CNN-LSTM Based Traffic Prediction Using Spatial-temporal Features

Zhen Zhao¹, Ze Li¹, Fuxin Li² and Yang Liu¹,*

¹Qingdao University of Science and Technology, Qingdao, Shandong, China
²Qingdao Civil Aviation Cares Co., Ltd, Qingdao, Shandong, China

*Corresponding author e-mail: yangliu@qust.edu.cn

Abstract. Aiming at the problem of traffic congestion prediction based on taxi big data, a CNN-LSTM based traffic prediction model using spatial-temporal trajectory topology is proposed. First, the trajectory information is abstracted into a trajectory topology map with spatial-temporal characteristics according to the time and space dimensions. The topology map solves the problem that the road network map does not have stationarity, and extracts a variety of road condition influence factors. Then, the spatial characteristics of the trajectory traffic flow are extracted by CNN, and the temporal characteristics of the trajectory traffic flow are extracted according to the memory characteristics of LSTM. The experimental results show that the RMSE, MAPE and Spearman correlation coefficients of the proposed method on JT-T809-2011 dataset have an absolute improvement of 1%~2% over state-of-the-art strategies.

Keywords: Space Time Division, CNN-LSTM, Trajectory Topology

1. Introduction

With the development of all kinds of sensors and 5G technology, traffic data shows explosive growth, and modern transportation industry is transforming to data-driven intelligent transportation. Taxi is an important way for residents to travel among them. The operating companies behind it excavate the value of taxi data and provide external data services, so as to carry out digital transformation. Therefore, it is very meaningful to use taxi big data for road condition prediction.

In the past a few decades, various methods have been proposed to forecast road conditions. For example, Vythoulkas et al. first proposed using artificial neural network and intelligent recognition system to predict the state of urban road traffic network [1]. Vlahogianni et al. used genetic algorithm to improve and optimize the neural network and solved the nonlinear traffic prediction problem [2].Zhao et al. used wavelet technology to decompose the passenger flow data and forecast them respectively, and then recombined the forecast results to get the final forecast results [3].Li et al. proposed a neural network prediction method based on PLS dynamic pruning algorithm, which realized the prediction of traffic flow [4].Li et al. proposed a multi-scale radial basis function (MSRBF) network model considering the temporal and spatial characteristics of passenger flow, which solved the irregular problem of subway passenger flow prediction in special cases [5]. Yu et al. used the DBN
traffic flow prediction model based on rainfall factor, and proved that the model has better effect than the traditional model [6]. Moretti et al. established the urban traffic flow prediction model by using neural network integrated hybrid modeling [7]. Zheng et al. used the differential integrated moving average autoregressive model to study and predict the short-term road traffic flow [8]. Yan et al. also used the autoregressive differential moving average model to predict the short-term traffic flow of subway [9]. However, the road congestion is depended on various factors, such as the information of associated intersection, holiday information, real-time road speed information and so on. Most of the above methods are based on their own historical data for prediction, without considering a variety of influence factors.

In this paper, a CNN-LSTM model based on spatial-temporal trajectory topology is proposed for traffic prediction. First, input data is transformed into space-time track topology map. On the one hand, the space-time track topology map solves the difficulty that the road network map does not have the stability and does not meet the premise of CNN application. On the other hand, the space-time track topology map preliminarily integrates the influence factors of road congestion extraction. Second, the CNN-LSTM model is constructed to extract the spatial-temporal features. Finally, the information is fused in the fully connected layer, and prediction result is obtained which is a vector with specified format. The model takes into account a variety of factors of road conditions, and the prediction results are more accurate.

2. Proposed Method

The model structure is shown in Figure 1. The network is mainly composed of four components, including topology construction layer, convolution pooling layer, long-term and short-term memory network layer and fully connected layer.

![Figure 1. Overall structure of proposed framework](image)

2.1. Construction of Spatial-Temporal Trajectory Topological Map

Firstly, data preprocessing is carried out to eliminate null values, illogical values and repeated values. According to the data characteristics, the second deep cleaning includes eliminating the values beyond the collection area and the confusion values caused by human factors. Then, the data is divided into 7 * 24 data subsets, corresponding to 24 hours every day from Monday to Sunday. Finally, the data in each data set is mapped to the road network map. Because the road network map does not have the smoothness of the usual picture, each intersection is unique and cannot be shared by different locations. Conventional CNN network structure can not effectively extract valuable features for traffic prediction, so it is difficult to achieve accurate traffic prediction. In order to solve this problem, we transform the road network graph into trajectory topology graph. Firstly, the road network graph is regarded as the backbone of the topology graph, and then each road segment composed of road condition information
is abstracted as an independent node, and they are connected as the turning information to describe the traffic correlation, and finally the corresponding track topology graph is obtained. At the same time, the construction process of spatial-temporal trajectory topological map is also the preliminary feature extraction process of road condition influencing factors, which integrates various factors into spatial-temporal trajectory topological map. Figure 2 shows the topological structure of the trajectory.

![Figure 2. Trajectory topological sketch map](image)

### 2.2. CNN-LSTM Prediction Model

We take the trajectory topological map with time label as the input of CNN-LSTM neural network model. We express the spatial-temporal trajectory topological map as:

$$X = \{x_1, x_2, ..., x_L\}, x_i \in \mathbb{R}^{d_s},$$

$L$ is the time length of the spatial-temporal trajectory topological map, $d_s$ is the dimension of the feature vector of the spatial-temporal trajectory topological map. Since the spatial-temporal trajectory topological map couples various influence factors of road congestion, we extract the spatial and temporal features of the spatial-temporal trajectory topological map. After fusing the characteristics of various influence factors in the full connection layer, the traffic prediction results are output.

#### 2.2.1. Convolution and pooling layer.

The convolution and pooling part of CNN is responsible for further extracting spatial features from the spatial-temporal trajectory topology. CNN uses convolution kernel $m$ with step size of 1 and window size of $3 \times 3$ to extract local context features on a series of spatial-temporal trajectory topological maps, and generates a series of feature maps with time labels, $H^{cnn} = \{h_1^{cnn}, h_2^{cnn}, ..., h_L^{cnn}\}$ among which, $h_i^{cnn} \in \mathbb{R}^{d_{cnn}}$. The $j$-th element in the feature graph is the inner product of the $j$-th filter and the $i$-th window in the spatial-temporal trajectory topology graph.

$$h_{i,j}^{cnn} = f (b_j^{cnn} + \text{win}(x, i, m))$$

(1)

$$< A,B >_F = \sum_i A_{i,j} \sum_j B_{i,j}$$

(2)

$$\text{win}(X, i, m) = [x_{i-\frac{m-1}{2}}; ...; x_i; ...; x_{i+\frac{m-1}{2}}]$$

(3)

Where, $< , >_F$ denotes matrix inner product operation, and $\text{win}(x, i, m)$ denotes the $i$-th window in inner product $X$, and the window size is $m$. We choose ReLU as the activation function. The adjustable parameters of the $j$-th filter are the size and content of the convolution kernel. Note the convolution before zero. Then, taking the output $w$ of convolution as the input, the maxpooling operation is performed. It can effectively reduce the length of the sequence, while retaining important features, which makes the subsequent LSTM network faster. The pooling result is as follows:

$$H_1^{pool} = \left\{ h_1^{pool}, ..., h_L^{pool} \right\}, h_i^{pool} \in \mathbb{R}^{d_{cnn}}$$

(4)
\[ h_i^{pool} = MaxPoolin(win(H^c_{mn}, i \times w - 1, w)) \]  

(5)

2.2.2. LSTM layer. LSTM network extracts the time characteristics of road condition by updating its hidden state, that is, the influence of road condition on current road condition in a certain period of time. It takes the output of convolution and pooling layer as the input, and outputs a series of hidden state matrices. Each hidden state matrix is a series of its previous hidden state matrix and its next hidden state matrix.

2.2.3. Fully connected layer. After the LSTM layer, the fully connected layer gathers the information from the LSTM hidden state matrix and generates a traffic prediction vector with a specific format. It takes the output of LSTM layer as the input, and fully expresses and fuses the spatial-temporal feature information. In order to construct the mapping relationship between fusion information and multi-point road condition, the spatial dimension of information is transformed and integrated by stacking multi-layer fully connected layer, so as to improve the learning ability of the model. Finally, the multi-point traffic forecast value of the specified period is output.

3. Experiments and Analysis

3.1. Data Set Introduction

Table 1 shows the dataset which complies with JT-T809-2011 protocol standard. Ten monitoring points with "upstream and downstream" relationship are selected to study the traffic flow. Because the time span of traffic flow prediction is not a very standard definition, the traffic data is usually divided into 15 minutes as the minimum unit. Therefore, the experimental data were divided into 47453 training samples and 11402 test samples according to 15 minutes.

| Id            | Longitude | Dimension | Speed | Direction | State | Time              |
|---------------|-----------|-----------|-------|-----------|-------|-------------------|
| 040201389608  | 36.577165 | 109.48305 | 27    | 145       | 3     | 2018-07-8 16:00:00 |
| 040201389583  | 36.599608 | 109.46005 | 41.299| 155       | 3     | 2018-07-8 16:00:00 |
| 040201389472  | 36.599441 | 109.476768| 40    | 212       | 3     | 2018-07-8 16:00:00 |

3.2. Evaluation Metrics

Road condition prediction model is to predict the traffic flow of the next time stamp. In order to judge the accuracy of prediction, certain evaluation indexes are needed to express the accuracy of prediction. The evaluation index is defined as the prediction accuracy index of the prediction value. The two indexes are root mean square error (RMSE) and mean absolute percentage error (MAPE). The calculation of RMSE and MAPE are shown as follows:

\[ \text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (\hat{y}_i - y_i)^2} \]  

(6)

\[ \text{MAPE} = \frac{1}{n} \sum_{i=1}^{n} \frac{|\hat{y}_i - y_i|}{y_i} \]  

(7)

When studying traffic problems, there is a certain correlation between the traffic flow of upstream and downstream intersections. This paper analyzes the similarity of traffic flow data at each intersection, and analyzes the correlation of traffic flow at adjacent intersections through Pearson
correlation coefficient of the same magnitude. The calculation formula of Spearman correlation coefficient is as follow:

\[ \rho = 1 - \frac{\delta \sum_{i=1}^{N} d_i^2}{N(N^2-1)} \]  

(8)

3.3. Experimental Setup

The road condition prediction model in this chapter is mainly composed of CNN and LSTM neural network. After extracting the spatial characteristics of passenger flow by CNN, the LSTM network layer is built to extract the temporal characteristics of passenger flow. The dimension of CNN layer is set to 16, step size as 1, LSTM layer neuron number as 32, step size as 8, iterates 500 times, single read neuron number as 128, selects activation function as relu, loss function as cross entropy loss function, optimizes algorithm to adjust Adam's learning rate to 3e-4, and adopts dropout. Considering the great difference of passenger flow distribution between weekdays and rest days, the working days and rest days are modeled separately, and the prediction results are analyzed.

In order to verify the effectiveness of CNN-LSTM model in predicting short-term subway passenger flow, this paper selects the following common prediction methods for comparison, including VAR, HA and ARIMA [10]. The results are shown in Table 2.

Table 2. Comparison results

| Model    | working day |     | Non working days |     |
|----------|-------------|-----|------------------|-----|
|          | MAPE(%)     | RMSE| MAPE(%)          | RMSE|
| VAR      | 28.114      | 27.101| 25.997          | 23.676|
| HA       | 35.217      | 34.102| 30.897          | 30.005|
| ARIMA    | 28.918      | 25.367| 28.138          | 27.881|
| LSTM     | 25.893      | 23.509| 24.750          | 22.163|
| CNN-LSTM | 24.339      | 22.010| 22.993          | 21.667|

As shown in Table 2, CNN-LSTM combination prediction model is the best, and the error values are the smallest, with RMSE value of 22.010 and MAPE value of 24.339 on weekdays, RMSE value of 21.667 and MAPE value of 22.993 on non weekdays. It further proves the superiority of the convolution long-term and short-term memory network CNN-LSTM model in the short-term prediction of passenger flow.

The experimental results show that CNN can capture the dependence between stations in the passenger flow forecasting problem, and LSTM can capture the time dependence in the passenger flow forecasting problem. Building the passenger flow characteristics from the dimensions of time characteristics and spatial symbol network, and establishing the CNN-LSTM combination model are helpful to short-term passenger flow forecasting from the two dimensions of time and space, so as to improve the forecasting effect in this experiment To achieve the best.

Table 3. Spearman coefficient of the intersection traffic flow

|        | Junction a | Junction B | Junction J |
|--------|------------|------------|------------|
| Junction a | 1.000      | 0.935      | 0.677      |
| Junction B | 0.935      | 1.000      | 0.483      |
| Junction J | 0.677      | 0.483      | 1.000      |

The data in Table 3 can be obtained by analyzing the traffic flow of some intersections in the experiment with Spearman coefficient. The mean value of Spearman correlation coefficient between the three intersections in this section is 0.79, which has strong correlation on the whole. It can be seen from Table 2 that the greater the Spearman correlation coefficient between intersections in the same time period, the more likely the traffic state in the same road section at the same time period has higher similarity.
4. Conclusion
In this paper, a CNN-LSTM road condition prediction model based on spatial-temporal trajectory topological map is proposed. By constructing spatial-temporal trajectory topological map, the comprehensive collection of road condition influence factors is realized, and the spatial-temporal trajectory topological map is imported into CNN-LSTM network model to solve the problem of road condition prediction based on taxi data. Through the comparative experiment on JT-T809-2011 dataset, we can see that our model has significantly improved in RMSE, MAPE and Spearman correlation coefficient compared with the other methods.

Acknowledgements
This work is supported by the Natural Science Foundation of Shandong Province (No.ZR2020QF007), Semi Virtual Robot Simulation Training System for IT Specialty (2018MS43) and Training Course Construction of Artificial Intelligence Application Development Based on Robot Teaching Platform (2025QN).

References
[1] Vythoulkas P. Alternative approaches to short term traffic forecasting for use in driver information systems[J]. Transportation and traffic theory, 1993, 12: 485-506.
[2] Vlahogianni E I, Golias J C, Karlaftis M G. Short - term traffic forecasting: Overview of objectives and methods[J]. Transport reviews, 2004, 24(5): 533-557.
[3] Zhao S z, Ni T H, Wang Y, et al. A new approach to the prediction of passenger flow in a transit system[J]. Computers & Mathematics with Applications, 2011, 61(8): 1968-1974.
[4] Li Haitao. Research on prediction of traffic flow based on dynamic fuzzy neural networks[J]. Neural Computing and Applications, 2016, 27(7): 1969-1980.
[5] Li Y, Wang X, Sun S, et al. Forecasting short-term subway passenger flow under special eventsscenarios using multiscale radial basis function networks[J]. Transportation Research PartC: Emerging Technologies, 2017, 77: 306 -328.
[6] Jia Y, Wu J, Xu M. Traffic flow prediction with rainfall impact using a deep learning method [J]. Journal of Advanced Transportation, 2017 (722):1-10.
[7] Moretti F, Pizzi S, Panzieri S, et al. Urban traffic flow forecasting through statistical and neural network bagging ensemble hybrid modeling [J]. Neurocomputing, 2015, 167(C):3-7.
[8] Zhao Zheng, Chen Weihai, Yue Haosong, et al. A novel short-term traffic forecast model based on travel distance estimation and ARIMA[C]//2016 Chinese Control and Decision Conference (CCDC).Northeastern University of China : IEEE Press, 2016: 6270-6275.
[9] YAN D ZHOU J , ZHAO Y, et al. Short-Term Subway Passenger Flow Prediction Based on ARIMA[C]. International Conference on Geo-Spatial Knowledge and Intelligence, 2017.
[10] Hu W, Yan L, Liu K, et al. A short-term traffic flow forecasting method based on the hybrid PSO-SVR [J]. Neural Processing Letters, 2016, 43 (1): 155-172.