Traffic flow prediction based on GRU-BP combined neural network

Xijun Zhang¹ and Jiwen Li²
1 School of Computer and Communication, Lanzhou University of Technology, Lanzhou, Gansu, 730050, China
2 School of Computer and Communication, Lanzhou University of Technology, Lanzhou, Gansu, 730050, China
¹Corresponding author’s e-mail: zxjun@lut.cn

Abstract. With the rapid growth of transportation, all kinds of traffic data show explosive growth, and accurate and timely traffic forecasting is also particularly important. Among them, traffic flow prediction is the basis for realizing reasonable traffic guidance and control, and it is also a prerequisite for intelligent transportation. Traffic flow data belongs to a typical chaotic time series with strong non-linearity. In the application of neural network, a two-layer neural network can theoretically approach any continuous function infinitely, so as to achieve very accurate prediction. This paper proposes a predictive traffic flow model based on the combination of GRU network and BP neural network, and uses the US Twin Cities traffic data experiment to verify that the model is feasible in traffic flow prediction. The experimental results show that the combined model has high prediction accuracy and can capture the volatility of traffic flow at rush hour. At the same time, the model has strong robustness.

1. Introduction
With the rapid economic development, motor vehicles have become more and more popular, and traffic congestion has become a worldwide problem. A series of problems such as environmental pollution and noise pollution caused by traffic congestion have brought huge obstacles to social development. The development of intelligent transportation provides new ideas for solving the problem of traffic congestion. As an important role in intelligent transportation, short-term traffic flow forecasting can provide a solid foundation and data support for the traffic management system[1].

In intelligent transportation, a real-time and accurate short-term traffic flow forecast is the key to successful traffic guidance, and it is also the goal of joint research by current practitioners and related scholars. Since the transportation system is a man-made, time-varying, nonlinear large-scale system related to many factors, it is a highly uncertain system[2]. At present, many experts and scholars have researched and proposed many prediction models to improve the accuracy of short-term traffic flow prediction, including historical average and smoothing, statistics and regression methods, methods for giving traffic flow theory, and machine learning methods[3,4]. Qiao Yihuan and others proposed a short-term traffic flow prediction method based on one-dimensional convolution neural network and long and short-term memory[5]. The spatial information in traffic data is obtained by convolution neural network, and the time information is obtained by LSTM. Then, they used the temporal and spatial characteristics of traffic flow as regression prediction, input to the fully connected layer, and finally got the corresponding prediction result. Xu Xianfeng and other scholars proposed a combined
model method that combines graph convolution networks and gated recurrent units[6]. Experiments have shown that the proposed GCN-GRU combined model method has higher prediction accuracy, and the prediction results are better than those of Benchmark forecasting methods such as ARIMA model and GRU model.

Short-term traffic flow prediction has strong non-linearity and time correlation is strong. Traditional linear prediction models are difficult to achieve high prediction accuracy and multivariable input and output, but neural networks can solve the problem of multivariate well. In this paper, GRU neural network and BP neural network combined model algorithm is used to predict short-term traffic flow. Compared with a single prediction model, the combined model can better extract the long-term dependency characteristics of the data center. Utilizing the advantages of GRU neural network for processing time series data and the ability of BP neural network to mine complex data information, the combined model can reflect the dependence of the data in the network again to continue training, thereby making short-term traffic flow predictions with higher accuracy. The combined model can be used not only in the transportation field, but also in other fields such as finance and wind speed.

2. Related algorithms

2.1. BP neural network

BP neural network is a two-way transmission multi-layer neural network based on gradient descent, and is essentially a kind of multi-layer perceptron. The main feature is that the signal propagates forward and the error propagates backward, and it has certain promotion and generalization capabilities, and can dig out the potential laws of the given data by itself. It realizes the mapping function from input to output, and theoretically it has been proved that it has the mapping function to realize any complex non-linearity, so it is suitable for solving the problem of complex internal mechanism of data. A BP neural network generally includes an input layer, a hidden layer, and an output layer. Usually, the hidden layer contains one or more. But the most used in practical applications is the three-layer neural network[7]. The structure is shown in Figure 1.

![Figure 1: Schematic diagram of BP neural network model](image)

The input layer of the network contains $d$ nodes, the hidden layer contains $q$ nodes, and the output layer contains 1 node. $V_{dh}$ represents the weight of the input signal $X_d$ corresponding to the $h$-th hidden layer neuron, $b_h$ is the output of the $h$-th hidden layer neuron, $w_{hq}$ represents the weight of the output $b_q$ of the hidden layer neuron corresponding to the $j$-th output layer neuron, $y_j$ is the output of the $j$-th output layer neuron. It can be concluded that the input of the $h$-th hidden layer neuron is $a_h = \sum_{i=1}^{d} v_{ih}X_i$, the output of the $h$-th hidden layer neuron is $b_h = f(\gamma_h + a_h)$, the input of the $j$-th output layer neuron is $\beta_j = \sum_{h=1}^{q} w_{hq}b_h$, the output of the $j$-th output layer neuron is $\hat{y}_j = f(\gamma_j + \theta_j)$. Among them, $\gamma_h$ and $\theta_h$ are the thresholds of hidden layer neurons and output layer neurons, respectively.

Explain the training process in detail according to the three-layer BP neural network. The process of BP neural network is mainly divided into two stages. The first stage is mainly the forward
propagation of the signal. The signal passes through the input layer, the hidden layer and then reaches the output layer. The second stage is error back propagation. According to the comparison of the output results, the calculated error passes from the output layer to the hidden layer and finally to the input layer. During this period, the weights and biases from the hidden layer to the output layer and the input layer to the hidden layer are adjusted. Because the prediction error of hidden layer nodes cannot be directly calculated, the BP neural network uses the error of the output layer node to reversely estimate the prediction error of the hidden node of the previous layer to realize the adjustment of the related link weight. When adjusting the weight, when the partial derivative of the error to the weight is greater than zero, the weight is adjusted to be negative. The actual output is greater than our expected output, and the weight is adjusted in a decreasing direction, so that the difference between the actual output and the expected output is reduced. On the contrary, when the partial derivative of the error to the weight is less than zero, the actual output is smaller than the expected output, and the weight is adjusted in the direction of increasing, so that the difference between the actual output and the expected value is reduced. The specific training process of BP neural network is shown in Figure 2.

![Specific training process of neural network model](image)

Figure 2 Specific training process of neural network model

1. Network initialization. This step mainly assigns a random number in the interval (-1, 1) to each connection weight of the built neural network, and the design error function E, given the calculation accuracy ε and the maximum number of learning M.
2. Randomly select k input samples and the corresponding output.
3. Calculate the input and output of each neuron in the hidden layer.
4. Use the expected output and actual output of the network to calculate the partial derivative of the error function to each neuron in the output layer, and correct the weight of each link.
5. Save the template after updating the weights, and calculate the global error.
6. Determine whether the network error is within the requirements or the number of iterations reaches the maximum. When the error reaches the preset accuracy or the number of iterations reaches the maximum, the algorithm ends. Otherwise, return to the second step to enter the next round of training.

2.2. GRU Recurrent neural network
LSTM neural network is an improved variant of recurrent neural network RNN. It is suitable for processing and forecasting important events with relatively long intervals and delays in time series. The overall structure of LSTM and RNN are basically the same, but the hidden layer design is
different. Through the three logic gates in the memory unit, namely the input gate \( i \), the forget gate \( f \) and the output gate \( o \), the problem of gradient disappearance and gradient explosion in the back propagation process is avoided[8]. The LSTM unit structure is shown in Figure 3.

As it can be seen from Figure 3, LSTM uses two gates to control the content of the cell state, one is the forget gate, which determines how much of the cell state \( C_{t-1} \) from the previous moment is retained to the current moment \( C_t \). The other is the input gate, which determines how much of the input \( x_t \) of the network at the current moment is saved to the unit state \( c_t \). The LSTM uses an output gate to control how much of the unit state \( c_t \) is output to the current output value \( h_t \) of the LSTM.

Gated Recurrent Unit[9] was proposed by K. Cho and other scholars in 2014. It is a simplified version of LSTM. GRU maintains the effect of LSTM while making the structure simpler. The GRU unit structure is shown in Figure 4.

As it is shown in Figure 4, the GRU has only two doors: the reset door and the update door. Among them, \( r_t \) represents the reset gate, \( Z_t \) represents the update gate, \( W_z \) represents the weight matrix of the update gate, and \( W_r \) represents the weight matrix of the reset gate. The reset gate determines whether to forget the previous state, and its effect is equivalent to merging the forget gate and input gate in LSTM. When \( r_t \) tends to 0, the state information \( h_{t-1} \) at the previous moment will be forgotten, and the hidden state \( \tilde{h}_t \) will be reset to the current input information. The update gate decides whether to update the hidden state to the new state \( h_t \)[10].

\[
\begin{align*}
z_t &= \sigma(W_z \cdot [h_{t-1}, x_t]) \\
r_t &= \sigma(W_r \cdot [h_{t-1}, x_t]) \\
\tilde{h}_t &= \tanh(W_z \cdot [r_t \odot h_{t-1}, x_t]) \\
h_t &= (1 - z_t) \odot h_{t-1} + z_t \odot \tilde{h}_t
\end{align*}
\]
Above all, both LSTM and GRU retain the useful information of the previous sequence through a gating mechanism, ensuring that it will not be lost during long-term propagation. At the same time, GRU has one less gate function than LSTM, and the number of parameters is less than that of LSTM, so the overall training speed of GRU is faster than that of LSTM.

3. GRU-BP combination model

3.1. Combination model overview
The combined forecasting model constructed in this paper is shown in Figure 5. The short-term traffic flow prediction model based on GRU neural network and the short-term traffic flow prediction model based on BP neural network are constructed respectively. The test set is used as the input set and the combined model is used to make predictions to obtain the prediction results of the combined model.

Modeling process:
1. Data preprocessing: use the historical average method to repair the missing data, and divide the training set and test set.
2. Train the GRU model: use the cross-experiment method to set model hyperparameters and tune them to obtain the optimal model and build a preliminary prediction model.
3. Train the BP model: use the multi-step prediction method to make predictions, select 8 as the prediction step size through experiments, predict the data at the next moment in the future, enter the training set for training, and adjust the hyperparameters to obtain the optimal model.
4. The test set is input into the trained GRU-BP combined prediction model, and the data simulation results are compared with known samples.
5. Calculate the evaluation index of the model, and compare the prediction results of the combined forecasting model and the single neural network forecasting model.

3.2. Combination model parameter setting
This article uses the traffic flow data of the Twin Cities in the United States as the research object, and selects the data from the 66 sensors from January to March 2018, with 10 minutes between the data. Using the past 8 consecutive steps of data to predict the future 1 step of data, the proportion of the training set is 0.8 of the total set, and the remaining data is used as the verification set. The middle layer of GRU-BP neural network is determined by trial and error method as 6 layers, of which 3 layers are GRU neural layers, and the remaining 3 layers are BP neural network connection layers. The error control rate is 0.001, the number of nodes is 100, and the maximum number of training times is 1000.

4. Experiment analysis

4.1. Data source and evaluation index
In order to evaluate the performance of the prediction model, this paper mainly uses the root mean square error, the average absolute error and the mean square error as the evaluation indicators of the GRU-BP combined model.

\[
RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (Y(i) - Y_p(i))^2} \tag{5}
\]

\[
MAE = \frac{1}{N} \sum_{i=1}^{N} |Y(i) - Y_p(i)| \tag{6}
\]

\[
MSE = \frac{1}{m} \sum_{i=1}^{m} (Y(i) - Y_p(i))^2 \tag{7}
\]

In the formula, \(Y(i)\) and \(Y_p(i)\) respectively represent the true value and predicted value of the time series. The smaller the root mean square error, the average absolute error, and the mean square error, the higher the model accuracy and the higher the model prediction accuracy.

### 4.2. Result analysis

After inputting the data set into a single neural network prediction model and the combined neural network prediction model proposed in this article, the results are obtained, and the root mean square error, mean absolute error and mean square error under different models are compared. To verify the prediction accuracy of the proposed GRU-BP combined neural network model. According to formula (11), formula (12) and formula (13), the RMSE, MAE and MSE results of the three models are calculated, as shown in Table 1. Compared with the evaluation indicators of the three prediction models, the GRU-BP combined prediction model has smaller RMSE, MAE and MSE. It can be seen from the Table 1 that the RMSE of GRU-BP is 14.9% higher than GRU and 33.8% higher than BP; the MAE of GRU-BP is 9.2% higher than GRU and 30.1% higher than BP; Compared with GRU, the MSE of GRU-BP is increased by 27.6% and compared with BP by 56.2%.

| PREDICTIVE MODEL | RMSE  | MAE   | MSE   |
|------------------|-------|-------|-------|
| GRU-BP           | 12.672| 9.426 | 160.593|
| GRU              | 14.890| 10.386| 221.703|
| BP               | 19.140| 13.487| 366.519|

Table 1 Analysis of Evaluation Indexes of Different Forecasting Models
As shown in Figure 6, it can be seen intuitively that the prediction performance of the GRU-BP combined neural network prediction model for time series and the generalization ability of data are better than the single GRU neural network prediction model and the BP neural network prediction model. Therefore, the GRU-BP combined neural network prediction model can be used as an effective short-term traffic flow prediction model.

In the actual prediction model, the prediction results of each model change every time because the initial parameters of the GRU neural network prediction model and the BP neural network prediction model are randomly generated during prediction. So each time the weights and parameters calculated by the gradient descent method have corresponding changes, resulting in changes in the prediction results. But such changes are fluctuations within a small range.

5. Conclusion
This paper proposes a GRU-BP combined neural network prediction model, which uses the keras framework to build and predict the network, and realizes the short-term traffic flow prediction based on the GRU-BP combined neural network prediction model. And through the comparison with the two single network model prediction results of GRU neural network prediction model and BP neural network prediction model, it is found that the RMSE of GRU-BP is increased by 14.9% compared with GRU and 33.8% compared with BP. It shows that the GRU-BP combined neural network model has high accuracy and stability in predicting short-term traffic flow, and this model can provide a certain reference and basis for traffic diversion. In the later stage, in order to further improve the accuracy of the prediction, the experimental data and the influence factors of the road network structure can be added accordingly to improve, and the accuracy and generalization ability of the model can be further improved.

Acknowledgement:
The authors would like to thank the National Natural Science Foundation of China (61461027) and National Natural Science Foundation of China (61465007) for financial support.

Funding Statement: This study was jointly supported by National Natural Science Foundation of China (61461027, 61465007).

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.
References

[1] Xu, X.F., Huang, L.Y., Gong, M. (2020) Short term traffic flow prediction based on convolution neural network and bidirectional long short term memory network. J. Industrial instrumentation and automation., 01: 13-18.

[2] Li, M.M., Lei, J.Y., Zhao, C.J. (2019) Short term traffic flow prediction based on LSTM-BP combination model. J. Application of computer system., 28(10):152-156.

[3] Shao, J.Q. (2016) Real-time traffic flow prediction based on wavelet fuzzy neural network. J. Application of computer system., 25(07):161-164.

[4] Feng, J., Pan, F. An LSTM-BP multi-model combined hydrological forecast method. J. Computer and modernization., 07:82-85+92.

[5] Qiao, Y.H., Wang, Y., Ma, C.X., Yang, J. (2021) Short-term traffic flow prediction based on 1DCNN-LSTM neural network structure. J. Modern Physics Letters B., 35(02)

[6] Xu, X.F., Xia, Z., Zhao, L.L. (2021) Short-term traffic flow prediction method based on combined model. J/OL. Measurement and Control Technology., 2021-02-24:1-6.

[7] Shen, F., Guo, J. (2005) Data Fusion Method Based on BP Neural Network. J. Automation and Instrumentation., 2005(05):63-65+68.

[8] Wu, J.Q., Li, K., Chan, Y.N. (2020) Intelligent prediction of road congestion time based on LSTM-BP neural network. J. Digital Technology and Application., 38(05):64-65.

[9] Zhang, G.W., Liu, D. (2020) Causal convolutional gated recurrent unit network with multiple decomposition methods for short-term wind speed forecasting. J. Energy Conversion and Management., 226.

[10] Li, N., Hu, L., Deng, Z.L., Su, T., Liu, J.W. (2021) Research on GRU Neural Network Satellite Traffic Prediction Based on Transfer Learning. J. Wireless Personal Communications., prepublish.