Traffic prediction based on GCN-LSTM model

Zhizhu Wu 1, a, Mingxia Huang 1,b*, Aiping Zhao 1,c, Zhixun Lan 1,d

1School of Transportation Engineering, Shenyang Jianzhu University, Shenyang 110168, Liaoning, China

ea-mail: 451863371@qq.com, * Corresponding author: b e-mail: 379419668@qq.com, c e-mail: 2148217196@qq.com, d e-mail: 1114346121@qq.com

Abstract. Traffic flow prediction is an important part of intelligent traffic management system. Because there are many irregular data structures in road traffic, in order to improve the accuracy of traffic flow prediction, this paper proposes a combined traffic flow prediction model based on deep learning graph convolution neural network (GCN), long-term memory network (LSTM) and residual network (RESNET). GCN is used to extract the features of topology structure in traffic data, LSTM is used to extract the features of time structure, combined with ResNet to optimize the overall model, reduce the occurrence of gradient disappearance or explosion in network degradation, and finally achieve the prediction of traffic flow. According to the experimental results, the combined traffic flow prediction model used in this paper is closer to the actual traffic flow occurrence than the traditional convolutional neural network model (CNN), and the accuracy is improved.

1. INTRODUCTION

In urban traffic network, traffic flow prediction is an important task for real-time operation of traffic system, which is applied to route planning and route guidance. In the traditional traffic prediction research, short-term traffic flow prediction based on mathematical model was proposed by he Guoguang and Li Yuma Shoufeng in 2000, namely model-driven method. Due to the strong hypothesis of such method, the practical application effect is not strong. After Box and Jenkins proposed ARIMA (autoregressed sum moving average), In 2004, Han Chao et al. proposed traffic flow prediction based on ARIMA model, combined ARIMA model with simulation examples, and obtained better simulation results. In 2010, Yang Jixiang et al. proposed a parallel SVR prediction method for real-time traffic flow prediction. By selecting SVR parameters, a real-time road network traffic flow prediction model was established and the experiment was carried out in combination with G-LB algorithm. The experimental results showed that the prediction accuracy was more in-depth. With the continuous optimization of traffic data, the coming of the era of big data, YanZhen and others from the time and spatial characteristics of traffic flow is put forward a kind of short-term traffic flow prediction method based on CNN + LSTM, including CNN using convolution kernels for the space of the data feature extraction and feature mapping, LSTM time feature extraction, and ARIMA and SVR models are compared, the results showed that the root mean square error (RMSE) was reduced by 9.8%, the average absolute percentage error (MAPE) was reduced by 7.6%.

Inspired by the above literature, this paper adopts THE GCN+LSTM+RES model to predict the traffic flow.
2. Model introduction

(1) Figure convolution neural network

The graph convolutional neural network uses the pre-defined Laplacian matrix based on the
distance of nodes to model the spatial dependence of nodes in a graph. GCN is a multi-layer neural
network used to process data sets of network structure. But in real life, the irregularity of data set exposes
the disadvantages of traditional convolutional neural network. The structure of graph is an infinite
dimensional data, so it has no translation invariability, and the structure around each node is unique.
GCN extracts features through node classification, graph classification, edge prediction and other
processing of graph data.

Suppose we have a batch of graph data, in which there are N nodes (nodes), each node has its own
characteristics, we assume that the characteristics of these nodes constitute an N*D dimension matrix
X, and then the relationship between the nodes will form an N*N dimension matrix A, also known as
the adjacency matrix. X and A are our input models.

GCN is a neural network layer, and the propagation mode between layers is as follows:

\[ H^{(l+1)} = \sigma(\tilde{D}^{-\frac{1}{2}} \tilde{A} \tilde{D}^{-\frac{1}{2}} H^{(l)} W^{(l)}) \]

In this formula:
- \( \tilde{A} = A + I \), I is the identity matrix
- \( \tilde{D} \) is the degree matrix of \( \tilde{A} \), Formula for:

\[ \tilde{D}_{ii} = \sum_j \tilde{A}_{ij} \]

H is a characteristic of each layer, and for the input layer, H is X. \( \sigma \) is a nonlinear activation function.

As shown in the figure, the characteristics of each node are output as Z after passing through the
hidden layer from the X of the input layer. No matter how many layers there are in the hidden layer, the
connection relationship between nodes, i.e., A in the above equation, is Shared.

Assuming that we construct a two-layer GCN and the activation function uses ReLU and Softmax
respectively, the overall forward propagation formula is as follows:

\[ Z = f(X, A) = \text{softmax}(\tilde{A} \Re L U(\tilde{A}XW^{(0)}))W^{(1)} \]

Finally, the cross entropy loss function is calculated for all labeled nodes:

\[ L = -\sum_{l \in y} \sum_{f=1}^{F} y_{lf} \ln Z_{lf} \]

You can train a node classification model.

(2) LSTM (Long-Term Short Term Memory) is a special RNN model. Since the traditional RNN
model cannot handle the dependence problem of long distance, the residual returned will decrease
exponentially, resulting in slower network weight. To solve this problem, the LSTM model replaces
each hidden cell in the RNN with a cell with memory function, and an additional information conveyor
named "Cell State" is added at the top of the LSTM.
The LSTM system is composed of three doors, namely forgetting door, input door and output door. The calculation formula of forgetting gate is as follows:

$$f_t = \sigma(W_f * [h_{t-1}, x_t] + b_f)$$

$W_f$ is the weight matrix of the forgetting gate, $[h_{t-1}, x_t]$ means to connect two vectors into a longer vector, $b_f$ forget the door is biased to the top, $\sigma$ is the sigmoid function.

The calculation formula of the input gate is:

$$i_t = \sigma(W_i * [h_{t-1}, x_t] + b_i)$$

$$\tilde{c}_t = \tanh(W_c * [h_{t-1}, x_t] + b_c)$$

$$c_t = f_t * c_{t-1} + i_t * \tilde{c}_t$$

The above formula combines the current memory $C_t$ and long-term memory $C_{T-1}$ to form a new cell state $C_t$. Thanks to the control of the forget gate, it can store information long, long ago. Thanks to the control of the input gate, it has the ability to keep irrelevant information from entering memory.

The calculation formula of the output gate is:

$$o_t = \sigma(W_o * [h_{t-1}, x_t] + b_o)$$

$$h_t = o_t * \tanh(c_t)$$

(3) ResNet

With the deepening of the network, there are more and more training errors and more serious information loss and consumption, resulting in gradient disappearance and gradient explosion. Kaiming He and four Chinese of Microsoft Research institute proposed Resnet, which greatly improves the accuracy of the model and speeds up the training speed of the neural network.

ResNet is composed of building blocks or Prevention.

Compared with the traditional convolutional structure, Building block has a residual branch and a short-cut branch, which is used to transmit low-level information and enable the network to be trained deeply.

Originally it reduces the number of channels by a 1 by 1 convolution, and tends to reduce the number of channels of intermediate convolution to 1/4; The ordinary convolution in the middle after the
convolution the number of output channels is equal to the number of input channels; The third convolution is used to increase the number of channels so that the number of output channels equals the number of input channels. These two 1 by 1 convolution effectively reduce the number of parameters and the amount of computation involved in the convolution.

3. GLR traffic flow prediction model
We built a model for predicting traffic flow, which is called GLR.

First of all, we define the traffic network as \( G=(V,E) \) to describe the topology of the road network. Each road is treated as a node, and \( V \) is a set of road nodes, \( V=\{v_1,v_2...,v_n\} \), \( n \) is the number of nodes, \( E \) is the set of edges. The adjacency matrix \( A \) represents the connection between roads, \( A \in R(N*N) \). The adjacency matrix contains only 0 and 1 elements.

The eigenmatrix \( X \in R(N*P) \). The traffic information network on the road is the attribute feature of nodes in the network. \( P \) represents the quantitative feature of node attributes (length of historical time series) \( X_t \in R(N* I) \), which is used to represent the speed of each road. The same node attribute can be any traffic information, such as traffic speed, traffic flow, and traffic density. Therefore, the space-time problem of traffic flow prediction can be considered as the mapping function \( f \) under the premise of road network topology \( G \) and feature \( X \), and then the traffic flow information at future times can be calculated as follows:

\[
(X_{t+1},...,X_{t+T}) = f(G; (X_{t-n},...,X_{t}))
\]

4. Model prediction process
(1) Conduct standardized preprocessing of the original traffic data.

(2) The data is divided into three parts, namely training set, verification set and test set.

(3) Input the data of the training set into the prediction model for training, calculate the predicted value through the model, compare it with the actual data, and calculate the loss function. Through Res Net, ensure that with the increase of training times, the value of the loss function is smaller and the predicted value of the model is closer to the real value. When the loss function converges to a small value, the model training is successful.

(4) Input test set data into my model, and standardize the output data results to get the predicted value of the test set.

5. Source of experimental data
The traffic data set FOR this experiment, Pemsd4, is from California. The data set was collected by the Caltrans Performance Measurement System (PeMs) (Chen et al. 2001) with real-time data at intervals of every 30 seconds. Traffic data is aggregated from the raw data every 5 minutes. The system has more than 39,000 detectors deployed on highways, mainly in California. (https://pan.baidu.com/s/1GlssEHKgf9agTRsdhPPuRA), our experiment mainly consider the total flow, average flow rate and the average share.

6. Experimental data processing and conclusions
In this paper, the traffic data set of the period from January 1, 2018 to February 15, 2018 is selected. The data of 5 minutes, 15 minutes and 30 minutes are selected and input into the model. It can be seen from the training curve in the figure below that with the increase of training times, the MAE of the model gradually decreases and the function converges, and the results of the model are effective.
Figure 4. Training curve

After data is input into the model, the following figure shows the predicted value and the true value of the data from January 1, 2018 to February 14, 2010.

Figure 5. True and predicted values

The following image shows the difference between the predicted value and the true value, which is the error image.

Figure 6. Error of the image

The evaluation index reflects the validity of the model, that is, the predictive ability of the model. The following three indexes are selected: fitting degree R2, average absolute error MAE, and root mean square error RMSE, which are defined as follows: Suppose the variable to be predicted is \( Y = (y_1, y_2, \ldots, y_n) \), the predicted value is \( \hat{Y} = (\hat{y}_1, \hat{y}_2, \ldots, \hat{y}_n) \), and \( n \) is the sample of the data set.
\[
MSE = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \\
MAE = \frac{1}{n} \sum_{i=1}^{n} |y_i - \hat{y}_i| \\
R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2}
\]

7. Experimental results and analysis

According to the comparison table of the following conclusion data, it can be seen that the GCN+LSTM model has the highest fitting degree between the tested value and the real value, and the mean absolute error and root mean square error are the lowest.

| Indicator | Model       | 5   | 15  | 30  |
|-----------|-------------|-----|-----|-----|
|           | MAE         | RMSE| R2  | MAE | RMSE| R2  |
| SVR       | 8.756       | 7.941|90.4%|9.134|8.746|89.9%|10.375|9.765|87.4%|
| LSTM      | 7.498       | 6.376|93.7%|8.254|7.403|90.6%|9.874|7.945|89.6%|
| CNN+LSTM  | 6.549       | 5.873|94.9%|7.037|6.519|92.5%|7.212|6.982|91.7%|
| GCN+LSTM  | 5.679       | 3.227|95.2%|5.698|3.564|94.7%|5.723|4.273|93.2%|

| Indicator | Model       | 5   | 15  | 30  |
|-----------|-------------|-----|-----|-----|
|           | MAE         | RMSE| R2  | MAE | RMSE| R2  |
| SVR       | 9.054       | 8.459|91.9%|9.984|9.074|89.9%|11.005|10.638|85.4%|
| LSTM      | 7.992       | 6.083|93.6%|8.779|7.341|90.6%|10.327|8.052|87.9%|
| CNN+LSTM  | 6.795       | 4.732|94.3%|7.135|5.972|92.5%|7.998|6.821|90.6%|
| GCN+LSTM  | 5.757       | 3.564|95.1%|6.094|4.076|94.7%|6.213|4.592|92.9%|

8. Conclusion

With the development of intelligent transportation, the accuracy of traffic flow prediction is constantly improving, which makes it easier for residents to travel. Depth based on the study of traffic flow prediction is presented in this paper, through the GCN extract spatial structure, LSTM extracting time structure characteristics, a model based on GCN + LSTM, finally and LSTM, CNN + LSTM model is compared, the results show that the model conclusion of this paper is more close to the real value, but in this paper, the generation of the conclusion of the graph is a static picture, later research direction, the adjacency matrix is generated by neural network, a dynamic conclusion figure.

REFERENCES

[1] Qu L, Li W, Li W, et al. Daily long-term traffic flow forecasting based on a deep neural network[J]. Expert Systems with Application, 2019, 121(MAY):304-312.
[2] Wang X, Xu L, Chen K. Data-Driven Short-Term Forecasting for Urban Road Network Traffic Based on Data Processing and LSTM-RNN[J]. Arabian Journal for Engineering, 2018:1-18.
[3] Wang J, Hu F, Li L. Deep Bi-directional Long Short-Term Memory Model for Short-Term Traffic Flow Prediction[J]. 2017.
[4] Monteil J, Dekusar A, Gambella C, et al. Deep Learning to Scale up Time Series Traffic Prediction[J]. 2019.
[5] Hua Y, Zhao Z, Li R, et al. Deep Learning with Long Short-Term Memory for Time Series Prediction [J]. IEEE Communications Magazine, 2018:114-119.

[6] Dai X, Fu R, Lin Y, et al. DeepTrend: A Deep Hierarchical Neural Network for Traffic Flow Prediction [J]. 2017.

[7] Koesdwiady A, Soua R, Karray F. Improving Traffic Flow Prediction With Weather Information in Connected Cars: A Deep Learning Approach [J]. IEEE Transactions on Vehicular Technology, 2016, 65(12):9508-9517.

[8] Lee T, Nam J, Han D, et al. Developer Micro Interaction Metrics for Software Defect Prediction [J]. IEEE Transactions on Software Engineering, 2016, 42(11):1015-1035.

[9] Liqiang, Xuedong, Wang, et al. Short-Term Traffic Flow Prediction Model of Wavelet Neural Network Based on Mind Evolutionary Algorithm [J]. International journal of pattern recognition and artificial intelligence, 2018, 32(12):1850041.1.

[10] Liu Y, Zheng H, Feng X, et al. Short-term traffic flow prediction with Conv-LSTM [C] // 2017 9th International Conference on Wireless Communications and Signal Processing (WCSP). IEEE, 2017.

[11] Liu Q, Wang B, Zhu Y. Short-term Traffic Speed Forecasting Based on Attention Convolutional Neural Network for Arterials [J]. Computer-Aided Civil and Infrastructure Engineering, 2018, 33(11):999-1016.

[12] Haiyang Y, Zhihai W, Shuqin W, et al. Spatiotemporal Recurrent Convolutional Networks for Traffic Prediction in Transportation Networks [J]. Sensors, 2017, 17(17):1501.

[13] Tian Y, Zhang K, Li J, et al. LSTM-based Traffic Flow Prediction with Missing Data [J]. Neurocomputing, 2018, 318(NOV.27):297-305.

[14] Yanlin Gao, Xuegang Zhan, Chengying Chi. Emotion analysis based on cnn-lstm model [J]. Journal of Liaoning University of science and technology, 2018, 41 (06): 71-76.

[15] Xian Lu, Yong Qi, Weibin Zhang. Traffic congestion prediction method based on long-term and long-term memory model [J]. Journal of Nanjing University of technology, 2020, V.44; no.230 (01): 30-36 + 52.

[16] Zhen Yan, Chongchong Yu, Lu Han, et al. Short term traffic flow prediction method based on CNN + LSTM [J]. Computer engineering and design, 2019, v.40; No.393 (09): 228-232 + 267.

[17] Zhihao Zhang, Wenzhong Yang, Tingting Yuan, et al. Traffic accident prediction based on LSTM neural network model [J]. Computer engineering and application, 2019, v.55; no.933 (14): 255-259 + 265.

[18] Huiying Wen, Dongran Zhang. Prediction of expressway traffic volume based on Bi LSTM model [J]. Highway engineering, 2019, 044 (006): 51-56.

[19] Yishi Liu, Xuefeng Guan, Huai Wu, et al. Short term traffic speed prediction of LSTM based on spatiotemporal correlation degree weighting [J]. Geographic information world, 2020 (1): 41-47.

[20] Xucheng Li, Xiaohui Tang, Zhuo Wang, et al. Traffic state prediction based on convolutional neural network and long-term and long-term memory model [C] // China urban traffic planning annual meeting 2019.