Research on short-term traffic flow prediction model based on RNN-LSTM

Shuying Liu, Li Zhang, Hangfu Li
School of Information Engineering, Anhui Xinhua University, Hefei, Anhui, 230088, China
29064581@qq.com

Abstract. With the rapid development of economy and the increasing number of cars, it is the first task to effectively predict the traffic flow of road traffic to alleviate road congestion and reduce traffic accidents. Aiming at the uncertainty and nonlinearity of traffic flow data, the traffic flow prediction model based on long-term and short-term memory is designed and compared with the traditional BP model and RNN model. The simulation results show that the LSTM model has a higher prediction accuracy and better prediction effect on traffic flow.

1. Introduction
With the rapid development of economy and the gradual improvement of people's living standards, the demand for travel is also growing higher. Therefore, cars are more and more appearing in people's daily life. Although the increase of car ownership brings convenience to people's life, it also increases environmental pollution, road congestion and frequent traffic accidents. Therefore, the most important task of the road management center is to plan the road reasonably, predict the traffic flow accurately and efficiently, and induce the traffic flow in time. Under the current road conditions, intelligent transportation system (ITS)[1] is applied. It is an intelligent traffic management system based on traffic control, guidance and traffic flow prediction algorithm. Therefore, it is of great significance to predict traffic flow in real time and accurately, because it provides timely road conditions for travelers and provides information guarantee for dynamic planning of travel routes.

2. Introduction to traffic flow prediction
According to different time span, traffic flow prediction is generally divided into long-term flow prediction and short-term flow prediction[2]. The long-term flow forecast is the traffic flow based on year, month and day, which mainly provides the basis for road planning and road design. The short-term flow forecast is generally no more than 15 minutes, or even less than 5 minutes, which provides the basis for the traffic control or guidance system to make the next decision.

2.1. Introduction to prediction model
Experts at home and abroad, after analyzing and studying the characteristics of traffic flow, such as strong uncertainty and weak regularity[3], put forward a variety of traffic flow prediction models. According to the different principles of prediction, these models are roughly divided into two categories. One is the data model represented by Kalman filter and ARIMA model. This kind of model is simple in structure and convenient in calculation, but it is greatly affected by the fluctuation of traffic flow, and it can not predict the non-linear and non-stationary data with strong randomness.
Therefore, it is difficult to meet the requirements of short-term traffic flow prediction; The other is the non mathematical model represented by support vector regression machine, non parametric regression model and neural network model, this kind of model has a strong dependence on historical data. As long as there is a large amount of historical data, there is no need to build a large redundant model. According to the nonlinear characteristics of short-term traffic flow, it increases the adaptability to nonlinear[4]. The prediction effect of short-term traffic flow is good.

2.2. RNN model introduction

The model of recurrent neural network (RNN) [5]for processing time series data, The biggest difference between it and the traditional neural network is that the traditional neural network layer is fully connected, and the nodes between each layer are unconnected, however, there are connections between the hidden layer nodes in RNN. In RNN, the hidden layer node not only receives the information transmitted by the input layer at the current time point, but also contains the information transmitted by the hidden layer at the previous time. Therefore, the network constructed by RNN retains the information of historical time points and applies the information left by memory to the output calculation of current neurons. It is a simple chain structure, that is, the same neural network is copied many times to enter the cycle structure.

![RNN unit expansion](image1)

A simple RNN structure is expanded as shown in Figure 1. Where st is the state of step t of the hidden layer, it is the memory unit of the network. It can be seen from the above figure that the hidden layer node at time t is not only affected by the current time XT, but also by the last time reserved ST-1, similarly, the state st retained at time t affects time t+1. Therefore, RNN is mostly used to forecast time series data, such as weather forecast, stock forecast and traffic flow forecast.

Although RNN has memory function for data processing, the time span of information that can be used in practice is limited. When the time span is small, RNN can effectively learn the information of historical time, but when the time span is large, RNN's learning ability will be reduced, That is, the problem of RNN gradient disappearance or gradient explosion. As shown in Figure 2. The memory of t1 is transferred to tn through long journey, RNN adopts the principle of error reverse transfer, error

![Gradient disappearance](image2)
x obtained at TN needs to be multiplied by the parameter W at each step, if W is less than 1, the error may be very small when it is transferred to T1, close to 0, so the error for the initial t1 disappears, the gradient disappears; On the contrary, if W is greater than 1, there will be gradient explosion.

To solve the problem of RNN gradient disappearance, experts and scholars have given a variety of solutions, among them, Hochreiter and Schmidhuber's long-term and short-term memory algorithm proposed in 1997 is the most popular, it can complete the long-term dependence more accurately and has the ability to learn long-term information better[6].

2.3. LSTM model
LSTM long short term memory network is an improved cyclic neural network[7], it overcomes the problem of gradient disappearing in RNN and the problem of long-term dependence which can not be dealt with in RNN.

LSTM evolved on the basis of RNN[8]. In the standard RNN network, the structure of cyclic module is simple, with only one tanh layer, As shown in Figure 3, where a represents the relationship between the input layer and the hidden layer.

![Figure 3. Repeated module chain in standard RNN](image)

LSTM is a special RNN. Its memory unit is mainly composed of three gates: An input gate, an output gate and a forget gate[9]. The three gates symbolize the gate of information. They control the transmission of neuron information. The memory function of the network is realized by these gate nodes. When the gate is opened, the information retained by the front neuron is transmitted to the current neuron, while when the gate is closed, the information of the front neuron will not affect the calculation of the current neuron. Therefore, by adjusting the gate switch, we can realize the influence of historical time series on the later neurons. If the current time information is not expected to affect the calculation of the back neurons, just close the gate. As shown in Figure 4, the input at time t1 affects the output of t4.
An important part of the LSTM is the door, which interacts with each other to control the operation of the module[10]. The opening and closing of the gate determines the retention or deletion of historical information. The basic working principle of the gate is shown in Figure 5. Under the action of activation function, three doors will generate a number between [0,1], which is used to determine the opening and closing of doors.

Figure 5. Basic working principle of LSTM

- the input gate determines how much information of the input XT of the network is retained to the unit state CT at the current time, so as to avoid the current irrelevant content entering the memory module.
- the forget gate determines how much of the cell state CT-1 of the previous time is reserved to the current time CT and what information is discarded.
- the output gate controls how much state CT of the unit outputs to the current output value ht of the LSTM, and controls the effect of long-term memory on the current input.

The calculation formula of forgetting gate is as follows:

$$F_t = \sigma (w_f [H_{t-1}, X_t] + b_f)$$  \hspace{1cm} (1)

The calculation formula of the input door is:

$$I_t = \sigma (w_i [H_{t-1}, X_t] + b_i)$$  \hspace{1cm} (2)

The calculation formula of current input cell state CT is:

$$C_t = F_t * C_{t-1} + I_t * \tilde{C}_t$$  \hspace{1cm} (3)
\[ \tilde{C}_t = \tanh(w_c \cdot [H_{t-1}, X_t] + b_c) \]  

\( \tilde{C}_t \) is the current memory. The current memory \( \tilde{C}_t \) is combined with the long-term memory \( C_{t-1} \) to form a new unit state \( C_t \), which is updated in the memory module without segments.

The calculation formula of the output gate is:
\[ O_t = \sigma(w_o \cdot [H_{t-1}, X_t] + b_o) \]  
\[ H_t = O_t \cdot \tanh(C_t) \]

In the above formula, \( F_t \) is the forgetting gate vector at time \( t \), \( I_t \) is the input gate vector at time \( t \), \( O_t \) is the output gate vector at time \( t \), \( H_t \) is the output vector at time \( t \), and \( X_t \) is the input vector at time \( t \). \( w \) and \( b \) are weights and offsets of different gates.

Two activation functions are used in the formula: sigmoid function and tanh function. The sigmoid function is also known as the S function, it can introduce nonlinear factors into LSTM. When a real value is input to the sigmoid function, it can be compressed to [0,1] output, in which the extremely large positive number is mapped to 1, and the extremely small negative number is supplemented to 0. The formula is as follows:
\[ S(x) = \frac{1}{1 + e^{-x}} \]  

Tanh function is the same as hyperbolic tangent function. It maps the real value of input to the range of [-1,1]. When the input value is 0, the mapping output of tanh function is 0, which conforms to the form of activation function. Its formula is as follows:
\[ \tanh(x) = \frac{\sinh(x)}{\cosh(x)} = \frac{e^x - e^{-x}}{e^x + e^{-x}} \]

3. Verification and simulation

3.1. Data sources
In this paper, the data used in the verification experiment is the measured data of the traffic flow at an intersection in Hefei. Select 5min as the time interval for data collection. Select the data from September 24, 2017 to September 30, 2017 from 8:00 a.m. to 12:00 a.m. to generate Figure 6. From Figure 6, it can be seen that the traffic volume and time distribution in this data set are regular and can be used as the prediction data set of traffic flow.

![Figure 6. One week traffic flow chart](image)

In the validation experiment, from September 18, 2017 to October 11, 2017 was selected as the training data set to train the model, and the data from October 12, 2017 to October 18, 2017 was used as the test set, to verify the validity of the model. For the data in the data set, abnormal data caused by equipment aging, instrument failure, storage failure and other reasons will affect the accuracy of the final prediction, so this experiment adopts the method of first deleting and then adding. For abnormal data, the mean value of the data in the first five days of the same period is used instead.
3.2. Evaluation index

The quality of a prediction model needs to be evaluated quantitatively. The evaluation indexes used in this simulation experiment are MAE, RMSE and MAPE.

(1) MAE——Mean Absolute Error

The formula is:

\[ MAE = \frac{1}{n} \sum_{i=1}^{n} |y_p - y_r| \]  

(9)

(2) RMSE——Root Mean Squae Error

The formula is:

\[ RMSE = \sqrt{\frac{\sum_{i=1}^{n} (y_p - y_r)^2}{n}} \]  

(10)

(3) MAPE——Mean Absolte Percentage Error

The formula is:

\[ MAPE = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{y_p - y_r}{y_r} \right| \times 100\% \]  

(11)

In the above formula, \( n \) is the number of samples, \( y_p \) is predict value, \( y_r \) is true value, The smaller the values of MAE, RMSE and MAPE are, the closer the predicted value is to the real value, the better the prediction effect is.

3.3. Verification results and analysis

In order to verify the effect of LSTM in traffic flow prediction, three models are constructed in this simulation: BP model, RNN model and LSTM model.

According to the verification results of reference 6, the BP model uses three hidden layers, and the number of hidden layer nodes is 11. The main parameter in the LSTM model is the number of cycles of the gates, which changes with the number of gates according to the loss function (MSE), as shown in Figure 7. As can be seen from the figure, when the number of cycles of the door is 400, the value of loss almost no longer changes with the increase of the number of cycles. Therefore, 400 gate control cycle units are selected in the hidden layer of the LSTM model simulation experiment.

![Figure 7. The change of loss function with the number of iterations](image)

After the verification of the three models, the corresponding evaluation index values are obtained,
as shown in Table 1.

| Model | MAE  | RMSE  | MAPE |
|-------|------|-------|------|
| BP    | 9.7372 | 12.4387 | 0.474 |
| RNN   | 8.149 | 10.6517 | 0.3039 |
| LSTM  | 7.2054 | 9.7062 | 0.2126 |

From the data in the table, it can be seen that among the three models, LSTM has the best verification effect, RNN takes the second place, and BP has the worst effect. It can be seen that LSTM model is suitable for traffic flow prediction. The real data comparison of the predicted data of the LSTM model is shown in Figure 8.

![Comparison of LSTM prediction data and real data](image)

**Figure 8. Comparison of LSTM prediction data and real data**

### 4. Conclusion

In this paper, the working principle of LSTM model is introduced in detail, and three models of BP, RNN and LSTM are constructed for simulation verification. After comparison, LSTM model can predict traffic flow more effectively and can be used for traffic flow prediction of urban roads. At the same time, in this experiment, only historical traffic data are used to simulate and predict traffic flow, but the influence of weather, holidays and other factors on traffic flow is not considered. In the later stage, based on the further optimization of the model, more influence factors are added to improve the prediction accuracy.

### Acknowledgments

This paper was financially supported by the quality Engineering Project of Anhui NO. 2018 mooc595 and the quality Project of Anhui Xinhua Universityunder Grant NO.2018 ppzyx03.

### References

[1] Chuntao Man, Danqing Kang. (2019) Considering the short-term traffic flow prediction of the upstream and downstream LSTM. Journal of Harbin University of technology, 5: 101-107.

[2] Chenghong Fu, Shumin Yang. (2019) Short term traffic flow prediction by improved support vector regression. Transportation system engineering and information, 11: 130-134 + 148.

[3] Xiaolei Ma, Zhimin Tao. (2015) Long short-term memory neural network for traffic speed prediction using remote microwave sensor data. Transportation Research, 3:158-162.

[4] Zhicong Han, Yanguo Fan, Huisheng Wu, Huiyan Liu. (2017) Research on short-term traffic flow prediction method based on GA-SVR model. Highway traffic technology, 1: 96-103

[5] Yu Tao. (2018) Short-term traffic flow forecasting and Realization Based on SVM and BP neural network [D]. Nanjing University of Posts and Telecommunications, NanJin.
[6] Hongju Cheng, Zhe Xie. (2019) Data prediction model in wireless sensor networks based on bidirectional LSTM. EURASIP Journal on Wireless Communications and Networking. 1:1-12.

[7] Yi Lin, Jian-wei Zhang, Hong Liu. (2019) Deep learning based short-term air traffic flow prediction considering temporal–spatial correlation. Aerospace Science and Technology, 1:93-97.

[8] Zhen Yan, Zhongzhong Yu, Lu Han, Weijun Su. (2019) Short term traffic flow prediction method based on CNN + LSTM. Computer engineering and design, 9: 2620-2624 + 2659.

[9] Yuelong Li, Dehua Tang, Guiyuan Jiang, Zhitao Xiao. (2019) Short term traffic flow prediction of residual LSTM based on dimension weighting. Computer Engineering, 6: 1-5

[10] Xiangxue Wang, Lunhui Xu. (2018) Short-term traffic flow forecasting based on in-depth learning. Transportation system engineering and information, 18: 81-88.