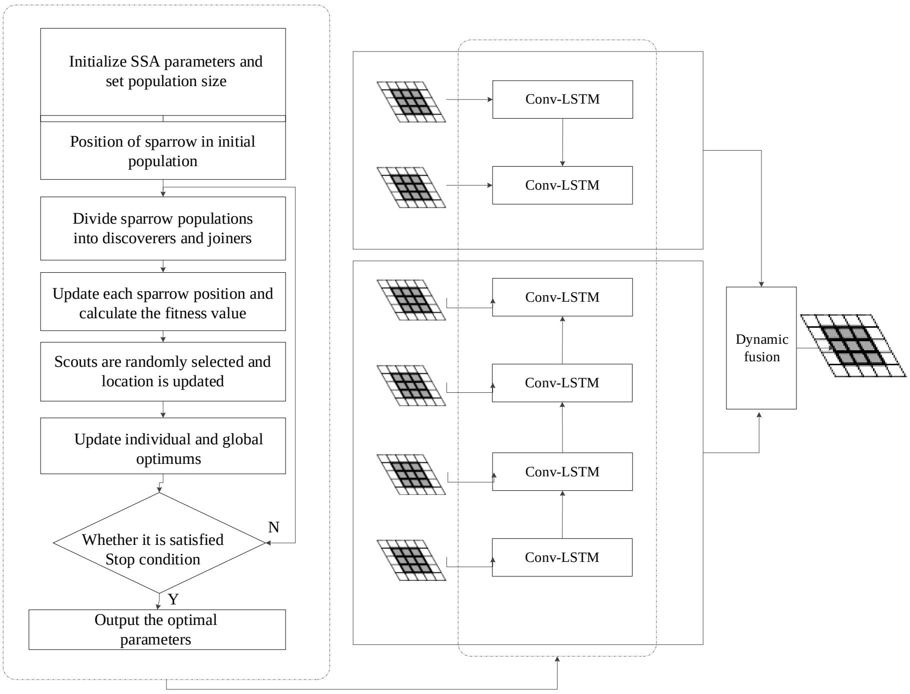


Fig. 5. The overall architecture of SSA-Conv-LSTM.

4. Discussion

4.1. Model training

As experimental data, the complete sample track data from Didi is used. The November 2016 track data from Chengdu's second ring road constitutes the majority of the information. Each vehicle is fitted with a GPS system. The sampling was carried out every 2 – 4 s, to obtain a total of 7.06 million items. After the extraction of OD data from the track data, a total of 300000 OD data pairs were obtained. The GPS sampling included order ID, driver ID, current time stamp, and current vehicle longitude and latitude. The initial data was then partitioned into a 7:3 ratio, with the first 70% serving as the training data set and the remaining 30% as the test data set.

In order to avoid the problem of data integration at different levels, the dimensions of the input data were normalized. The calculation formula is as follows:

$$x' = \frac{x - min}{max - min} \tag{12}$$

To evaluate the accuracy of the predictions, we used two evaluation indicators: Mean Absolute Percentage Error (MAPE) and Root Mean Square Error (RMSE). MAPE refers to the average absolute percentage error, a relative error measure. It employs absolute values to prevent the offsetting of positive and negative errors and is typically used to assess the precision of various time series models. RMSE is the root mean square error, which represents the average deviation value and is used to measure the model's stability. The two evaluation indicators may be stated as follows:

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{Y_i - Y_i^{'}}{Y_i}\right| \tag{13}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(Y_i - Y_i'\right)} \tag{14}$$

where, Y_i is the predicted value and Y_i' is the actual value.

The Python framework was used to write the SSA-Conv-LSTM prediction model. Herein, two Conv-LSTM units and a convolution layer were connected in series to create a spatiotemporal feature extraction unit, which was used to extract spatiotemporal features from the traffic flow map. The first Conv-LSTM unit accepted the tensor data as the input. In addition, the output hidden state and tensor data were used as the in-put to the second Conv-LSTM unit, and the final output hidden state was used to predict the flow map inference. Following experiments utilizing the Sparrow Search Algorithm (SSA), the batch size of the Conv-LSTM demand prediction model, the size of the con-volution kernel in the Conv layer, the extracted input tensor data, and the number of residual units were found to be 48, 3, and 8, respectively. Other prediction model parameters are listed in Table 4.

Fig. 6 depicts the fitness curve of the Conv-LSTM prediction model optimized by the SSA algorithm as a function of the number of iterations.

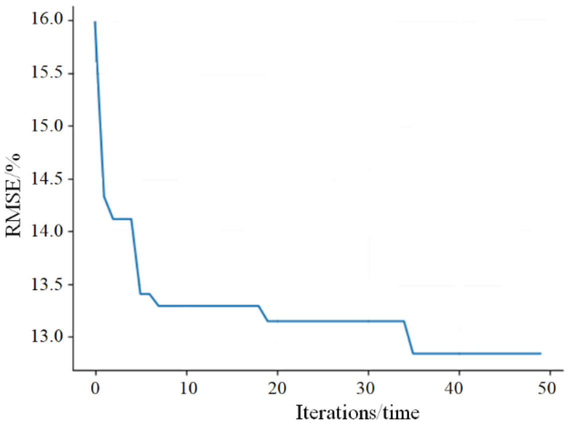


Fig. 6. SSA-Conv-LSTM training network parameters optimization error curve.

Table 4

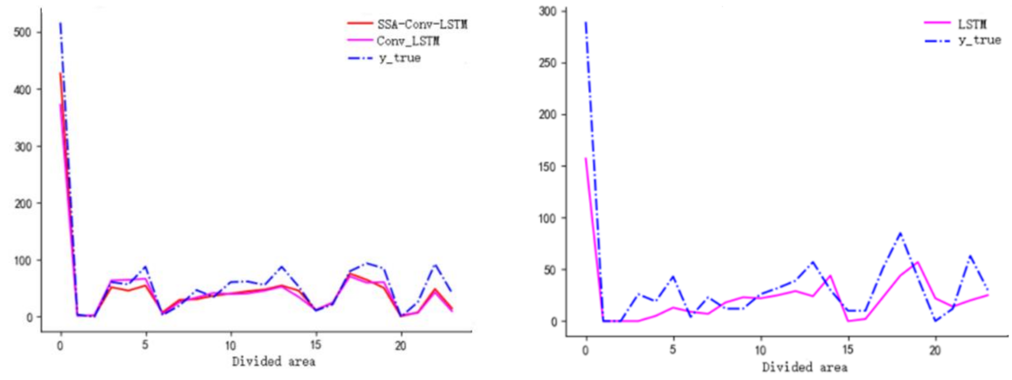
Parameter table of SSA-Conv-LSTM model

Algorithm name	Hyperparameter	Value
SSA algorithm	Population size	20
	Maximum Iterations	50
	Safety value	0.8
	Proportion of foragers	0.2
	Dimensions to be optimized	3
Conv-LSTM based on parameter optimization	Conv-LSTM network input step size of space-time characteristics under periodic conditions	2
	Input step size of Conv-LSTM network with spatiotemporal characteristics under the condition of proximity	4
	Number of LSTM channels	16
	Input data dimension	24*24
	Learning rate	0.003
Conv-LSTM prediction without considering parameters Model	Batch size	64
	Convolution kernel of Conv layer	5
	Extract residual unit of input tensor data	12
LSTM prediction Model	time step	10
	Number of neurons in hidden layer	6
	Input and output characteristics	24

4.2. Comparative analysis of different model results

The SSA-Conv-LSTM network car appointment prediction model was then compared to other prediction models to determine its accuracy. The non-optimized Conv-LSTM prediction model was selected for comparison with the LSTM model that is commonly used for predicting time series. All models predict the demand for online car rental in each region within the next 15 minutes based on the trip data of Chengdu Didi. The specific parameters of the comparison model are shown in Table 4.

One can compare the afore-mentioned model with the SSA-Conv-LSTM prediction model to forecast future demand for online car rental in a specific area of Chengdu. Fig. 7 depicts the comparison between the predicted value and the actual value of the model.



(**a**) (**b**)

Fig. 7. Comparison Chart of Forecast Values of Online Car hailing Demand Forecast Models.(a) Description of the Comparison between the SSA-Conv-LSTM Model and Real Results; (b) Description of Comparing Models with Real Results..

The blue dashed line represents the actual value, the solid line represents the value predicted by various models, the abscissa represents various regions, and the ordinate represents regional demand. From the prediction of each model, the LSTM prediction value was found to be relatively stable, and the prediction performance of the demand at the peak value between regions was found to be poor; in addition, the Conv-LSTM prediction model fit the trend of the high peak and normal value of demand in each region, although, with a certain difference between the predicted value and the actual value. On the basis of meeting the demand trend of online car rental in various regions, SSA-Conv-LSTM can better grasp the demand law of high-demand areas, and the predicted value is closer to the actual value, which can better reflect the space-time change law of online car rental.

In Table 5, the performance evaluation indices for SSA-Conv-LSTM, Conv-LSTM, and LSTM are compared. SSA-Conv-LSTM was found to maintain a low value in the prediction of inflow and outflow, as determined by experimental results. In addition, the root mean square error of the model increased by 0.8% when compared with Conv-LSTM. However, it was significantly superior to the LSTM prediction model, and its predictive effect was found to be substantial.

Table 5

Evaluation form of prediction results

prediction model	RMSE	MAE
SSA-Conv-LSTM	4.8729	0.793
Conv-LSTM	5.6745	1.004
LSTM	47.5987	13.719

5. Conclusion

In this study, we developed a car reservation demand prediction model, named SSA-Conv-LSTM, by utilizing the Conv-LSTM neural network and the SSA algorithm to optimize the neural network parameters. We also conducted a comparative analysis with the Conv-LSTM prediction model and the LSTM prediction model, which were not subjected to parameter optimization. The results of the analysis demonstrate that the stability and prediction accuracy of SSA-Conv-LSTM is significantly superior to those of competing models. In addition, provide data basis for relevant departments to develop reasonable and effective regional network car scheduling schemes. Through the advanced deployment of dispatching vehicle services, it saves fuel costs for online car booking operators and effectively improves the efficiency of green travel. When generating the OD matrix by data processing, the region division method based on the Tyson polygon is used to solve the problem of sparse and uneven data in the previous OD generation matrix and complete the processing of the input data. Regarding the demand forecast for online car reservations, external factors such as the weather and holidays will influence the effect of the forecast. Although, in terms of parameter optimization, the Sparrow algorithm has the disadvantage of slow con-vergence in the later stages, these aspects of the model can be enhanced accordingly.

Funding Statement

This research has been jointly supported by the Science and Technology Research Project of Jilin Education Department (JJKH20230306CY)

R E F E R E N C E S

[1]. *Jamil M S, Akbar S.* Taxi passenger hotspot prediction using automatic ARIMA model// 2017 3rd International Conference on Science in Information Technology (ICSITech). Bandung, Indonesia: IEEE, 2017:23-28.
[2]. *Yan Z, Lv S.* Short-term forecast model of taxi demand based on time and space heterogeneity. Journal of Intelligent & Fuzzy Systems, 2021, 41(2): 4175-4186.
[3]. *Xu R.* Machine learning for real-time demand forecasting. Massachusetts Institute of Technology, 2015.
[4]. *Poongodi M, Malviya M, Kumar C, et al.* New York City taxi trip duration prediction using MLP and XGBoost. International Journal of System Assurance Engineering and Management, 2022: 1-12.
[5]. *Gupta B, Awasthi S, Gupta R, et al.* Taxi travel time prediction using ensemble-based random forest and gradient boosting model//Advances in Big Data and Cloud Computing. Springer Singapore, 2018: 63-78.
[6]. *Liu Z, Chen H, Li Y, et al.* Taxi demand prediction based on a combination forecasting model in hotspots. Journal of Ad-vanced Transportation, 2020, 2020: 1-13.
[7]. *Rajendran S, Srinivas S, Grimshaw T.* Predicting demand for air taxi urban aviation services using machine learning algorithms. Journal of Air Transport Management, 2021, 92: 102043.

[8]. *Guo D, Wang J, Li S.* Research on Short-term Traffic Demand of Taxi in Large Cities Based on BP Neural Network Algorithm//2021 IEEE International Conference on Artificial Intelligence and Computer Applications (ICAICA). IEEE, 2021: 1161-1166.

[9]. *Wang N, Jia S, Liu Q.* A user-based relocation model for one-way electric car sharing system based on micro demand prediction and multi-objective optimization. Journal of Cleaner Production, 2021, 296: 126485.

[10]. *Ku D, Na S, Kim J, et al.* Real-time taxi demand prediction using recurrent neural network//Proceedings of the Institution of Civil Engineers-Municipal Engineer. Thomas Telford Ltd, 2021, 174(2): 75-87.

[11]. *Lai Y, Zhang K, Lin J, et al.* Taxi demand prediction with LSTM-based combination model//2019 IEEE Intl Conf on Parallel & Distributed Processing with Applications, Big Data & Cloud Computing, Sustainable Computing & Communications, Social Computing & Networking (ISPA/BDCloud/SocialCom/SustainCom). IEEE, 2019: 944-950.

[12]. *Lin C Y, Tung S L, Lu P W, et al.* Predictions of taxi demand based on neural network algorithms. International Journal of Intelligent Transportation Systems Research, 2021, 19: 477-495.

[13]. *Huang F, Yi P, Wang J, et al.* A dynamical spatial-temporal graph neural network for traffic demand prediction. Information Sciences, 2022, 594: 286-304.

[14]. *Shu P, Sun Y, Zhao Y, et al.* Spatial-Temporal Taxi Demand Prediction Using LSTM-CNN// 16th International Conference on Automation Science and Engineering. Hong Kong, China: IEEE, 2020: 1226-1230.

[15]. *Lv J,Sun Q ,Li Q ,et al.* Multi-Scale and Multi-Scope Convolutional Neural Networks for Destination Prediction of Trajectories. IEEE Transactions on Intelligent Transportation Systems, 2020, 21(8):3184-3195.

[16]. *Liu Z, Liu Y, Lyu C, et al.* Building personalized transportation model for online taxi-hailing demand prediction. IEEE Transactions on Cybernetics, 2020, 51(9): 4602-4610.

[17]. *Dong C, Zhang K, Wei X, et al.* Spatiotemporal Graph Attention Network modeling for multi-step passenger demand prediction at multi-zone level. Physica A: Statistical Mechanics and its Applications, 2022, 603: 127789.

[18]. *Myers L, Sirois M J.* Spearman correlation coefficients, differences between. Encyclopedia of statistical sciences, 2004, 12.