Traffic Flow State Prediction Based on Deep Learning - Taking Hohhot As an Example

Jiaqi Zhang* Hongye Yang* Shuchen Gao* and Zituo Li*
*Inner Mongolia University of Technology, China

***Abstract*—Traffic flow prediction is an important part of a smart city. With the continuous development of machine learning and artificial intelligence, it has been widely used in the field of traffic engineering. This paper chooses the Gated Recurrent Unit model as the research object based on the bus GPS data of Hohhot. Through the comparative analysis with LSTM model, the results show that GRU model has better prediction performance than LSTM model in traffic state prediction.**

I. INTRODUCTION

With the rapid development of social economy, according to the statistics released by the National Bureau of Statistics in 2018, Reference [1] the possession of civil vehicles in China increased from 6280.61 million in 2009 to 23122 million in 2018, as shown in Fig. 1.The rapid growth of motor vehicles makes public travel more convenient, production and life more efficient. What's more, it not only provides new opportunities for urban economic development, but also brings a series of problems such as traffic congestion, environmental pollution and so on.

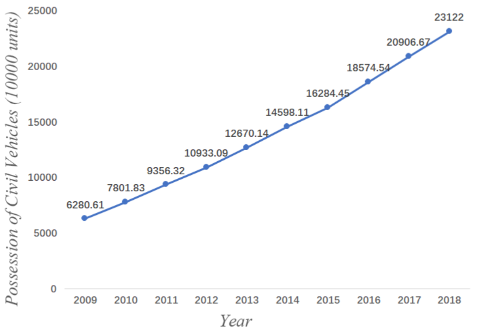


Fig. 1 Possession of civil vehicles (10000 units).

On the one hand, accurate prediction of traffic flow state can provide traffic information for travelers and save their travel time. On the other hand, traffic manager departments can conduct traffic guidance in advance by predicting results to avoid excessive congestion. Traffic flow state prediction refers to the real-time prediction of the traffic state at that time for the next time t+Δt and even at some later times. Because the traffic is a complex system with time-varying and non-linear, it is difficult to predict the traffic flow state. At present, commonly used traffic flow forecasting methods including Reference [2] Han Chao et al. ARIMA model applied to short-term traffic flow prediction shows that the predicted results are good; Reference [3] R.E.Kalman proposes a linear filtering method—traffic flow forecasting method of Kalman filtering method; Reference [4] Fu Gui et al. use support vector machine (SVM) to predict traffic flow; Reference [5] Dochy Thierry et al. use neural network to realize traffic flow forecast, ensuring the high accuracy of the results; Reference [6] Qian Wei et al. use a combination prediction model for traffic flow. Each model has its own adaptive conditions and advantages, therefore, the combined model combines advantages of more than two models, and its performance is better than that of single model. The advantages and disadvantages of each method are shown in Table I.

TABLE I
ADVANTAGES AND DISADVANTAGES OF EACH MODEL

Model	Advantages	Disadvantages
Historical average model	Simple structure and fast	Lower prediction accuracy
Time series model	Mature technology. High accuracy when data is sufficient	Only for single section studies
Kalman model	Can better adapt to non-stationary data	Large amount of calculation and slow operation
Nonparametric regression model	Less affected by abnormal data and higher accuracy	Need a lot of data support
Neural network model	Good effect on nonlinear data processing	Easy to fall into local convergence
Combined model	Complementary advantages	Need to take the appropriate combination method

From the above, the prediction of traffic flow state was and still is the focus of scholars' research. In recent years, with the rapid development of machine learning and artificial intelligence, new ideas and great opportunities have been brought to the field of traffic flow prediction, and its application in urban traffic guidance, network optimization and improvement of urban traffic congestion has become the focus of current research. This paper, based on the neural network structure of machine learning, taking Hohhot as an example, introduces the Gated Recurrent Unit model in detail, and compares it with LSTM model. Then, it studies and discusses the applicability of this method in the field of traffic flow state prediction, and provides reference for the traffic flow state prediction of cities in northern and northwestern border areas of China.

II. The Gated Recurrent Unit Model

In the traditional neural network model, the nodes in the input layer, the hidden layer and the output layer are connected with each other, but there is no connection between the nodes in the same layer. The traditional neural network model is shown in Fig. 2.

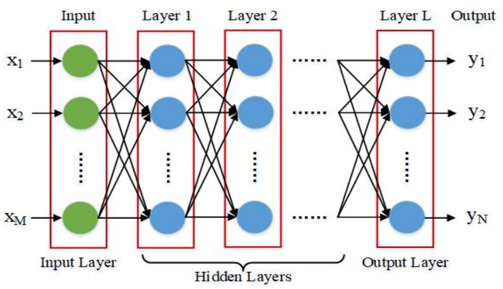


Fig.2 Traditional neural network model.

The circles in Fig.2 represent the nodes of the neural network, and the arrows represent the flow direction of information. According to Fig.2 — the traditional neural network model, information can only be transmitted between different layers, and the nodes of the same layer are independent of each other.

Differing from the traditional neural network model, the cyclic unit network is an improved multi-layer perceptron network which has the concept of time sequence. The output of the hidden layer not only enters the output end of the current moment, but also enters the hidden layer of the next moment as feedback information. The network model structure of deep loop element is shown in Fig.3.

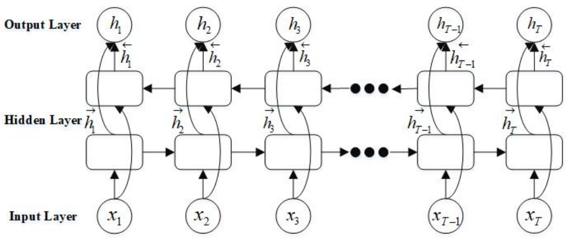


Fig.3 Cyclic neural network model.

As can be seen from Fig.3, differing from the traditional neural network, information in the deep circulation unit network model can be transmitted between different layers and between adjacent nodes of the hidden layer to each other. However, when the distance between the previous information and the current predicted position is far, the cyclic unit network will lose this learning ability, that is, the gradient disappears. In order to solve the problem of gradient disappearance and long-term dependence, Reference [9] Hochreiter and others proposed a long-term and short-term memory unit, namely LSTM model, which has been proved to be very effective in solving long-term dependence and other problems. Reference [10] In 2014, K. Cho and others improved the LSTM model and proposed the gated recurrent unit, namely GRU model, which is easier to calculate and realize.

The biggest difference between GRU and LSTM is that in LSTM, the model determines how much information of the previous layer is saved to the cell layer by setting the parameter values of the forgetting gate, output gate and reset gate, so as to determine the influence of the previous sequence on the current output, while in GRU algorithm, the three gate structures and cell units in LSTM are removed, and the update gate and reset gate are used instead. Moreover, the output of the model has only one hidden layer, which reduces the weights to be adjusted in the training process and simplifies the algorithm.

On the problem of extending time memory and eliminating the disappearance of gradient, GRU algorithm, like LSTM algorithm, uses two gate structures to determine the current input and the weighted output of the hidden layer of the previous node to determine how much information enters the current node, thus eliminating the problem of small gradient and time memory existing in RNN algorithm. The GRU neural network structure is shown in Fig.4.

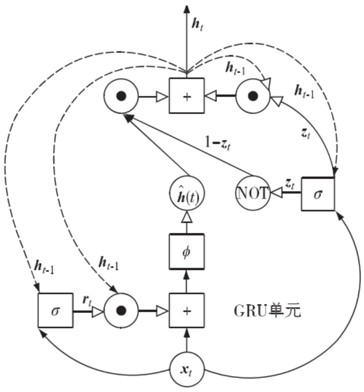


Fig.4 Gated Recurrent Unit Structure.

When input x_t enters the GRU unit, it is divided into three directions:

First: Entering the reset door R together with the hidden layer output h_{t-1} of the previous input. After weighting and adding, the output r_t of the reset door is formed after the activation function σ. This mechanism can make the model discard some useless information, in (1).

$$r_t = \sigma\left(W_{rx} x_t + W_{rh} h_{t-1}\right) \quad (1)$$

Second: Together with the reset gate output r_t, the hidden layer output h_t of the previous node is added with the reset gate signal, and the weight value is added with the current output to form the output $\hat{h}_t$ of GRU, in (2).

$$\hat{h}_t = \varphi\left(W_{rx}x_t + r_t \odot W_{rh}h_{t-1}\right) \quad (2)$$

Third: Through the update gate, the weight z_t of the update gate is obtained, which is used to determine how much specific gravity in $\hat{h}_{t-1}$ can enter the output of the data at the moment, and then weight with the output of the reset gate to get the final output, in (3) and (4).

$$z_t = \sigma\left(W_{zx}x_t + W_{zh}h_{t-1}\right) \quad (3)$$

$$h_t = (1 - z_t) \odot \hat{h}_t + z_t \odot h_{t-1} \quad (4)$$

Where x_t is the input of the current neural network and h_{t-1} is the activation value of the output of the previous hidden node. σ denotes the sigmoid function, φ denotes the tanh function, $\odot$ denotes the Hadamard product, and all W are weight parameter matrices to be trained by the model.

III. Traffic Flow Status Prediction

A. *The Data*

The research data in this paper is the bus-vehicle GPS data provided by the public transportation company of Hohhot City. We know that when the road traffic is smooth, the driving speed of the bus does not change too much, and the azimuth change tends to be stable; when the traffic is congested, the speed of the bus changes and fluctuates, the azimuth angle is erratic, fluctuations appear more obvious, acceleration, deceleration, starting and braking of buses appears more frequently. That is to say, the movement speed and azimuth of the bus depend on the road congestion level and traffic flow, so the external shape and intrinsic characteristics of the vehicle-related signals reflect the road traffic status. This paper selects the data from September 1st to 6th, 2017 as the historical data set training model, and the data of September 7, 2017 is used as the test set of the model to verify the model validity. The bus speed curve chart at the morning peak on September 2, 2017 is shown in Fig.5.

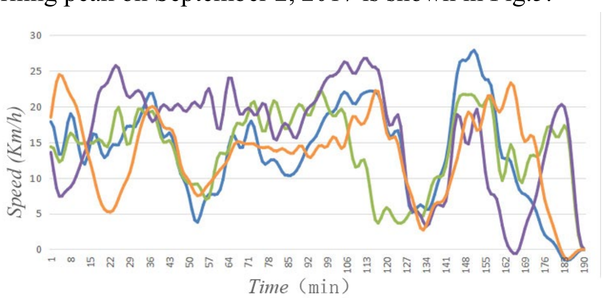


Fig.5 Speed Curve during Morning Peak Hours.

B. *Data Preprocessing*

The data studied in this paper is the bus historical data collected by GPS equipment from Hohhot bus corporation. Due to abnormal GPS equipment, unstable transmission or other factors, abnormal value and missing value will appear in the data. Therefore, it is necessary to pre-process the data. The first step is to remove the abnormal data such as stopping at the station and waiting for the traffic light. And then is to fix the abnormal and missing data according to the historical data of the same period.

C. *The Evaluation Index*

In order to evaluate the performance of the prediction model, this paper selects goodness of fit (R) and square pairwise percentage error (MAPE) to measure the prediction accuracy of the model, in (5) and (6).

$$R = [1 - \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 / \sum_{i=1}^{n}(y_i - \overline{y})^2] \times 100\% \quad (5)$$

$$MAPE = [\frac{1}{n}\sum_{i=1}^{n}\left|\frac{(y_i - \hat{y}_i)}{y_i}\right|] \times 100\% \quad (6)$$

Where n is the number of samples; y_i is the true value of the ith sample; $\hat{y}_i$ represents the predicted value of the ith sample; $\overline{y}$ is the average.

D. *Experimental Content and Results*

In order to evaluate the effect of GRU in traffic flow prediction, this paper also constructs the LSTM model. The changes in the MAPE index of the LSTM model and the GRU model are shown in Fig.6 and Fig.7.

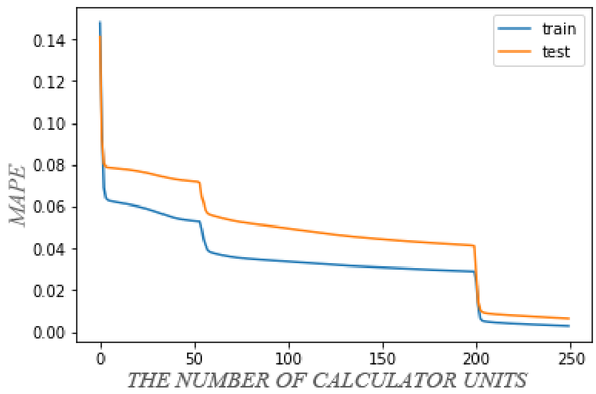


Fig.6 MAPE of LSTM Varying with the Number of Calculator Units.

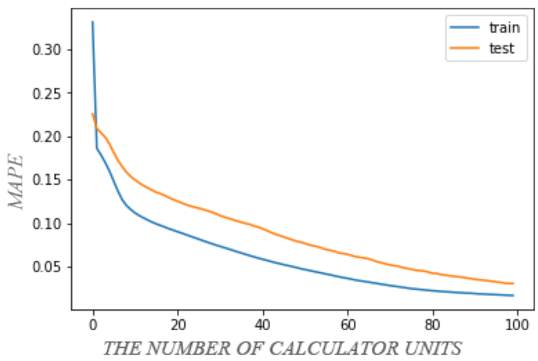


Fig.7 MAPE of GRU Varying with the Number of Calculator Units.

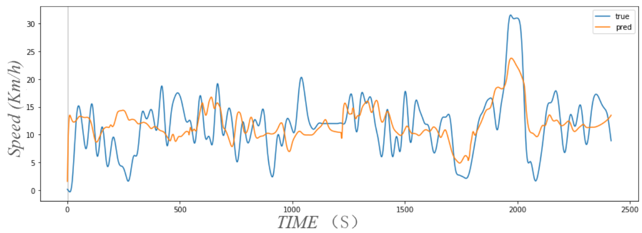
The prediction performance indicators, prediction results of the LSTM model and the GRU model under different training times are shown in Table II and Fig.8.

TABLE II
THE PERFORMANCE INDEXES OF EACH MODEL AFTER MULTIPLE TRAINING

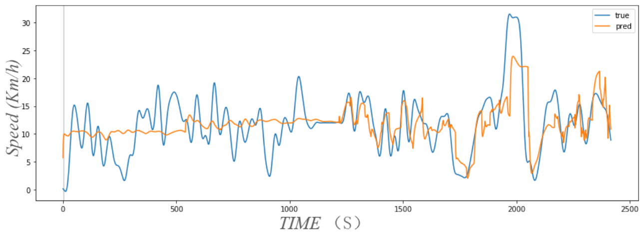
		LSTM	GRU
The 1st Training	MAPE	0.0060	0.0153
	R	68.80	76.4
	Elapsed Time	32min56s	36min10s
The 2nd Training	MAPE	0.0136	0.0154
	R	75.91	77.2
	Elapsed Time	38min43s	38min18s
The 3rd Training	MAPE	0.0060	0.0153
	R	71.40	71.4
	Elapsed Time	39min15s	38min17s
The 4th Training	MAPE	0.0063	0.0155
	R	55.76	76.8
	Elapsed Time	34min18s	34min14s
The 5th Training	MAPE	0.0053	0.0154
	R	70.17	75.4
	Elapsed Time	39min12s	39min24s
Mean	MAPE	0.0074	0.0153
	R	68.40	75.44
	Elapsed Time	36min52s	37min16s

By comparing the performance indicators, it can be seen that the prediction performance of GRU model is 7.04% higher than that of LSTM model. GRU model can self-adaptively consider the correlation between time series data and selectively use historical data to predict. Therefore, GRU model has a good performance in the field of urban traffic flow prediction.

Finally, conbining the predicted speed results of buses in various sections of the city with Reference [13] the "Urban Traffic Management Evaluation Index System" published by



(a) LSTM Model.



(b) GRU Model.

Fig. 8 Comparison of Predictive Value of Different Models with True Value.

the Ministry of Public Security in 2002, which proposes to use the average speed of motor vehicles on urban trunk roads to measure the road traffic operation status, as shown in Table III. That is to say, a more accurate traffic flow state can be obtained in this section.

TABLE III
THE RELATIONSHIP BETWEEN ROAD TRAFFIC CONDITION AND VEHICLE SPEED

The vehicle speed	Road traffic condition
More than 30 Km/h	Unimpeded
Between 20 Km/h and 30 Km/h	Mild Congestion
Between 10 Km/h and 20 Km/h	Congestion
Below 10 Km/h	Serious Congestion

IV. CONCLUSION

This paper introduces the concept of threshold recursive neuron and the basic flow of algorithm, constructs GRU neural network, and explores its applicability in traffic flow prediction. In this paper, the prediction results of LSTM and GRU are compared by various performance indicators and the results show that the GRU model in the traffic flow prediction of goodness of fit (R) and square pair percentage error (MAPE) are better than the LSTM model, that is, the accuracy is higher. Experimental results show that the GRU model can simulate the real-time changes of traffic flow state more accurately and can be used to predict the state of urban traffic flow.

The application of GRU in the traffic state forecasting model has achieved good results, but for now, this article only uses the bus speed as the input data instead of the diverse bus data such as the azimuth data, weather conditions, holidays or other factors into account. The writer believes that the accuracy of traffic flow state prediction will be further improved if more variabilities are considered. In this way, more precise early warning and corresponding solution to traffic management can be produced more timely which will improve the efficiency of network traffic and city traffic in the future.

References

[1] National Bureau of Statistics of China, "*China Statistical Yearbook 2018*," China Statistics Press, pp. 545-546. 2018.

[2] Han Chao, Song Su, Wang Chenghong, "A Real-time Short-term Traffic Flow Adaptive Forecasting Method Based on ARIMA Model," JOURNAL OF SYSTEM SIMULATION, Beijing, Vol. 16 No. 7, pp. 1530-1532+1535, July 2004.

[3] Okutani I, Stephanedes Y J, "Dynamic prediction of traffic volume through Kalman filtering theory," Transportation Research Part B: Methodological, 1984, 18(1):1-11.

[4] Fu Gui, Han Guoqiang, Lu Feng, Xu Zixin, "Short-Term Traffic Flow Forecasting Model Based on Support Vector Machine Regression," Journal of South China University of Technology (Natural Science Edition), Guangzhou, Vol. 41 No. 9, pp. 71-76, September 2013.

[5] Dochy T, Danech-Pajouh M, Lechevallier Y, "Short-Term Traffic Forecasting Using Neuronal Network," 1994.

[6] QIAN Wei, YANG Kuangli, YANG huihui, XU qingzheng, "Short-Term Traffic Flow Prediction Based on Combined Models," Computer Simulation, Beijing, Vol. 32 No. 02, pp. 175-178+193, February 2015.

[7] Hui Gao, Yujian Zhao, Lei Jia, "Summary of Short-time Traffic Flow Forecasting Methods," Journal of University of Jinan(Science and Technology), Jinan, Vol. 22 No. 01, pp. 88-94, January 2008.

[8] HE Guoguang, LI Yu, MA Shoufeng, "Discussion on Short-Term Traffic Flow Forecasting Methods Based on Mathematical Models," Systems Engineering-Theory & Practice, Beijing, Vol. 12, pp. 51-56, December 2000.

[9] Hochreiter S , Schmidhuber, Jürgen, "Long Short-Term Memory," Neural Computation, Vol. 9 No. 8, pp. 1735-1780, December 1997.

[10] Cho K , Van Merrienboer B , Gulcehre C "Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation," Computer Science, June 2014.

[11] Yi Lin,Jian-wei Zhang,Hong Liu, "Deep learning based short-term air traffic flow prediction considering temporal–spatial correlation," Aerospace Science and Technology, 2019.

[12] Liu Mingyu, Wu Jianping, "Traffic Flow Prediction Based on Deep Learning," JOURNAL OF SYSTEM SIMULATION, Beijing, Vol. 30 No. 11, pp. 4100-4105+4114, November 2018.

[13] "Evaluation index system of urban road traffic management (2002 edition) list of various urban evaluation indexes," Road Traffic management, Beijing, Vol. 06, pp. 45-46, 2002.