

A bagging ensemble learning traffic demand prediction model based on improved LSTM and transformer

Tianhuai Wang, Yipeng Li, Wenbing Chang, Shenghan Zhou*
School of Reliability and Systems Engineering, Beihang University, Beijing 100191, China

ABSTRACT

The paper aims to build a traffic prediction model for online car-hailing demand with Improved LSTM and Transformer. There are many factors that affect demand, such as temporal features, spatial features, high signal-to-noise ratio, and so on. In this study, LSTM and Transformer are used to extract the temporal and spatial features of data. The temporal and spatial features are used for bagging ensemble learning to predict the online car-hailing orders. An improved LSTM suitable for Traffic data sets is proposed in this paper. In LSTM, we use wavelet decomposition and reconstruction to extract the small-amplitude high-frequency signal from data. The extracted signal is translated and superimposed into the results of the model. The average MAE of 21.24 is obtained on Online Car-hailing Information Data Sets. Compared with other methods, the results suggest the proposed traffic demand prediction model has better accuracy.

Keywords: Traffic demand prediction model, wavelet decomposition and reconstruction, LSTM, transformer, bagging ensemble learning

1. INTRODUCTION

In the 21st century, with the rise of mobile Internet, online car-hailing has become an important part of the smart city. Due to weather, holidays, geographical location, and other factors, orders in different regions are difficult to predict in the business scenario of online car-hailing. At present, traffic demand prediction based on big data analysis has become the main research direction. With the help of big data analysis, features from the data are extracted by machine learning to build the traffic demand prediction model.

Time series prediction methods are used by the traffic demand prediction model, which includes linear models and nonlinear models. Linear models, such as ARIMA, ARIMAX, and so on, are widely used. The linear time series is well fitted by linear models. However, Traffic demand prediction models are usually nonlinear time series. And linear models are ineffective. Therefore, many nonlinear models based on machine learning, such as RNN, LSTM, and so on, are proposed in recent years. Shahid et al.¹ proved that LSTM has better performance as a prediction model. Song et al.² divided LSTM into two subnetworks: temporal attention subnetwork and spatial attention subnetwork. Hu et al.³ proposed AB-ConvLSTM to predict traffic speed. The proposed model is based on LSTM, which includes two Bi-LSTM, a Convolutional LSTM, and an attention mechanism. Tian et al.⁴ used multiscale time smoothing method for missing data inference. And the residual is learned based on LSTM. Zhang et al.⁵ proposed LSTM-XGBoost. It combined LSTM and XGBoost for road short-term traffic flow prediction. Dogan et al.⁶ proved that enlarging the size of data sets increases the performance of LSTM in traffic flow prediction. And a clustering model is proposed to adjust the size of the data sets. Ma et al.⁷ proposed a Convolutional LSTM to predict short-term traffic flow. Mou et al.⁸ proposed an improved LSTM based on temporal information enhancement for traffic flow prediction. Luo et al.⁹ proposed a traffic flow prediction model that combined LSTM and KNN for spatiotemporal prediction. In the proposed model, KNN extracts spatial features of traffic flow and selects neighboring stations. Liu et al.¹⁰ proposed a novel ocean-temperature prediction model. In the proposed model, the matrix is fused based on historical observation data. Wang et al.¹¹ proposed a novel model, including a seven-layer Bi-LSTM and an output layer for respiration motion prediction. Chen et al.¹² proposed an improved LSTM for rainfall-runoff prediction. Trend features of time series are extracted and the nonlinear mapping relationship is identified. Zheng et al.¹³ select the important influencing factors by grey correlation analysis. LSTM is improved by a linear structure for temperature prediction. Chen et al.¹⁴ proved an improved LSTM to predict the deformation mirror voltages.

* zhoush@buaa.edu.cn

Multiple features in the data of Online Car-hailing Information Data Sets include temporal and spatial features. LSTM and Transformer extract the temporal and spatial features of data respectively. The spatial features of the data include weather, temperature, PM2.5, traffic jam information, POI value, holidays, and so on. The temporal and spatial features are used for Ensemble Learning for the online car-hailing orders prediction. In LSTM, the high-frequency signal in the data will affect the feature extraction of the model. We use wavelet decomposition and reconstruction to extract the small-amplitude high-frequency signal in the data, which is similar to the random Gaussian signal. It is not necessary to predict the random Gaussian signal, but only to translate and superimpose the signal. The method of predicting the low-frequency signal and superimposing the high-frequency signal overcomes the problem of low signal-to-noise ratio in data sets. The proposed method is verified in Online Car-hailing Information Data Sets. And the results suggest the proposed prediction model has better performance.

2. ONLINE CAR-HAILING INFORMATION DATA SETS AND DATA PROCESSING

2.1 Data overview

Traffic data sets used in this study are named Online Car-hailing Information Data Sets, which are from the DIDI Taxi GAIA Initiative. Online car-hailing information of a city in January 2016 is recorded by Traffic data sets. Online Car-hailing Information Data Sets include Area Data Sets, Order Data Sets, Weather Data Sets, Traffic Jam Information Data Sets, and POI Data Sets. Area Data Sets contain the area list and area hash value. Order Data Sets records the order data of online car-hailing. Traffic demand is affected by other factors. The weather is recorded in Weather Data Sets. The traffic jam information is recorded in Traffic Jam Information Data Sets. POI value represents the attributes of area, and POI value is recorded in POI Data Sets. Online Car-hailing Information Data Sets contain about 8.5 million data records, and the information contained in each data set is shown in Table 1.

Table 1. The information contained in each data set.

Data sets	Information
Area data sets	Area hash value, Area ID
Order data sets	Order ID, passenger ID, driver ID, order price, order start area hash value, order destination area hash value, time
Weather data sets	Time, weather information, temperature, PM2.5
Traffic jam information data sets	Area hash value, traffic jam information, time
POI data sets	Area hash value, POI value

2.2 Data cleaning

We analyzed the data sets and eliminated invalid data and unreasonable data. For example, some Area hash values on Order Data Sets do not have corresponding area IDs.

2.3 Data preprocessing

For data analysis, the data is normalized by the min-max method as:

$$x'_i = \frac{x_i - \min(x_i)}{\max(x_i) - \min(x_i)} \quad (1)$$

where x'_i is the normalized data, x_i is the original data in the data sets, $\min(*)$ is a function, which returns the minimum value of the input value, and $\max(*)$ is a function, which returns the maximum value of the input value.

In Weather Data Sets, Traffic Jam Information Data Sets, and POI Data Sets, multiple data exist at the same time or in the same area. We take the average of multiple data as:

$$x' = \frac{x_1 + x_2 + \dots + x_n}{n} \quad (2)$$

where x_1, x_2, \dots, x_n is multiple data at the same time or in the same area, n is the size of multiple data, and x' is the average value of multiple data. In Order Data Sets, the order whose driver ID is NULL represents a gap order. We count the size of gap orders at each time, which needs to be predicted. If some data is NULL, we use the average of the values near it.

2.4 Data augmentation

183032 groups of data can be obtained after data cleaning and data preprocessing. To enlarge features of Traffic data sets and improve the generalization ability of the proposed traffic demand prediction model, data augmentation is used. A group of data includes gap, weather information, temperature, PM2.5, traffic jam information, POI value, and holiday. MixUp method is used, and coefficient α and coefficient β of beta function are respectively taken as:

$$\begin{cases} \alpha = 0.5 \\ \beta = 0.5 \end{cases} \quad \begin{cases} \alpha = 0.3 \\ \beta = 0.7 \end{cases} \quad \begin{cases} \alpha = 0.1 \\ \beta = 0.9 \end{cases} \quad (3)$$

After data augmentation, 3666064 groups of data can be obtained.

3. IMPROVED LSTM BASED ON WAVELET ANALYSIS

In the traffic prediction model, the high-frequency signal in the data affects the extraction of data features by LSTM. Wavelet analysis can effectively extract the high-frequency signal in the data and separate the signals with different frequencies. Therefore, we use wavelet analysis to decompose and reconstruct the signal. And the paper proposes a new traffic prediction model based on LSTM.

3.1 LSTM

LSTM¹⁴ is used for time series prediction, which effectively solves the long-term dependencies problem of RNN. LSTM includes an output gate, an input gate, and a forget gate. LSTM specific structure is shown in Figure 1.

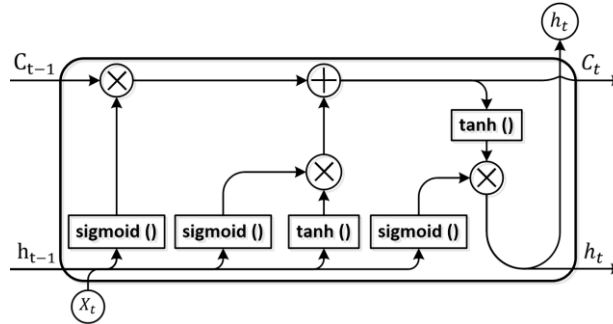


Figure 1. LSTM specific structure.

where h_t is the output value at time t , x_t is the input value at time t , C_t is the state value at time t , $\text{sigmoid}()$ represents the feedforward neural network with a sigmoid activation function, $\text{tanh}()$ represents the feedforward neural network with a tanh activation function.

3.2 Wavelet decomposition and reconstruction

Wavelet analysis¹⁵ is used to effectively extract important information from data. Through the operations of stretching and translation, the signal can be analysed in multi-scale detail, and then the detailed features of the signal can be focused. In recent years, more and more people begin to use wavelet analysis to process time series.

The time series is decomposed by wavelet decomposition, which is divided into two parts: the low-frequency coefficient and the high-frequency coefficient. The wavelet decomposition of time series is as follows:

$$\begin{cases} y_{t+1} = h_0 * y_t \\ z_{t+1} = h_1 * z_t \end{cases} \quad (4)$$

where h_1 is the high-pass decomposition filter, h_0 is the low-pass decomposition filter, z_t is the high-frequency coefficient, y_t is the low-frequency coefficient, and $*$ is the convolution operator.

Time series can be reconstructed by wavelet reconstruction. Wavelet reconstruction can calculate the low-frequency component from the low-frequency coefficients and calculate the high-frequency component from the high-frequency coefficients. The wavelet reconstruction of time series is as follows:

$$\begin{cases} Y_t = g_0 * y_t \\ Z_t = g_1 * z_t \end{cases} \quad (5)$$

where g_1 is the high-pass reconstruction filter, g_0 is the low-pass reconstruction filter, Z_t is the high-frequency signal, and Y_t is the low-frequency signal.

Using wavelet decomposition and reconstruction, the relationship between the low-frequency signal Y_t , the high-frequency signal Z_t and time series X_t is as follows:

$$X_t = Y_t + Z_t \quad (6)$$

3.3 Improved LSTM

The paper improves LSTM by wavelet decomposition and reconstruction. The high-frequency signal in the data sets is extracted by wavelet decomposition and reconstruction. Db10 wavelet base and soft threshold function are used, and the number of wavelet layers is 3. The process of wavelet decomposition and reconstruction is shown in Figure 2.

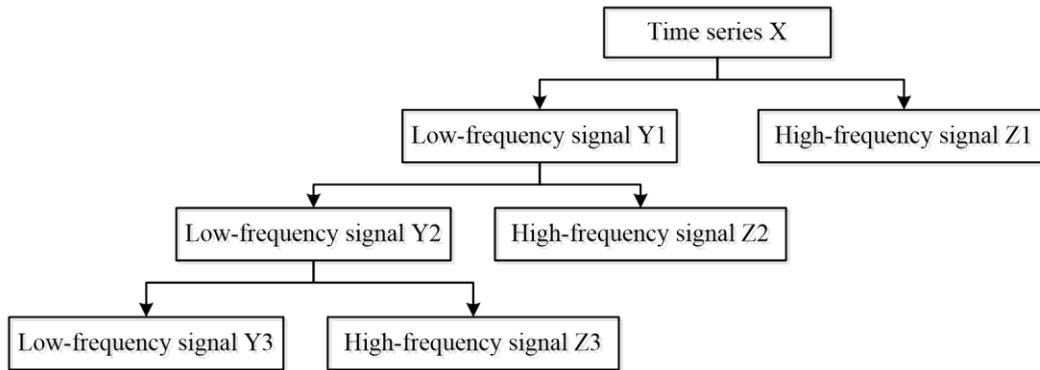


Figure 2. The process of wavelet decomposition and reconstruction.

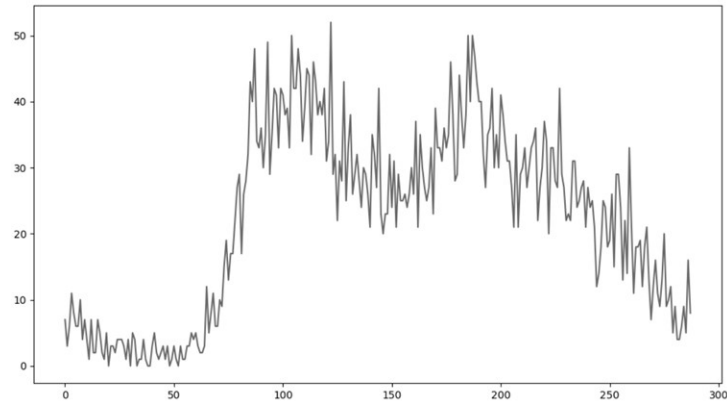
After three-layer wavelet decomposition and reconstruction, the high-frequency signal $Z3$ is obtained. The soft threshold function is used to filter the high-frequency signal $D3$. The soft threshold function is as follows:

$$\overline{w_{j,k}} = \begin{cases} \text{sgn}(w_{j,k})\{|w_{j,k}| - T\lambda\} & |w_{j,k}| > \lambda \\ 0 & |w_{j,k}| < \lambda \end{cases} \quad (7)$$

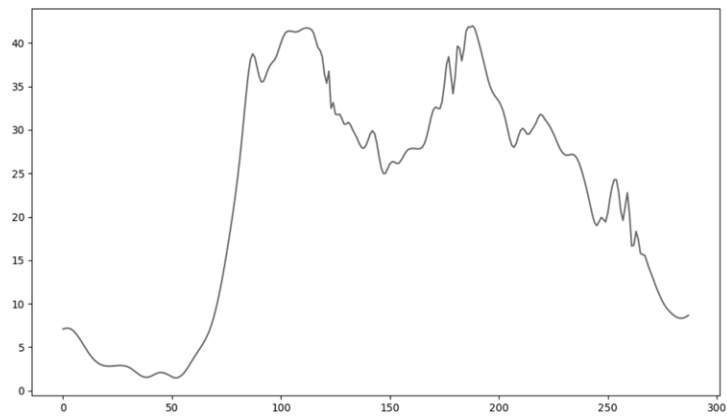
where λ is a threshold, $\overline{w_{j,k}}$ is estimated wavelet coefficients, $w_{j,k}$ is wavelet coefficients after decomposition, T is taken between 0 and 1, and $\text{sgn}(\ast)$ is the symbolic function.

Part of the data is selected as an example. Wavelet decomposition and reconstruction divides the signal into two parts: the low-frequency signal and the high-frequency signal. We believe that the small-amplitude high-frequency signal should not be focused on by the prediction model. Therefore, it is filtered to prevent it from affecting temporal features extracted by LSTM. The time series and the low-frequency signal are shown in Figures 3a and 3b respectively.

The filtered small-amplitude high-frequency signal is shown in Figure 4.



(a)



(b)

Figure 3. Time series and the low-frequency signal.

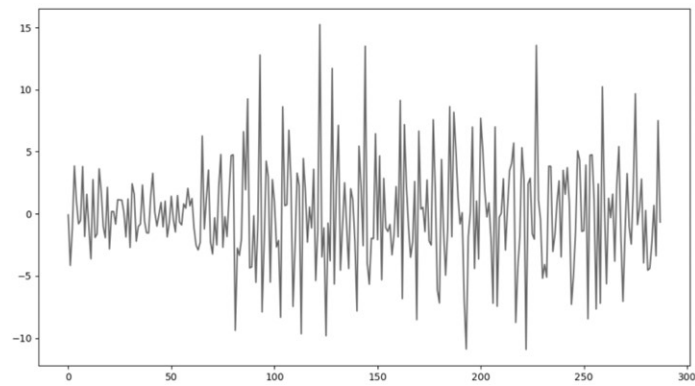


Figure 4. The filtered small-amplitude high-frequency signal.

From Figure 4, the small-amplitude high-frequency signal is similar to the Gaussian signal. There are a few valuable features to extract. And it will affect the extraction of temporal features in signals. In contrast, the low-frequency signal contains almost all the features of the data and directly determines the trend of the predicted signal. Therefore, LSTM should focus on the low-frequency signal, and only need to translate the high-frequency signal and superimpose them on the prediction results. The flow chart of the proposed Improved LSTM is shown in Figure 5.

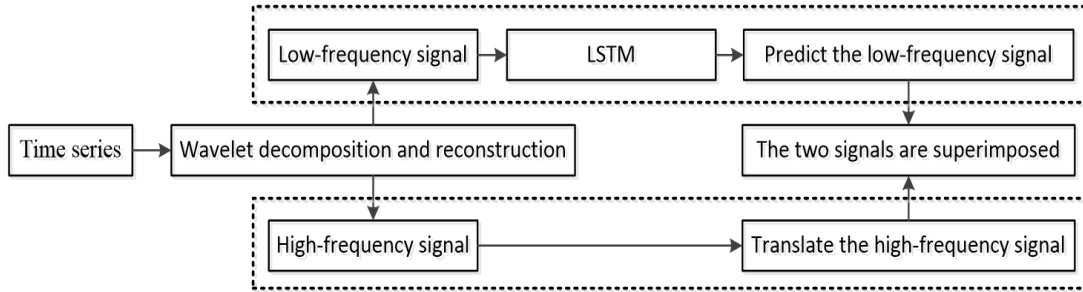


Figure 5. The flow chart of the proposed Improved LSTM.

The steps of Improved LSTM are as follows:

Step 1: The time series is decomposed and reconstructed by wavelet as:

$$X_{input}(t) = Y_{input}(t) + Z_{input}(t) \quad (8)$$

where X_{input} is the time series, which is input into the model, Y_{input} and Z_{input} are the low-frequency signal and the high-frequency signal respectively.

Step 2: LSTM is used to predict the low-frequency signal as:

$$Y_{output}(t+n) = f(Y_{input}(t)) \quad (9)$$

where $f(*)$ is LSTM, n is the length of the prediction signal, and Y_{output} is the prediction signal output by LSTM.

Step 3: The high-frequency signal is translated as:

$$Z_{output}(t+n) = Z_{input}(t) \quad (10)$$

Step 4: The two signals obtained in the previous two steps are superimposed as:

$$X_{output}(t+n) = Y_{output}(t+n) + Z_{output}(t+n) \quad (11)$$

where X_{output} is the prediction series.

The pseudo-code of Improved LSTM is shown in Table 2.

Table 2. The pseudo-code of Improved LSTM.

Algorithm: Improved LSTM
Input: Time series $X_{input} = (X_1, X_2, \dots, X_t)$
Output: Prediction series $X_{output} = (X_{t+1}, X_{t+2}, \dots, X_{t+n})$
The signal is decomposed and reconstructed by wavelet as equation (8).
Train LSTM by X_{input}
for all $X_i \in X_{input}$
LSTM is used to predict the low-frequency signal as equation (9).
Translate the high-frequency signal as equation (10).
The two signals are superimposed as equation (11).
end for
Return X_{output}

4. BAGGING ENSEMBLE LEARNING

Traffic data sets have significant temporal and spatial features. To make the best use of features, the paper proposes a traffic demand prediction method based on Bagging Ensemble Learning. Improved LSTM proposed above is used in the method.

The time series in the data is used, and the time series prediction model is proposed as:

$$x_t = f(x_{t-1}, x_{t-2}, \dots, x_1) \tag{12}$$

where x_t is the gap at time t , and $f(*)$ is the time series prediction model, proposed Improved LSTM can be used. The temporal features of the data are fully extracted by LSTM.

The features of other dimensions in the data are used, and the regression model is proposed as:

$$x_t = h(p_1, p_2, \dots, p_6) \tag{13}$$

where x_t is the gap at time t , and p_1, p_2, \dots, p_6 are weather information, temperature, PM2.5, traffic jam information, POI value, and holiday, $h(*)$ is the regression model, Transformer can be used. The spatial features of the data are fully extracted by Transformer. Online car-hailing demand is related to these factors, and their impact should be considered. Transformer¹⁶ consists of N self-attention modules, and each module includes a multi-head self-attention model and a feedforward network layer. multi-head self-attention model as:

$$V = \text{attention}(Q, \{k_i, v_i\}_M) \tag{14}$$

where Q is a query value, $\{k_i, v_i\}_M$ is the set of a key-value pair, $\text{attention}(*)$ is the multi-head self-attention model, and V is the output.

The above two models have limitations and are difficult to analyze all the information of the data sets. To use information as much as possible, we establish the time series prediction model and the regression model respectively. The two models extract all the features of the data. In order to obtain their advantages and better results, the two models are trained through Bagging Ensemble Learning as:

$$X(t) = \lambda X_{LSTM}(t) + (1 - \lambda) X_{Transformer}(t) \tag{15}$$

where λ is the coefficient, which determines the proportion of the two models, $X_{LSTM}(t)$ is the result of the time series prediction model, $X_{Transformer}(t)$ is the result of the regression model, and $X(t)$ is the result of Bagging Ensemble Learning. The flow chart of the proposed traffic prediction demand model is shown in Figure 6.

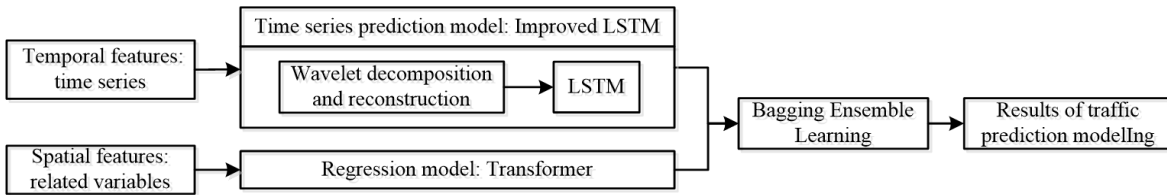


Figure 6. The flow chart of the proposed traffic demand prediction model.

5. EXPERIMENT AND RESULTS

To verify the effectiveness of the proposed traffic demand prediction model, Traffic data sets are verified by experiments. The proposed method is used in Online Car-hailing Information Data Sets. The training of the model has achieved good results. The parameters are set for the Improved LSTM and Transformer are shown in Table 3.

The results of the two models are used for Bagging Ensemble Learning as:

$$X(t) = 0.6X_{LSTM}(t) + 0.4X_{Transformer}(t) \tag{16}$$

The results of the validation set are shown in Figure 6a. The correlation between variables is shown in Figure 6b.

Table 3. The parameters are set for the improved LSTM and Transformer.

	LSTM	Transformer
Epochs	30	100
Batch size	16	16
Loss function	MAE	MAE
Optimizer	Adam	Adam
Learning rate	0.003	0.0002

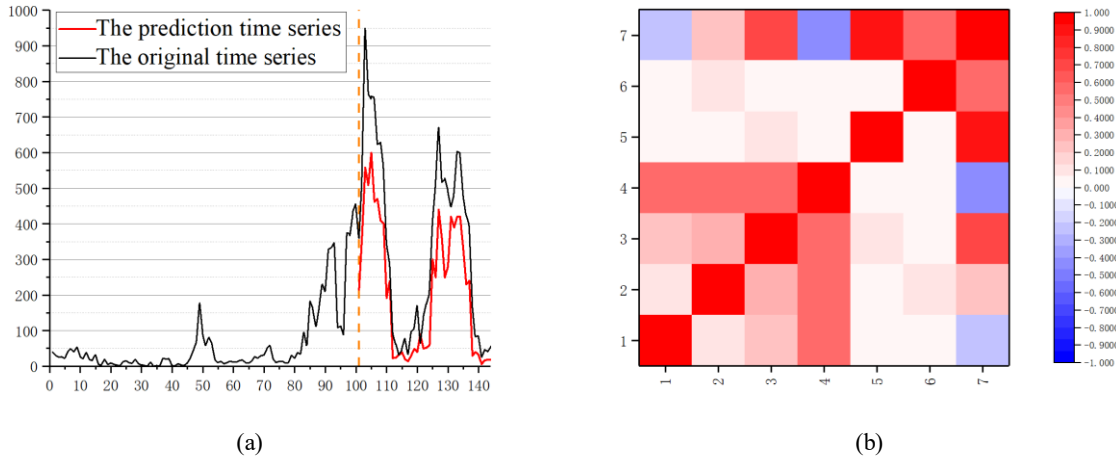


Figure 6. The results of the validation set and the correlation between variables.

where 1, 2, 3, 4, 5, and 6 in Figure 6b represent weather information, temperature, PM2.5, traffic jam information, POI value, and holiday respectively, and 7 in Figure 6b represents gap. From Figure 6b, gap has a strong relationship with PM2.5 and POI value. More attention should be paid to these variables in data analysis.

To verify the effectiveness of the proposed method, we experiment with Online Car-hailing Information Data Sets. A variety of time series prediction models are selected for comparison, including ARIMA, ARIMAX, RNN, and LSTM. MSE is used as the evaluation standard. The verification results on Online Car-hailing Information Data Sets are shown in Table 4.

Table 4. The verification results on Online Car-hailing Information Data Sets.

	Area 1	Area 2	Area 3	...	Area 65	Area 66	Average
ARIMA	28,94	24.39	25.42	...	28.78	26.43	27.45
ARIMAX	26.53	22.89	23.21	...	23.31	26.32	24.51
RNN	23.34	24.52	34.96	...	32.32	25.31	23.52
LSTM	20.43	21.13	21.23	...	22.32	24.53	22.43
The proposed method	19.31	22.34	20.42	...	21.31	22.13	21.24

From Table 4, the proposed method achieves better results in most situations.

6. CONCLUSIONS

In this study, a novel traffic demand prediction model based on improved LSTM and Transformer is proposed. Compared with the other research, this study provides the following innovations:

- Multiple features of data are used, and different models are proposed to extract temporal and spatial features. Bagging Ensemble Learning is used to predict online car-hailing orders.
- An Improved LSTM based on wavelet decomposition and reconstruction is proposed, which is suitable for data sets with the high-frequency signal.

In summary, this study analyzes the temporal and spatial features of Traffic data sets and proposed a novel traffic demand prediction model. The proposed model has better performance.

ACKNOWLEDGMENTS

This research was funded by the National Natural Science Foundation of China (Grant No.71971013) and the Fundamental Research Funds for the Central Universities (YWF-22-L-943). The study was also sponsored by the Graduate Student Education & Development Foundation of Beihang University.

REFERENCES

- [1] Shahid, F., Zameer, A. and Muneeb, M., "Predictions for COVID-19 with deep learning models of LSTM, GRU and Bi-LSTM," *Chaos, Solitons & Fractals*, 140, 110212(2020).
- [2] Song, S., Lan, C., Xing, J., Zeng, W. and Liu, J., "An end-to-end spatio-temporal attention model for human action recognition from skeleton data," *Association for the Advance of Artificial Intelligence*, 1-7(2017).
- [3] Hu, X., Liu, T., Hao, X. and Lin, C., "Attention-based Conv-LSTM and Bi-LSTM networks for large-scale traffic speed prediction," *The Journal of Supercomputing*, 10, 1-24(2022).
- [4] Tian, Y., Zhang, K., Li, J., Lin, X. and Yang, B., "LSTM-based traffic flow prediction with missing data," *Neurocomputing*, 318, 297-305(2018).
- [5] Zhang, X. and Zhang, Q., "Short-term traffic flow prediction based on LSTM-XGBoost combination model," *Computer Modeling in Engineering & Sciences*, 125, 95-109(2020).
- [6] Dogan, E., "LSTM training set analysis and clustering model development for short-term traffic flow prediction," *Neural Computing and Applications*, 33, 11175-11188(2021).
- [7] Ma, Y., Zhang, Z. and Ihler, A., "Multi-lane short-term traffic forecasting with convolutional LSTM network," *IEEE Access*, 8, 34629-34643(2020).
- [8] Mou, L., Zhao, P., Xie, H. and Chen, Y., "T-LSTM: A long short-term memory neural network enhanced by temporal information for traffic flow prediction," *IEEE Access*, 7, 98053-98060(2019).
- [9] Luo, X., Li, D., Yang, Y. and Zhang, S., "Spatiotemporal traffic flow prediction with KNN and LSTM," *Journal of Advanced Transportation*, 1, 537-546(2019).
- [10] Liu, J., Zhang, T., Han, G. and Gou, Y., "TD-LSTM: Temporal dependence-based LSTM networks for marine temperature prediction," *Sensors*, 18, 3797(2018).
- [11] Wang, R., Liang, X., Zhu, X. and Xie, Y., "A feasibility of respiration prediction based on deep Bi-LSTM for real-time tumor tracking," *IEEE Access*, 6, 51262-51268(2018).
- [12] Chen, Y. and Xu, J., "Rainfall-runoff short-term forecasting method based on LSTM," *Journal of Physics: Conference Series*, 012005(2021).
- [13] Zheng, L, Ye, X. and Chen, F., "Research on main steam temperature prediction model based on improved LSTM algorithm," *Journal of Physics: Conference Series*, 012055(2020).
- [14] Chen, Y., "Voltages prediction algorithm based on LSTM recurrent neural network," *Optik-International Journal for Light and Electron Optics*, 220, 164869(2020).
- [15] Lu, J., Hong, L., Dong, Y. and Zhang, Y., "A new wavelet threshold function and denoising application," *Mathematical Problems in Engineering*, 5, 1-8(2016).
- [16] Zheng, S., Lu, J., Zhao, H. and Zhu, X., "Rethinking semantic segmentation from a sequence-to-sequence perspective with transformers," *IEEE Conf. on Computer Vision and Pattern Recognition*, 6881-6890(2021).