Trend modeling and multi-step taxi demand prediction

Shan Jiang
School of Computer Science and Engineering
Chongqing Three Gorges University
Wanzhou, Chongqing, 404100
China
jiangshan19920402@qq.com

Yuming Feng*
School of Computer Science and Engineering
Chongqing Three Gorges University
Wanzhou, Chongqing, 404100
China
yumingfeng25928@163.com
ymfeng@sanxiau.edu.cn

Xiaofeng Liao
College of Computer Science
Chongqing University
Chongqing, 404004
China
xfliao@cqu.edu.cn

B. O. Onasanya
Department of Mathematics
University of Ibadan
Ibadan, 234002
Nigeria
bo.onasanya@ui.edu.ng

Abstract. At present, there is a serious mismatch between the supply and demand of taxis, and reasonable demand forecasting can effectively reduce the supply-demand gap, which is an important foundation for taxi scheduling. This article proposes three modeling methods for taxi demand cycle trends, namely the Fourier series based method, the principal component analysis trend based method, and the average trend based method. Finally, based on a weighted combination of three periodic features, a multi-step prediction model for taxi demand was established. On actual data, the method proposed in this paper achieved an MAE error of 1.91, indicating that it can effectively predict taxi multi-step demand. Furthermore, after comparison, the method proposed in this paper outperforms other comparative methods in predicting taxi demand.

Keywords: trend modeling, fourier series, principal component analysis, average trend method, multi-step prediction.
MSC 2020: 68T09

*. Corresponding author

Introduction

The taxi industry provides urban transportation services. But, due to the information lag, there is an imbalance in the supply and demand of taxis in some areas. For example, there are many taxi queues waiting for passengers in areas such as airports, train station, etc., while in some places there are few taxis waiting for service. Therefore, analyze taxi passenger data is of great importance.

Conducting real-time and detailed statistical analysis of the taxi carrying data by the use of big data, cloud computing and artificial intelligence among others, to predict taxi demand will greatly help improve the efficiency of taxi operation [1, 2], which is of great significance for alleviating urban traffic pressure.

Traffic data is essentially a type of time series data, and traffic prediction problems actually belong to a type of time series prediction problem. Therefore, some researchers use basic time series prediction methods, such as exponential smoothing [3], Kalman filter algorithm [4], spectral analysis [5], differential integrated moving average auto regressive model [6], as the basis for traffic prediction.

With the continuous deepening of research on traffic prediction problems, researchers have begun the use of machine learning methods for traffic prediction. The main methods include: *support vector regression* (SVR) [7], *genetic algorithm support vector machine model* (GASVM) [8], *k-nearest neighbor* (KNN) [9]. With the continuous breakthroughs in deep learning technology in tasks such as speech recognition and image recognition, researchers have begun to attempt to use deep learning methods [10, 11, 12] to solve traffic prediction problems. The main methods include: *multi-layer perceptron* (MLP) [12], *deep belief network* (DBN) [13], *convolution neural network* (CNN) [14], *recurrent neural network* (RNN) [15], *long short term memory network* (LSTM) [16], etc.

The popular demand forecast these days for taxis is mainly the short-term forecasting, which predicts the demand for taxis at such a time as 5 minutes, 10 minutes, or 30 minutes. While this article will continue to study the short-term forecasting, to predict the demand for taxis for the whole next day in the sequence of 1 hour, 2 hours, 3 hours, $\cdots$, 24 hours, due to the strong daily periodicity of taxi demand, we will first extract the cycle characteristics of taxi demand based on Fourier series [17], *principal component analysis* (PCA) trend [18] and average trend methods [19]. Finally, based on weighted combinations, a prediction model is established. Based on the cycle characteristics of the past 5 days as input, and the data of the last day as model output, the weighted parameters of the model can be calculated. Finally, this model can be used to predict the demand data for taxis for the next day.

1. Method

1.1 General idea

In the transportation industry, through the observation of the taxi demand curve, we can find that the taxi demand curve of different working days has certain similarity and periodic regularity. This paper will propose three different taxi demand cycle trend calculation methods to predict taxi demand for the next day. These three trends are based on Fourier series, principal component analysis and average trend method. This article will propose a multi-step taxi demand prediction model based on periodic trend weighted combination.

The specific definition of the taxi demand prediction model is as follows

$$d(t) = \theta_1 \times fourier(t\%T) + \theta_2 \times pca(t\%T) + \theta_3 \times at(t\%T) + e(t), \tag{1}$$

where T is the interval numbers of one day, $fourier(t\%T)$ is periodic term of Fourier series, $pca(t\%T)$ is periodic term based on PCA trend, $at(t\%T)$ is average trend period term, $\theta_1, \theta_2, \theta_3$ are weight coefficients of three periodic terms, $d(t)$ is taxi demand value at time t, $e(t)$ is error term. The meaning of this model is to first extract daily cycle trends based on three methods, and then find the weighted sum of the demand values at the same time point of the three daily trends to predict the demand value at a certain time in the future day.

In time series prediction problems, Fourier series method, average trend method, and principal component analysis method have achieved good results in periodic trend modeling. Therefore, this article combines these three methods and integrates them based on ridge regression to predict taxi demand.

1.2 Periodic trend based on Fourier series

In this paper, we will use Fourier transform, a mathematical tool widely used in the field of signal processing, to build a periodic term with time t as the variable, which is used to describe the periodic law of the taxi demand curve [20]. The construction of this cycle item can help us describe the changes in the daily taxi demand curve more accurately and predict the number of taxis that citizens may need in a specific working day. Its specific definition is as follows

$$fourier(t) = a_0 + \sum_{m=1}^{M} a_m \cos(m \times \omega \times t) + b_m \sin(m \times \omega \times t), \tag{2}$$

where a_m and b_m are parameters to be solved. T is the number of time intervals included in a day, or the number of time intervals included in a cycle, M is the order term of the Fourier series, and we usually take this value as 10. $\omega = 2\pi/T$ is the fundamental frequency component of the Fourier series.

The a_m and b_m can be found as follows

$$a_0 = \frac{1}{T} \sum_{t=1}^{T} f(t), \tag{3}$$

$$a_m = \frac{1}{T}\sum_{t=1}^{T} f(t)\cos(m \times \omega \times t), \tag{4}$$

$$b_m = \frac{1}{T}\sum_{t=1}^{T} f(t)\sin(m \times \omega \times t). \tag{5}$$

1.3 Periodic trends based on principal component analysis

In this article, we will explore how to use principal component analysis [21] to extract periodic patterns in demand time series, with the aim of discovering patterns from data and transforming them into actionable information. In this article, we will focus on the application of principal component analysis methods and how to select appropriate periodic parameters to analyze time series data. Firstly, construct a data matrix D so that $d_{i,j}$ represents the taxi demand for the i-th time period on the j-th day. So,

$$D = \begin{bmatrix} d_{1,1} & \cdots & d_{1,N} \\ \cdots & \cdots & \cdots \\ d_{T,1} & \cdots & d_{T,N} \end{bmatrix}. \tag{6}$$

Due to the wide use of PCA algorithm, we can easily find relevant information in the literature. Here, we will not elaborate on the principles and steps of the PCA algorithm. Based on the data matrix D obtained, we can call the PCA algorithm module to reduce its dimensionality. Firstly, normalize the data

$$d_{i,j} = \frac{d_{i,j} - \mu_j}{\sigma_j}, \tag{7}$$

where μ_j is the average value of the data on day j-th, and σ_j is the standard deviation of the data on day j-th. Then calculate the covariance matrix of the data matrix as follows

$$Cov = \frac{1}{N} D^T D. \tag{8}$$

Find all the eigenvalues of the covariance matrix Cov and arrange them from maximum to minimum. Select the eigenvectors corresponding to the first K features and arrange them in rows to form a transformation matrix W. Use the transformation matrix W to reduce the dimensionality of the data. In the process of using PCA algorithm, we can obtain the fractional matrix $S \in R^{T\times N}$ and coefficient matrix $C \in R^{N\times N}$. To obtain the final trend, we can take the first K columns of the matrix S and then the first K rows of the coefficient matrix C. Then, we perform matrix multiplication on these two matrices to obtain the results of the PCA periodic law. The calculation formula is as follows

$$pca = mean(S^K \times C^K). \tag{9}$$

1.4 Periodic trend based on average method

We will propose a modeling method for taxi demand cycle patterns based on the average trend method [22]. The core idea of the average trend method is to calculate the average of demand time series from different days, and ultimately obtain the periodic trend of the time series. Firstly, we construct a data matrix so that $d_{i,j}$ represents the taxi demand for the i-th period on the j-th day. So there is a data matrix

$$D = \begin{bmatrix} d_{1,1} & \cdots & d_{1,N} \\ \cdots & \cdots & \cdots \\ d_{T,1} & \cdots & d_{T,N} \end{bmatrix}. \tag{10}$$

The formula for calculating the periodic pattern using the average trend method is as follows

$$at = [\frac{1}{N}\sum_{n=1}^{N} d_{1,n}, \frac{1}{N}\sum_{n=1}^{N} d_{2,n}, \cdots, \frac{1}{N}\sum_{n=1}^{N} d_{T,n}]. \tag{11}$$

1.5 Working days and holidays

Through long-term data analysis and statistics of the transportation industry, we find that there are obvious differences between the taxi demand curve on weekdays and weekends, which is due to the different travel needs of people on weekdays and weekends. On the other hand, we also find that there are similarities between taxi demand curve in different working days or weekends, because people travel regularly in different working days or weekends.

Therefore, in the calculation process, we divide the data into two parts: one is for weekdays, and the other is for holidays. When calculating the cycle pattern of working days, we only consider the data of working days, and when calculating the cycle pattern of weekends and holidays, we only consider the data of holidays.

1.6 Periodic model integration

We propose a periodic model integration method based on ridge regression [23]. The method of integrating three periodic terms is usually through weighted summation. Therefore, linear regression is used as the ensemble model. Due to the problem of singular matrix values, this paper adopts ridge regression to solve this problem. Therefore, the final integrated model is ridge regression. After calculating the three periodic trends, we need to determine the weights of the three periodic trends. The following data matrix is defined, and the matrix A of the three periodic trend calculation results is defined as follows

$$A = \begin{bmatrix} fourier(1) & pca(1) & at(1) & 1 \\ fourier(2) & pca(2) & at(2) & 1 \\ \cdots & \cdots & \cdots & \cdots \\ fourier(T) & pca(T) & at(T) & 1 \end{bmatrix}. \tag{12}$$

Define the vector d containing demand values as follows

$$d = [d(\tau + 1), d(\tau + 2), \cdots, d(\tau + T)]^T, \tag{13}$$

where τ is the starting date of the training set to predict the demand. The goal is to minimize the error between the weighted curve of three trends and the actual curve as much as possible [23]. The minimization problem is as follows

$$\min_{\theta}(||d - A\theta||^2 + \lambda||\theta||^2). \tag{14}$$

This paper employed the least squares method, obtain the coefficients and optimize the above objective function. The analytical solution for calculating the coefficients is as follows

$$\hat{\theta} = (A^T A + \lambda I)^{-1} A^T d. \tag{15}$$

1.7 Evaluation index

The evaluation index used in this paper is the *mean absolute error* (MAE). The MAE value of area is calculated as follows

$$MAE = \frac{1}{T}\sum_{t=1}^{T} |d_{i,t} - dr_{i,t}|, \tag{16}$$

where $dr_{i,t}$ is the observed demand for taxis in the i-th area on day t, and $d_{i,t}$ is the predicted demand for taxis in the i-th area on day t. In order to measure the relative value of error relative to the time series, we also defined the following indicators to measure the size of demand

$$MAX_i = max\{dr_{i,t}|t = 1, 2, \cdots, T\}, \tag{17}$$

where $dr_{i,t}$ is the observed demand for taxis in the i-th area on time t, and MAX_i is the maximum value of the demand time series.

2. Numerical experiment

2.1 Dataset

The dataset for this article is New York green taxi travel data [24], which covers the period from June 1st, 2017 to June 30th, 2017. There are 265 areas in the data. Table 1 shows the main fields of this data.

Group the above table according to the departure time (time interval of 1 hour) and boarding location ID, and aggregate the number of passengers to obtain our final demand matrix. The Figure 1 shows the transformation rule of the demand curve of 166th area from June 5th, 2017 to June 9th, 2017. It can be seen from the figure that this demand curve has a very strong daily cycle law. There is strong similarity in time series of different days.

Table 1: Field description of the New York green taxi dataset

Field name	Field Meaning
VendorID	A code indicating the LPEP provider that provided the record.
lpep_pickup_datetime	The date and time when the meter was engaged
lpep_dropoff_datetime	The date and time when the meter was disengaged.
Passenger_count	The number of passengers in the vehicle. This is a driver-entered value
Trip_distance	The elapsed trip distance in miles reported by the taximeter
PULocationID	TLC Taxi Zone in which the taximeter was engaged
DOLocationID	TLC Taxi Zone in which the taximeter was disengaged

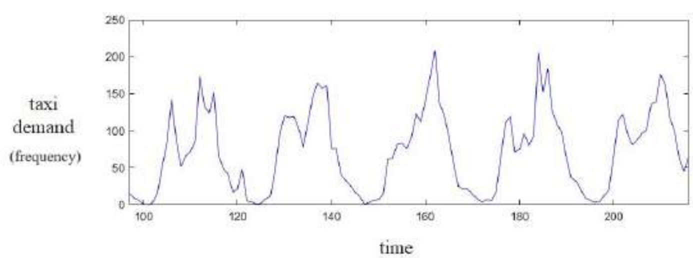

Figure 1: Taxi demand for 166th area from June 5th, 2017 to June 9th, 2017

2.2 Fourier period term

The nonlinear least square method is used to solve the parameters of Fourier series. The input variable is time, and the output variable is demand value. Taking the 166th area as an example, the fitting effect is shown in the figure 2, where the red day curve is the result of Fourier period calculation, and the blue data points are the actual value of taxi demand.

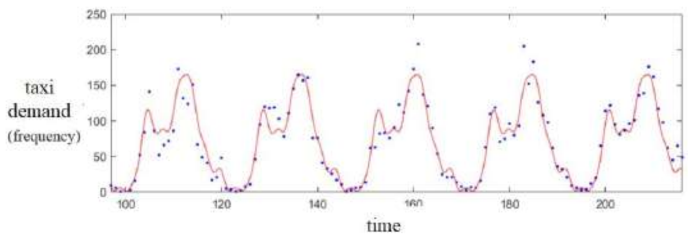

Figure 2: Fitting diagram of Fourier series period component. Taxi demand data is from 166th area, from June 5th, 2017 to June 9th, 2017. The red curve is the fitting result of Fourier series, and the blue data point is the actual taxi demand value

The residual plot of the original time series after removing the fourier period trend is shown as Figure 3, from which it can be seen that the Fourier series method can effectively extract the periodic trend of taxi demand.

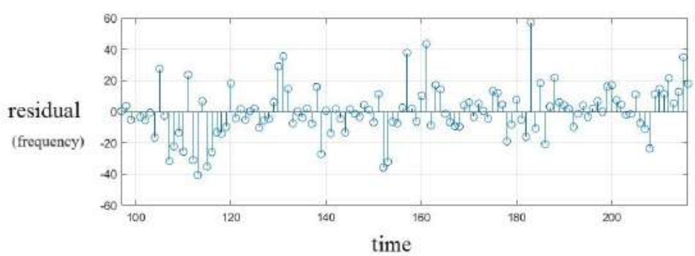

Figure 3: Residual curve after removing the Fourier period term. The taxi demand data comes from the 166th area, from June 5th, 2017 to June 9th, 2017.

2.3 PCA cycle term

The PCA method is used to extract the cycle trend of taxi demand on workdays, and the calculation results are shown in Figure 4. The green curve is the calculation results of cycle items, and the blue data points are the actual data of taxi demand:

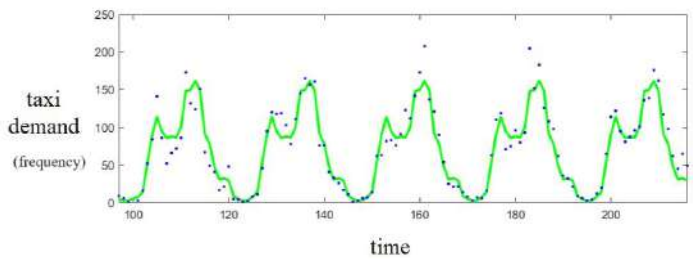

Figure 4: PCA cycle term of taxi demand curve. Taxi demand data comes from 166th area, from June 5th, 2017 to June 9th, 2017. The green curve is the fitting result of Fourier series, and the blue data points are the actual taxi demand values.

The residual plot of the original time series after excluding the PCA cycle trend is in Figure 5. From the graph, it can be seen that the principal component analysis method can effectively extract the periodic trend of taxi demand.

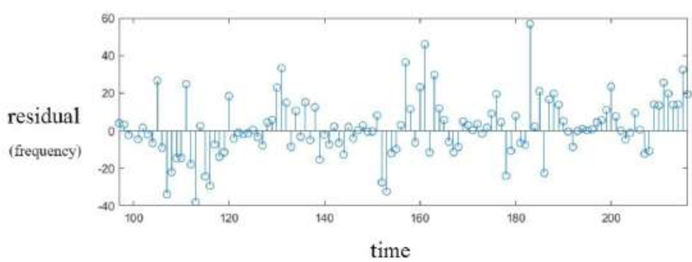

Figure 5: Taxi demand curve residual items after removing PCA cycle items. Taxi demand data comes from 166th area, from June 5th, 2017 to June 9th, 2017

2.4 Average trend method

The average trend method is used to extract the cycle trend of taxi demand on workdays. The calculation results are shown in the Figure 6. The green curve is the calculation results of cycle items, and the blue data points are the actual data of taxi demand.

The residual plot of the original time series after removing the average trend period term is as in Figure 7. From the graph, it can be seen that the average trend method can effectively extract the periodic trend of taxi demand.

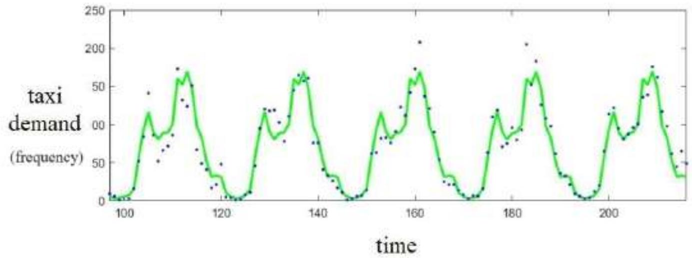

Figure 6: Simple average cycle term of the taxi demand curve. Taxi demand data comes from 166th area, from June 5th, 2017 to June 9th, 2017. The green curve is the fitting result of Fourier series, and the blue data points are the actual taxi demand values.

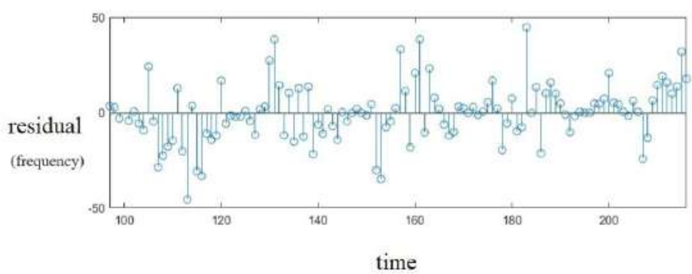

Figure 7: Residual sequence after removing the simple average period component. The taxi demand data comes from 166th area, from June 5th, 2017 to June 9th, 2017.

2.5 Prediction accuracy

We used three methods to calculate the periodic trend term, and then used the least squares method to obtain the weights of each periodic term. Finally, after calculation, we obtained the predicted and observed values on the test set, as shown in Figure 8.

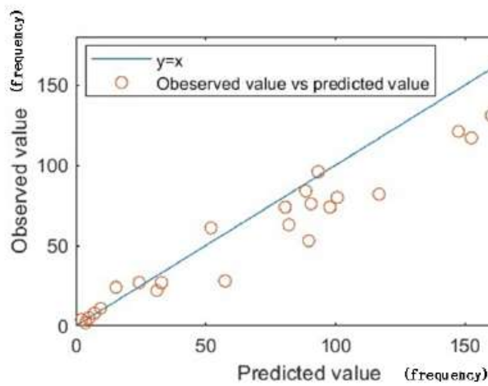

Figure 8: Predicted and true values on the test set. Taxi demand data comes from 166th area.

This paper uses data from June 5th, 2017 to June 9th, 2017 as the training set for trend extraction, uses data from June 12th, 2017 as the weight solution dataset, and uses data from June 13th, 2017 as the test set, periodic trend extraction is performed, and the weight obtained by the least squares method is used to calculate the predicted value of taxi demand. Figure 9 shows the prediction results of our method on the test set. The red curve represents the predicted results, and the blue dots represent the observed values.

In order to provide a more detailed analysis of the accuracy of the prediction algorithm in this article, we conducted experiments in most areas. This paper uses data from June 5th, 2017 to June 9th, 2017 as the trend extraction dataset, and uses data from June 12th, 2017 as the weight solution dataset. This paper uses data from June 13th, 2017 as the test set. Table 2 shows the prediction accuracy of some areas. It can be seen that the method proposed in this article can predict the time series of taxi demand for the next day. It can also be seen that the method proposed in this article achieved an average MAE prediction accuracy of 1.91, with an average of 14.93 for the maximum observed value.

To verify the predictive effect of our method on non-working days, this paper uses data from June 3th, 2017, and June 4th, 2017 as trend extraction datasets, and uses data from June 10th, 2017 as weight solving datasets. This paper uses data from June 11th, 2017 (non-working days) as the test set. Table 3 shows the prediction accuracy of some areas. It can be seen that the method proposed in this article can predict the time series of weekend taxi demand. It can also

Table 2: Prediction accuracy of the first 80 areas for the next day. The data from June 5th, 2017 to June 9th, 2017 was used as the trend extraction dataset, the data from June 12th, 2017 was used as the weight solution dataset, and the data from June 13th, 2017 (working days) was used as the test dataset.

Area id	MAE (frequency)	Max value (frequency)	Area id	MAE (frequency)	Max value (frequency)
1	0.03	0	41	11.52	169
2	0.00	0	42	9.24	101
3	0.67	5	43	5.88	65
4	0.00	0	44	0.00	0
5	0.00	0	45	0.00	0
6	0.06	0	46	0.07	0
7	17.06	154	47	1.39	7
8	0.29	2	48	0.00	0
9	0.12	1	49	5.43	39
10	0.48	2	50	0.00	0
11	0.10	1	51	0.73	4
12	0.00	0	52	5.21	33
13	0.00	0	53	0.62	3
14	1.13	6	54	0.94	7
15	0.08	0	55	3.37	8
16	0.48	2	56	1.80	8
17	5.02	24	57	0.57	3
18	1.56	6	58	0.08	0
19	0.16	1	59	0.02	0
20	1.41	4	60	0.47	2
21	0.25	1	61	5.40	32
22	0.41	5	62	1.47	11
23	0.13	0	63	0.30	2
24	2.63	23	64	0.06	0
25	10.97	82	65	8.36	74
26	0.95	6	66	9.57	93
27	0.00	0	67	0.57	5
28	0.88	4	68	0.00	0
29	0.60	4	69	3.02	13
30	0.00	0	70	1.74	7
31	0.80	4	71	0.88	5
32	0.81	2	72	0.74	5
33	11.11	97	73	0.60	6
34	1.14	5	74	10.84	206
35	0.94	4	75	16.70	213
36	1.73	17	76	1.24	7
37	2.15	18	77	1.01	8
38	0.00	0	78	1.61	5
39	0.75	2	79	0.00	0
40	5.64	25	80	5.60	33

Table 3: Prediction accuracy of non-working days in the first 80 areas. The data from June 3rd, 2017 to June 4th, 2017 was used as the trend extraction dataset, the data from June 10th, 2017 was used as the weight solution dataset, and the data from June 11th, 2017 (non-working days) was used as the test dataset.

Area id	MAE (frequency)	Max value (frequency)	Area id	MAE (frequency)	Max value (frequency)
1	0.09	1	41	30.50	176
2	0.00	0	42	12.10	118
3	0.67	5	43	3.71	30
4	0.00	0	44	0.00	0
5	0.00	0	45	0.07	1
6	0.07	0	46	0.09	0
7	25.70	249	47	1.54	7
8	0.58	7	48	0.00	0
9	0.29	1	49	8.28	63
10	0.13	1	50	0.00	0
11	0.15	1	51	0.84	5
12	0.00	0	52	6.15	45
13	0.00	0	53	0.43	2
14	1.37	7	54	0.79	6
15	0.17	1	55	2.17	12
16	0.42	1	56	1.53	12
17	6.62	57	57	0.70	4
18	2.11	8	58	0.00	0
19	0.19	1	59	0.09	0
20	1.30	7	60	0.58	3
21	0.85	8	61	4.79	50
22	0.51	3	62	2.53	19
23	0.46	6	63	0.42	1
24	3.39	19	64	0.08	1
25	15.57	98	65	8.41	64
26	1.33	6	66	11.17	83
27	0.00	0	67	0.14	1
28	1.54	13	68	0.00	0
29	0.78	6	69	3.41	12
30	0.00	0	70	2.67	11
31	1.83	17	71	0.72	3
32	0.83	3	72	1.04	3
33	11.99	115	73	0.39	2
34	0.67	5	74	20.27	221
35	1.22	11	75	20.65	119
36	4.37	70	76	1.46	8
37	5.49	36	77	0.42	2
38	0.14	1	78	1.34	5
39	1.18	6	79	0.00	0
40	7.42	38	80	6.50	97

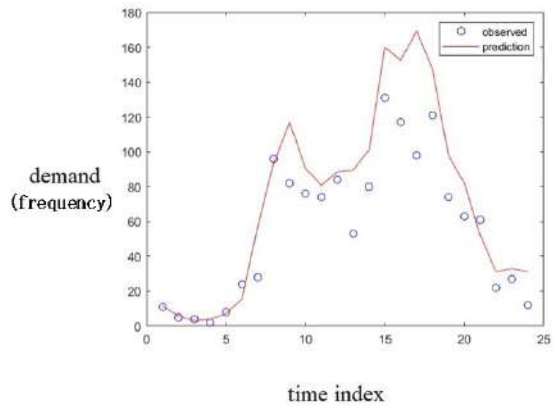

Figure 9: Predicted and actual values for 166th area on the test set as of June 13th, 2017.

be seen that the method proposed in this article achieved an average MAE prediction accuracy of 2.52, with an average of 19.53 for the maximum observed value. For the current taxi demand prediction, the MAE value is around 10% of the maximum value, which is considered a relatively good result [25, 26, 27]. Therefore, it can be seen that the method proposed in this article can effectively predict multi-step demand on both working and non-working days.

2.6 Comparison with other models

In the field of taxi demand prediction, many scholars have conducted research using CNN and RNN [25, 26, 27]. The long-term taxi demand prediction model in this article can also be predicted using CNN and RNN. Due to the differences between the issues in these literature and those in this article, there are certain differences between the comparative model and these methods used in this article. This article compares the proposed method with multi-step prediction models based on CNN [28] and RNN [29]. Figure 10 is the structural diagram of CNN and RNN models.

The CNN network consists of a convolutional layer, a fully connected layer, and a reshape layer. The RNN network consists of a recurrent layer, a fully connected layer, and a reshape layer. The input and output are shown in Figure 11. The input of the CNN network consists of the requirements of all areas in the past 72 steps (each representing one hour), and the output is the requirements of each area in the next 24 steps. The input and output of RNN also take the same form.

At the same time, we will also compare the method proposed in this article with the multi-step prediction model in Facebook's Prophet method [30, 31]. The Prophet model is a universal method for modeling periodic time series. In this article, the prophet model consists of three components. The first item is the growth trend item, the second item is the daily cycle item, and the third item is the weekly cycle item. The definition form of its model composition is as follows

$$y(t) = g(t) + day(t) + week(t) + \epsilon_t, \tag{18}$$

where $y(t)$ is the value of the time series, $g(t)$ is the trend term, $day(t)$ is the daily periodic term, $week(t)$ is the weekly periodic term, and the last term is the error term. These three items are combined through addition. These three parameters are fitted based on historical data and Bayesian methods. After obtaining the parameters, the model can make long-term predictions. Figure 12 shows the fitting results of the prophet model on 166th area. From the graph, it can be seen that the model can extract daily trends and distinguish between working and non-working days. In Figure 12, the minimum value is 60 and the maximum value is 68. That is to say, the demand for taxis remains basically unchanged within a month, and its slope is a random number close to zero. The first term of the Prophet model is a linear term, the second term is a weekly cycle term, and the third term is a daily cycle term. The first item is its mean, the third item has a negative value, and the third item reflects the fluctuation of the demand curve near the mean.

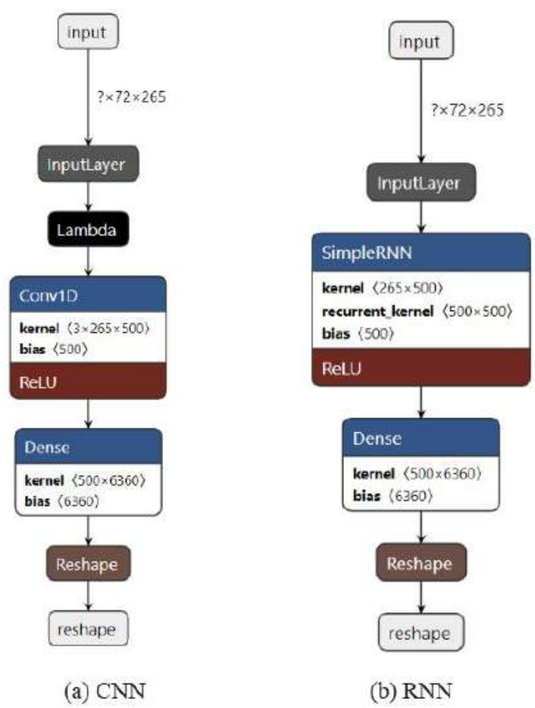

Figure 10: Structure diagram of CNN and RNN models

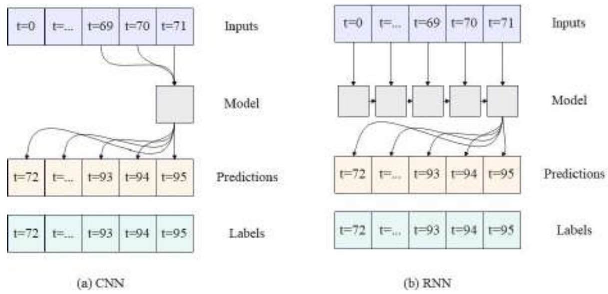

Figure 11: Input and output structure diagram of CNN and RNN models

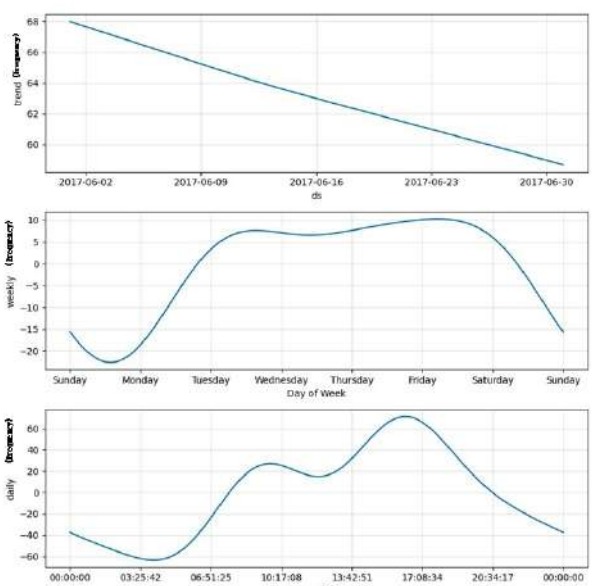

Figure 12: The fitting results of the prophet model on 166th area

This article uses data from June 1st to 29th, 2017 as the training set, and data from June 30th, 2017 as the testing set. We conducted tests on all areas and obtained the average MAE values of different methods across all areas. For convolutional neural networks, the convolutional layer adopts 1-dimensional convolution. The convolutional kernel width is set to 3, and the output size of the convolutional layer is set to 500. The activation function is set to Relu, and the output size of the fully connected layer is set to 24×265. The final output size is set to (24, 265). Set the batch size to 32 and the number of iterations to 300. For recurrent neural networks, the size of the hidden layer is set to 500, and the activation function is Relu. The output size of the fully connected layer is set to 24×265. The final output size of the model is set to (24, 265). Set the batch size to 32 and the number of iterations to 300. For the Prophet model, we set the cycle to daily and weekly, with a time interval of 1 hour, to predict data for the next 24 hours. The growth trend method adopts a linear trend, and the periodic model adopts a trigonometric function. The comparison results are shown in Table 4. The average MAE value of this method in all areas is 2.22, which is relatively smaller than other methods. It can be seen that the method proposed in this article is relatively superior in predicting taxi demand.

Table 4: Average MAE of different methods across all 265 areas.

Method	MAE
The method of this article	2.22
CNN	2.54
RNN	2.31
Prophet	2.56

3. Conclusion

The three taxi demand trend modeling methods proposed in this article can effectively extract the periodic trends of taxi demand time series. These three methods are: Fourier series based method, principal component analysis based method, and average trend based method. This article integrates three trend features for multi-step prediction of taxi demand. The average absolute error of this method reached a prediction accuracy of 1.91 on weekdays and 2.52 on weekends. The method proposed in this article can effectively predict the demand for taxis in the future. Besides, the method proposed in this article outperforms several currently popular methods in predicting taxi demand.

Acknowledgement

This work is supported in part by the Science and Technology Research Program of Chongqing Municipal Education Commission (KJZD-M202201204), in part by the National Natural Science Foundation of China (12201086), and in part by Youth Project of Science and Technology Research Program of Chongqing Municipal Education Commission (KJQN202201209).

References

[1] S. Jiang, W. Chen, Z. Li, H. Yu, *Short-term demand prediction method for online car-hailing services based on a least squares support vector machine,* IEEE Access, 7 (2019), 11882-11891.

[2] B. Zhang, S. Chen, Y. Ma, T. Li, K. Tang, *Analysis on spatiotemporal urban mobility based on online car-hailing data,* Journal of Transport Geography, 82 (2020), 102568.

[3] K.Y. Chan, T.S. Dillon, J. Singh, E. Chang, *Neural-network-based models for short-term traffic flow forecasting using a hybrid exponential smoothing and Levenberg CMarquardt algorithm,* IEEE Transactions on Intelligent Transportation Systems, 13 (2011), 644-654.

[4] T. Zhou, D. Jiang, Z. Lin, G. Han, X. Xu, J. Qin, *Hybrid dual Kalman filtering model for short-term traffic flow forecasting,* IET Intelligent Transport Systems, 13 (2019), 1023-1032.

[5] T.T. Tchrakian, B. Basu, M. O'Mahony, *Real-time traffic flow forecasting using spectral analysis,* IEEE Transactions on Intelligent Transportation Systems, 13 (2011), 519-526.

[6] C.F. Li, J. X. Huang, S.C. Wu, *Prediction traffic flow with combination arima and PageRank,* In Smart Grid and Internet of Things: Third EAI International Conference, SGIoT 2019, TaiChung, Taiwan, December 5-6, 2019 (2020), Proceedings, 129-140.

[7] W. C. Hong, *Traffic flow forecasting by seasonal SVR with chaotic simulated annealing algorithm,* Neurocomputing, 74 (2011), 2096-2107.

[8] X. Chen, *Railway passenger volume forecasting based on support vector machine and genetic algorithm,* In 2009 ETP International Conference on Future Computer and Communication, 2009, 282-284. IEEE.

[9] H. Hong, W. Huang, X. Zhou, S. Du, K. Bian, K. Xie, *Short-term traffic flow forecasting: multi-metric KNN with related station discovery,* In 2015 12th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD), 2015, 1670-1675. IEEE.

[10] G. Hinton, L. Deng, D. Yu, G.E. Dahl, A.R. Mohamed, N. Jaitly, B. Kingsbury, *Deep neural networks for acoustic modeling in speech recognition: the shared views of four research groups,* IEEE Signal Processing Magazine, 29 (2012), 82-97.

[11] D. Li , R. Li, S. Zhang, *A deep learning image recognition framework accelerator based parallel computing,* In Proceedings of the 2018 2nd International Conference on Deep Learning Technologies, (2018), 16-20.

[12] G. Xiaojian, Z. Quan, *A traffic flow forecasting model based on BP neural network,* In 2009 2nd International Conference on Power Electronics and Intelligent Transportation System (PEITS), 3 (2009), 311-314.

[13] W. Huang, G. Song, H. Hong, K. Xie, *Deep architecture for traffic flow prediction: deep belief networks with multitask learning,* IEEE Transactions on Intelligent Transportation Systems, 15 (2014), 2191-2201.

[14] X. Ma, Z. Dai, Z. He, J. Ma, Y. Wang, *Learning traffic as images: a deep convolutional neural network for large-scale transportation network speed prediction,* Sensors, 17 (2017), 818.

[15] R. Fu, Z. Zhang, L. Li, *Using LSTM and GRU neural network methods for traffic flow prediction.* In 2016 31st Youth academic annual conference of Chinese association of automation (YAC), 2016, 324-328. IEEE.

[16] X. Shi, Z. Chen, H. Wang, D.Y. Yeung, W.K. Wong, W.C. Woo, *Convolutional LSTM network: a machine learning approach for precipitation nowcasting,* Advances in Neural Information Processing Systems, 2015, 28.

[17] S. Kohtoh, Y. Taguchi, N. Matsumoto, M. Wada, Z. L. Huang, Y. Urade, *Algorithm for sleep scoring in experimental animals based on fast Fourier transform power spectrum analysis of the electroencephalogram,* Sleep and Biological Rhythms, 6 (2008), 163-171.

[18] L. Li, X. Su, Y. Zhang, Y. Lin, Z. Li, *Trend modeling for traffic time series analysis: an integrated study,* IEEE Transactions on Intelligent Transportation Systems, 16 (2015), 3430-3439.

[19] E. Cascetta, *Transportation systems engineering: theory and methods* (Vol. 49), Springer Science and Business Media, 2015.

[20] T. Omay, D. Baleanu, *Fractional unit-root tests allowing for a fractional frequency flexible Fourier form trend: predictability of Covid-19,* Advances in Difference Equations, 2021 (2021), 167.

[21] Y. Zhang, Y. Wang, *Forecasting crude oil futures market returns: a principal component analysis combination approach,* International Journal of Forecasting, 39 (2023), 659-673.

[22] R. Kyung, L. S Cho, *Analysis of trend in traffic flow using computational and statistical approaches,* In 2019 IEEE 10th Annual Ubiquitous Computing, Electronics and Mobile Communication Conference (UEMCON), (2019, October), 1003-1009.

[23] G. C. McDonald, *Ridge regression. Wiley interdisciplinary reviews,* Computational Statistics, 1 (2009), 93-100.

[24] New York taxi dataset compilation, Zhihu, December 9, 2022, https://zhuanlan.zhihu.com/p/343918278.

[25] P. Shu, Y. Sun, Y. Zhao, G. Xu, *Spatial-temporal taxi demand prediction using LSTM-CNN,* In 2020 IEEE 16th International Conference on Automation Science and Engineering (CASE), (2020, August), 1226-1230.

[26] X. Guo, *Prediction of taxi demand based on CNN-BiLSTM-attention neural network,* In Neural Information Processing: 27th International Conference, ICONIP 2020, Bangkok, Thailand, November 23-27, 2020, Proceedings, Part III, 27 (2020), 331-342.

[27] H. Yao, F. Wu, J. Ke, X. Tang, Y. Jia, S. Lu, Z. Li, *Deep multi-view spatial-temporal network for taxi demand prediction,* In Proceedings of the AAAI Conference on Artificial Intelligence, 32 (2018).

[28] Z. Li , F. Liu, W. Yang, S. Peng, J. Zhou, *A survey of convolutional neural networks: analysis, applications, and prospects,* IEEE Transactions on Neural Networks and Learning Systems, (2021).

[29] S. Li, W. Li, C. Cook, C. Zhu, Y. Gao, *Independently recurrent neural network (indrnn): building a longer and deeper rnn,* In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, 5457-5466.

[30] S. J. Taylor, B. Letham, *Forecasting at scale*, The American Statistician, 72 (2018), 37-45.

[31] M. Navratil, A. Kolkova, *Decomposition and forecasting time series in the business economy using prophet forecasting model,* Central European Business Review, 8 (2019), 26.

Accepted: December 30, 2023