The Prediction of Traffic Flow Based on Long Short-Term Memory Network for All Weather

Lin Peiqun1,a, Zhou Chuhao2,b,
1School of Civil Engineering and Transportation South China University of Technology, Guangzhou, Guangdong, China
2School of Civil Engineering and Transportation South China University of Technology, Guangzhou, Guangdong, China
apqlin@scut.edu.cn, bshusoko@163.com

Abstract—To set up a schedule of toll collector in advance and reduce the queue length and traffic congestion of toll stations, this paper proposes a short-term forecasting method for toll stations based on long short-term memory network. First, we analyze the quality of highway toll data. Then, a data preprocessing method based on the cubic smoothing algorithm with five-point approximation is designed. Moreover, we establish the traffic data set which is associated with the toll stations information and time. Then we construct a traffic flow prediction model. Taking the Airport Station of Airport Highway in Guangzhou as an example, then test the validity and real-time performance of our model. The results show that the mean absolute percentage error of the prediction is about 3.6% when the forecast horizon of prediction is 5 minutes; when the forecast horizon of prediction is 10 minutes, the mean absolute percentage error is about 6.07%; when the forecast horizon of prediction is 15 minutes, the mean absolute percentage error is about 8.68%, therefore, the model can accurately predict the traffic flow of toll stations. At the same time, compared with the KNN algorithm and GBDT algorithm, the model of this paper not only has higher prediction accuracy, but also has better adaptability to predict the peak, and when weather is adverse, the algorithm of this paper also can predict accurately through extracting the relevance of time in data set effectively.

1. INTRODUCTION
With the development of our country's economy, the mileage of highways has been increasing year by year. It is the top of the world at the end of 2019. However, the congestion problem of highway toll stations is becoming increasingly serious. Therefore, how to ensure smooth roads has attracted the attention of many scholars. The effective organization of personnel scheduling is the key to solving the problem, and the short-term forecast of highway traffic flow is an important way for scientific scheduling.

For a long time, highway traffic volume prediction has focused on long-term, long-span cycle prediction, of which annual traffic volume prediction is typical. There are many representative studies. For example, Wang Huiyong et al. [1] built a gray linear regression combined model to predict the traffic volume of highways. Their method avoids the local deviation of a single prediction model; Lei Dingyou et al. [2], considering the characteristics of the non-linearity of highway traffic and the influence of multiple factors, used the non-linear principal component analysis to reduce dimensions...
and GA-RBF neural network to predict; Ding Zhikun et al. [3] added economic and transportation factors to the "four-stage method" to improve the accuracy of traffic volume prediction.

In recent years, the intelligent transportation system has been gradually introduced into the management of highways. The highway department has paid more and more attention to the precision of management. As an important component of the intelligent transportation system, short-term traffic flow prediction occupies an important position in the highway traffic management system. Compared with long-term and long-span traffic volume prediction, short-term traffic volume prediction focus on timeliness. There are some researches on short-term prediction which points out the importance of short-term prediction. Yu Bin et al. [4] used the KNN model to predict according to different state vectors. Their research shows that when the state vector using space-time parameters and the distance measurement method combined with exponential weights, the KNN model has higher prediction accuracy. Although the setting of weight reflects the order of the data, the reasonableness of the setting of weight’s size is insufficient, and it may not achieve the best prediction performance; Wang Jian et al. [5] constructed the combined prediction model of the Bayesian network which can improve the accuracy of a single method, but its step size k may not be fixed for different periods, which may increase the complexity of calculation. If a certain model has high accuracy, it may cause a slight decline in overall accuracy due to other models that have low accuracy; Yang Zhaosheng et al. [6] used Kalman filtering to establish a traffic flow prediction model, but the results are relatively volatile and the time-lag phenomenon occurred; Li Li et al. [7] used support vector machine regression to predict traffic volume, which overcomes the defects of neural networks to a certain extent, but this method is difficult to deal with large samples and is slightly insufficient in the context of big data.

Recurrent Neural Network (RNN) is mainly used for the analysis and processing of sequence data, such as text, speech. The long short-term memory network (LSTM) is a special type of RNN. Some studies demonstrate its usefulness. Simoncini et al. [8] used low-frequency GPS data to classify vehicle models, based on the long short-term memory network model. The performance is better than traditional machine learning algorithms, such as support vector machine; Zhou et al. [9] proposed a model named RNNa, which can effectively capture the traffic oscillation and accurately predict the car's following trajectory. Highway traffic flow data is related to time and has periodicity, therefore, LSTM is a good model for traffic prediction.

To solve the problems that traditional methods have and predict the traffic flow accurately even in adverse weather, first, we analyze the characteristics of highway data and design a data preprocessing method based on the cubical smoothing algorithm with five-point approximation, then establish a traffic volume data set based on the information of toll station sites and time. Finally, we build the long short-term memory network for the traffic volume prediction of the toll station. To validate the effectiveness of our model, we compare the performance of KNN and GBDT [10] with our method.

2. TRAFFIC DATA PROCESSING

2.1. Data Cleaning
The toll data of the highway contains lots of information, such as the time for each vehicle to enter and exit, license plate number, driving distance, and vehicle weight. There may be some abnormal information in traffic data. Generally, the proportion of abnormal data is small, but it may cause large errors to the model. Therefore, to count the traffic volume accurately, it is necessary to eliminate abnormal data. In this paper, the rule of thumb is adopted for data cleaning that if a vehicle’s travel time is not within three standard deviations of the mean of all data, we remove it.

2.2. Noise Data Processing
Due to the stochastic disturbance of toll data which may make the model difficult to converge, and the existence of noise may also let the model learn from noisy data, we employ a smoothing algorithm, named cubical smoothing algorithm with five-point approximation, to remove the stochastic disturbance. The formula is as follows:
Where $y_j$ denotes the traffic volume at the time interval $j$, $\overline{y}_j$ denotes the smoothing value of $y_j$. The demand for this algorithm is $n \geq 5$.

3. ALGORITHM

3.1. Long Short-Term Memory

The long short-term memory (LSTM) is proposed by Hochreiter and Schmidhuber [11] in 1997. LSTM is a special network with three "gate" structures. LSTM effectively solves the long-term dependence of standard recurrent neural networks, gradient exploding problem, gradient vanishing problem, and others [12]. The structure is as follows.

First, we use $x^t$ denotes the value that concatenates $h^t$ and $x^t$.

$$x^t = [h^t, x^t]$$  \hspace{1cm} (2)

The equations of the input gate and the forget gate are as follows.

$$i_t = \text{sigmoid}(W_i x^t + b_i)$$  \hspace{1cm} (3)

$$f_t = \text{sigmoid}(W_f x^t + b_f)$$  \hspace{1cm} (4)

$W_i, h_i, W_f, h_f$ denote the weight and the bias of the input gate and the forget gate, respectively and it uses the function of sigmoid as the activation function.

Then update the memory state. The formula is as follows.

$$c_t = f_t c_{t-1} + i_t x^t$$  \hspace{1cm} (5)

Next, filter the information expressed in the current memory state through the output gate. The formula is as follows.

$$o_t = \text{sigmoid}(W_o x^t + b_o)$$  \hspace{1cm} (6)

$$h_t = o_t c_t$$  \hspace{1cm} (7)
3.2. The procedure of model training and prediction

TABLE I. PROCEDURE

| Step | Description |
|------|-------------|
| 1    | Split the total sample set $\mathcal{S}$ into a training set $\mathcal{S}_{\text{train}}$ and a test set $\mathcal{S}_{\text{test}}$; |
| 2    | Normalize $\mathcal{S}_{\text{train}}$ and $\mathcal{S}_{\text{test}}$ to acquire $\mathcal{S}^*_{\text{train}}$ and $\mathcal{S}^*_{\text{test}}$; |
| 3    | Get $m$ samples from $\mathcal{S}^*_{\text{train}}$ and input into the model, then output the predicted value; |
| 4    | Judge whether the number of iterations is less than the maximum number, if less, then Step 5, otherwise Step 8; |
| 5    | Denormalize the predicted value and compare it with the actual value for calculating the mean square error; |
| 6    | Judge whether the error is less than the threshold, if less, go to Step 8, otherwise Step 7; |
| 7    | Error back propagates to correct the parameters of LSTM, then go to Step 3; |
| 8    | Save the parameters of the model; |
| 9    | Read the parameters of the model. Input $\mathcal{S}^*_{\text{test}}$ into the trained model; |
| 10   | Denormalize the predicted value and output. |

Through continuous testing and adjustment, we select $m=64$ as the number of training samples for each round. The maximum number of iterations is set to 50000. We make the learning rate reduce from 0.8 to 0.01 gradually. The error threshold is set to $10^{-7}$.

3.3. Error Evaluation

This paper uses the following four indicators to evaluate the prediction effect. Mean Absolute Error (MAE), Mean Square Error (MSE), Mean Square Percentage Error (MSE), Mean Absolute Percentage Error (MAPE).

4. CASE STUDY

This paper takes the Guangzhou Airport Express Airport Station for the experiment and selects traffic volume data from September 1 to September 30, 2017. The interval is 5 minutes, and there are $30 \times 288$ data points.

The method in this paper is compared to KNN and GBDT in different prediction horizons (5min, 10min, 15min). The results show that LSTM is better than the others.

Next, the performance will be demonstrated in detail. The comparison of three methods when the prediction horizon is 5min is shown below.

![Figure 2. Comparison of prediction results (5 minutes)](image)

It can be observed from the figure that the three methods can predict the trend of traffic volume. However, KNN is not as good as LSTM in predicting inflection points, peaks, etc., while GBDT’s prediction effect is better than KNN, and its prediction of inflection points is more accurate than KNN. GBDT’s judgment on peak value is also slightly inferior to LSTM. Therefore, LSTM can better catch the trend of traffic data and make accurate predictions. The following table is the prediction error.
### TABLE II. COMPARISON OF ERROR

| date   | algorithm | MAE  | MSE  | MSP  | MAPE  |
|--------|-----------|------|------|------|-------|
| 2017/9/28 | KNN       | 14.975 | 21.930 | 0.110 | 0.082 |
|         | GBDT      | 9.338  | 12.101 | 0.087 | 0.061 |
|         | LSTM      | **6.323** | **8.321** | **0.049** | **0.037** |
| 2017/9/29 | KNN       | 12.685 | 15.974 | 0.104 | 0.078 |
|         | GBDT      | 8.630  | 10.863 | 0.079 | 0.054 |
|         | LSTM      | **6.381** | **8.350** | **0.049** | **0.037** |
| 2017/9/30 | KNN       | 13.449 | 17.875 | 0.120 | 0.089 |
|         | GBDT      | 7.523  | 9.506  | 0.074 | 0.053 |
|         | LSTM      | **5.422** | **6.923** | **0.043** | **0.034** |

It can be observed from the table that LSTM is the best from four error indicators. The large MAE value of KNN also shows that it cannot well reflect the changing trend of the data, resulting in large fluctuations in the prediction and the large MAE and MSP values of KNN also reflect the reliability of its prediction is low. The overall error of GBDT is good but slightly worse than LSTM. The errors of LSTM are all at a small level, and 95.9% of the predicted values of LSTM whose absolute percentage errors are less than 10%, and none of them whose absolute percentage errors exceed 20%. However, KNN has 8% of the predicted values whose errors exceed 20%, and the proportion of errors within 10% is only 70%. 3.47% of the predicted values of GBDT whose errors exceed 20% and the proportion of the predicted values whose errors less than 10% is 86.46%.

The following is a comparison of different prediction horizons.

**Figure 3. Comparison of prediction results (10 minutes)**

**Figure 4. Comparison of prediction results (15 minutes)**

As the prediction horizon increases, the prediction accuracy of LSTM decreases. The MAPE value of 10 minutes is 6.07%, and that of 15 minutes is 8.68%, while the prediction error of KNN only changes a little, being 8.82% and 9.1%, respectively. The errors of GBDT are 7.72% and 9.42%. It can be seen that the increase in the prediction horizon causes the LSTM's forecast to fluctuate, resulting in a decline in forecast accuracy, but the overall trend is almost the same. Even if the prediction horizon is 15 minutes, the performance is better than KNN and GBDT. It can be concluded from the error analysis
of KNN that this method is not very sensitive to the change of data. The accuracy of GBDT also has a significant decrease. For the 15 minutes, the performance of GBDT is lower than KNN, which also shows that its stability is insufficient. Therefore, for short-term prediction, LSTM has better accuracy, stability and can reflect data changes in time.

Finally, under adverse weather, the prediction performance of LSTM is tested, and the prediction results are shown below.

![Figure 5. Prediction under adverse weather (5 minutes)](image)

The date is August 27, 2017, with moderate to heavy rain and strong wind on that day. It can be seen from the data that the peak value of that day was only about 220, which was nearly a quarter lower than 300 on September 28-30. The forecast error (MAPE) on that day was 4.75%. Although it was about 1% lower than the error on September 28-30, it also reflects that LSTM can still maintain a good performance under adverse weather.

Through the above analysis, it is shown that LSTM can effectively predict and judge the peaks and inflection points of the traffic volume data, and it can maintain stability for the prediction of different prediction horizons, and it can still maintain high accuracy for different scenarios, such as adverse weather.

5. CONCLUSION

Because of the periodicity of traffic data, we proposed LSTM to predict traffic volume in different prediction horizons and different weather. To improve the network training speed and the generalization ability of the model, we employed the cubical smoothing algorithm with five-point approximation to smooth data. This paper takes Guangzhou Airport Express Airport Station for the experiment to predict the traffic volume, and the results show that the average absolute percentage error is 3.6% when the prediction horizon is 5 minutes, and the prediction of peaks and inflection points is accurate and can be applied to various weather. Compared to KNN and GBDT, LSTM is more time-sensitive and accurate. This paper has a positive significance for improving the toll station service level.

In the future, we will conduct research in the case of multiple data sources, analyze the correlation between different data, mine potential information, and further improve the prediction accuracy under a large prediction horizon.

ACKNOWLEDGMENT

This research has been supported by the National Natural Science Foundation of China (61572233) and the Science and Technology Program Project of Guangdong Province (2016A040403045).

REFERENCES

[1] WANG Hui-yong, YAN Qiu. Forecast of Expressway Traffic Volume Based on the Grey Linear Regression Combined Model [J]. Journal of Transportation Engineering and Information, 2016, 14(01):53-57.

[2] LEI Ding-you, MA Qiang, XU Xin-ping, LIU Qing-yi. Forecasting method of expressway traffic volume based on NPCA and GA-RBF [J]. Journal of Traffic and Transportation Engineering, 2018, 18(03):210-217.
[3] DING Zhikun, ZHU Menglian, SONG Yiyong. Traffic Forecast of Highway Based on Improved “Four-Stage Method” [J]. Journal of Chongqing Jiaotong University (Natural Science), 2017, 36(05):86-90.

[4] B.Yu, S.H. Wu, M.H. Wang, Z.H. Zhao. K-nearest neighbor model of short-term traffic flow forecast [J]. Journal of Traffic and Transportation Engineering, 2012, 12(2): 109-115.

[5] J. Wang, W. Deng, J.B. Zhao. Short-Term Freeway Traffic Flow Prediction Based on Multiple Methods with Bayesian Network [J]. Journal of Transportation Systems Engineering and Information Technology, 2011, 11(04):147-153.

[6] Zhu Zhong, Yang Zhaosheng. A real-time traffic volume prediction model based on the kalman filtering theory [J]. CHINA JOURNAL OF HIGHWAY AND TRANSPORT, 1999, 12(3):63-67.

[7] LI Li, Ren Qi-liang, Luo li. SVMR Model on Short-time Forecasting of City Road Traffic Flow [J]. Communications Standardization, 2006(09):158-161.

[8] Simoncini M, Taccari L, Sambo F, et al. Vehicle classification from low-frequency GPS data with recurrent neural networks[J]. Transportation Research Part C: Emerging Technologies, 2018, 91: 176-191.

[9] Zhou M, Qu X, Li X. A recurrent neural network based microscopic car following model to predict traffic oscillation[J]. Transportation research part C: emerging technologies, 2017, 84: 245-264.

[10] Friedman J H. Greedy function approximation: A gradient boosting machine [J]. Annals of Statistics, 2001, 29(5):1189-1232.

[11] Hochreiter S, Schmidhuber J. Long short-term memory [J]. Neural computation, 1997, 9(8): 1735-1780.

[12] WANG Xiang-xue, XU Lun-hui. Short-term Traffic Flow Prediction Based on Deep Learning [J]. Journal of Transportation Systems Engineering and Information Technology, 2018, 18(1):81-88. M. Young, The Technical Writer's Handbook. Mill Valley, CA: University Science, 1989.