Short-term Traffic Flow Prediction Based on Deep Neural Network Considering Spatiotemporal Correlation

Tian Yu [1, a], Weixiang Xu[2, b]

[1] *School of traffic and transportation, Beijng Jiaotong University, Beijing 100044, China*
[2] *School of traffic and transportation, Beijng Jiaotong University, Beijing 100044, China*
[a]*15120910@bjtu.edu.cn*, [b] *wxxu317@126.com*

Keywords: Short-term traffic flow prediction, Deep Neural Network, Spatiotemporal correlation.

Abstract: This study aimed to improve accurate of short-term traffic flow prediction. A considering spatiotemporal correlation traffic flow prediction method which is based on the deep neural network (ST-DNN) was proposed. In the aspect of temporal correlation, considering the influence of the lag operator and the seven days from Monday to Sunday on traffic flow prediction. In the aspect of spatial correlation, considering the impact on the current section traffic flow by the section of upstream and downstream of the road and the section with higher spatial correlation. Using data provided by Caltrans PEMS, results showed that the proposed model fits traffic flow prediction. In addition, a better performance compared with four existing methods, including ARIMA, shallow neural network (SNN), regression tree, and wavelet neural network. Clearly, this work has demonstrated the effectiveness of ST-DNN in the field of traffic flow prediction.

1. Introduction

Accurate and timely Traffic flow information, is an important factor in measuring freeway traffic conditions. The purpose of traffic flow prediction is to provide such traffic flow information. Due to the impact of traffic infrastructure, traffic demands, weather factors, and data collection, traffic flow prediction can exhibit sensitive fluctuations in different traffic conditions, which is a complex and challenging process. Therefore, proposing accurate traffic flow prediction algorithms is crucial.

In the existing traffic flow prediction literature, some fundamental assumptions or the predefined relationships between independent variables and dependent variables were applied. That is, the correctness of this assumption is sensitive to the accuracy of traffic flow prediction and is relatively random [1]. Data-driven models not only avoid the limitations of model assumptions, but also extract potential traffic flow characteristics and predict short-term traffic flow, such as neural networks (NNs), Bayesian networks [2], hybrid models [3], random forests [4], etc. Among them, NNs have good performance for nonlinear characteristics, it is appropriate for traffic flow prediction. However,

Published by CSP © 2018 the Authors
DOI: 10.23977/iceccs.2018.036

most of them are shallow traffic models, and one hidden layer is designed with poor traffic flow prediction [5]. This prompted us to use deep architecture models for traffic flow prediction.

In this paper, we propose a considering spatiotemporal correlation traffic flow prediction method which is based on the deep neural network (ST-DNN). The main objectives of this study are to (1) prove DNN has good performance for traffic flow prediction, and (2) to examine spatiotemporal correlation can improve the accuracy of traffic flow prediction. To achieve these goals, first, we study the single section traffic flow with DNN method to testify the DNN can predict traffic flow and to determine some hyper-parameters, such as lag operator, the number of iterations, etc. Then, consider temporal correlation. Since daily traffic flow trends are different, adding 1 to 7 (1, 2, 3, 4, 5, 6, 7) these seven numbers represent the day of the week. Moreover, the calculation amount will be large if every section of the entire road network is used as a whole to predict the traffic flow, and the influence of the section farther away from the current section can be negligible. Therefore, dividing entire road network into smaller road networks by clustering the sections with high correlation, with these small road networks as the research object, not only improve the prediction accuracy, but also save the computational cost.

2. The short-term traffic flow prediction model

2.1 Deep neural network

Artificial neural networks (ANN) simulate the behavior characteristics of biological neural networks, using similar transmission between the brain synapses for information processing. As described in [6], NN can be used for not only cluster, compress and filter, but also regression analysis, predictive modeling and so on. In general, biological neurons receive signals from synapses and perform weighted summation on them. When reached a certain threshold, the neurons are activated and the output of the neuron will be the input of the next neuron [7]. So is ANN.

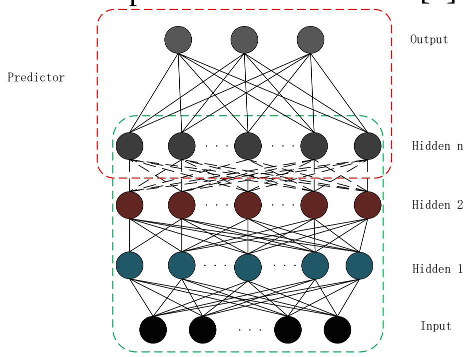

Fig. 1 The structure of DNN

Neural network has three layers structures, they are input layer, hidden layer and output layer. x (x={X_1, X_2, …, X_{n-1}, X_n}) is the input value, Y is the output value, Deep neural network (DNN) contains multiple hidden layers. The structure of a DNN is shown in Fig. 1. The lower layer of the DNN is used to extract the traffic flow characteristics and the top layer to predict the traffic flow. Because it contains multiple hidden layers, the compute formula of the hidden layer of the DNN can be represented as:

$$h_l = \begin{cases} R(xW_1^T + b_1) & l = 1 \\ R(h_{l-1}W_l^T + b_l) & 1 < l \le L \end{cases} \tag{1}$$

The compute formula of top output layer can be represented as:

$$y^{\wedge} = R(h_L W_{L+1}{}^T + b_{L+1}) \tag{2}$$

Where h_l is the output value of the l th hidden layer, and W_l is the weight matrix, b_l is the bias value. Formula (2) is the output of the top layer of the DNN, which is the value of traffic flow prediction. The input is the output of the L th hidden layer, which is represented as h_L. For the activation function, we use the *relu* as activation function. The activation functions is defined as:

$$R(x) = \max(0, x) \tag{3}$$

We train the network so that the model can fit the training data well, but as long as it is predictive, there is an error. Using the core of the least-squares method to optimize, defined as the loss function expressed as:

$$\mathrm{L(y, y^{\wedge})} = \sum (y - \mathrm{y}^{\wedge})^2 \tag{4}$$

Where y is the observed actual traffic flow data set, $y^{\wedge}$ is the traffic flow data set predicted by the neural network. In order to avoid overfitting, we add $L1$ norm regularization to the loss function to constraint model, and it is expressed as:

$$\mathrm{L(y, y^{\wedge})} = \sum (y - \mathrm{y}^{\wedge})^2 + \alpha \,|\, \mathrm{W} \,| \tag{5}$$

Training DNN is a process to minimize the loss function, shown as:

$$\Theta = \arg\min_{\Theta} L(y, y^{\wedge}) \tag{6}$$

We obtain the parameters W_l and b_l (l=1, 2, …, L+1), and here are denoted as Θ. The parameter update is based on an adaptive rate learning method (ADAELTA) proposed by Matthew D. Zeiler [8] in 2012.

2.2 Spatiotemporal correlation based on the deep neural network (ST-DNN)

Traffic flow prediction is affected by many factors. According to the time variation of traffic flow time series, the traffic flow in the current section is affected by the traffic flow in the previous sections, which are called lag operator. It can be expressed as $x=\{X_{t\text{-}k}\}$, k=1,2,…,n. Due to the different trend of traffic flow in the same section from Monday to Sunday, the symbol of the days of the week 1 to 7 is added to the input data to improve the prediction accuracy, which is shown as D={1,2,3,4,5,6,7}. The sliding window is used to entry data. So the input data can be shown as:

$$x = \{D,\ \ X_{\text{t-k}}\},\ k = 1, 2, ..., \mathrm{n} \tag{7}$$

There is a certain relationship between the upstream and downstream traffic flow and the current section traffic flow. Therefore, the spatial correlation analysis can be used to divide the road network into several smaller-scale road networks. These smaller networks can be analyzed as a whole. From the perspective of spatial correlation, not only the prediction accuracy is improved, but also the calculation complexity is reduced. Fig. 2 is a simulation of a road network structure, the red dot in the figure shows the road section, we consider section B as the main section, that affects the traffic flow in section B as the secondary section, and here are represented as $S=\{A,C,D,E,F\}$. The traffic flow of secondary section in n time periods is determined as $X_{S(t\text{-}k)}$, The method of data fusion is studied in [9], and it was proved by examples that the traffic flow of n time periods is added to the dimension of input data with higher accuracy. Therefore, data fusion to the dimension is used in this paper. The input data Fusion shown as:

$$x = \{D, X_{\text{t-k}}, X_{\text{S(t-k)}}\},\ k = 1, 2, \ldots, \text{n} \tag{8}$$

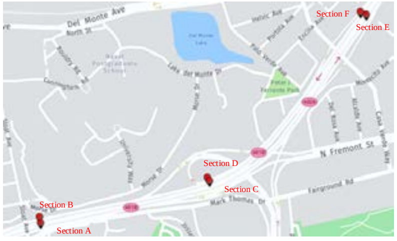

Fig. 2 The section of the road network

Training a ST-DNN can be described in Algorithm.

Algorithm: Training ST-DNN

Require: training data set x={X_1, X_2, …, X_{n-1}, X_n}, lag operator k, the number of hidden layers l, the number of hidden layer nodes n_l, activation function *relu*, the iterations T

Initialize:

step 1: input data normalization;

step 2: determine the lag operator k and the iterations T, and then add the day of the week to the input dat;

step 3: determine the secondary section using System clustering;

step 4: xavier initialization initialize $W_{1,}$ truncated normal distribution random initialize W_l, $b_{l.}$

for t=1 to T **do**

  Compute the predicted value: y^

  Compute loss function: $L(y, y\text{^})$

  Compute ΔΘusing ADADELTA

    Update Θ: Θ=Θ+ΔΘ

end for

3. Experiment and comparative analysis

3.1 Data description

In order to verify the effectiveness of the algorithm, we use traffic flow data provided by the Caltrans Performance Measurement System dataset (PEMS) [10]. Caltrans PEMS provides real-time data to the PEMS every 30s, and then accumulates 30s of data to an output every 5min with 288 data points a day. In this paper, the traffic flow date of August 1, 2016 to October 30, 2016 from District 5: Central Coast was used for experiments. The data of the first two months (August 1 to October 2) were selected as training data set, and the remaining one month’s data (October 3 to October 30) were selected as testing data set. There are 243 detectors in District 5, but because of the missing and error data is too large, we screened the data of 110 detectors. In the following study, we used the detector 500010092 as the principal section.

3.2 Impact of temporal and spatial factors

In order to reduce the computational complexity, we directly selected the hidden layer number of 6 as our deep architecture. The number of hidden layer nodes in each layer was {64, 128, 256, 512, 256, 128} respectively. The lag operator and the number of iterations were calculated in this section.

3.2.1 Using data collected from single VDS

We assumed that lag operator *k* was 10, running a single VDS deep architecture model, using sliding window entered traffic flow data. Based on the experimental results and considering the computational complexity and accuracy, we determined the number of iterations was 800.

Then we determined the lag operator *k* based on the MAE, RMSE and MRE of the training data set. The relationship between k and model error is shown in Table 1. To prevent overfitting, we chose *k* as 12.

Table 1 Effect of lag operator

k	MAE	RMSE	MRE
8	9.19	13.15	0.301
9	8.94	13.18	0.262
10	9.80	13.83	0.312
11	9.22	13.31	0.265
12	9.20	13.19	0.276

Next, we added the day of the week traffic flow data as an input data attribute. The two experiments and the corresponding result of index of performance are listed in Table 2. The results showed that the data added with the day of the week constraints was better than the original data.

Table 2 Performance of considering temporal correlation

Input data	MAE	RMSE	MRE
Original data	15.1	22.6	0.241
Added the day of the week data	13.1	19.6	0.241

3.2.2 Using data collected from upstream and downstream VDSs

The traffic flow of upstream and downstream VDS has a strong correlation with the current section traffic flow of VDS. Therefore, in this section, we will discuss the impact of upstream and downstream traffic flow on the current section traffic flow. Based on the data with the day of the week to input data, the upstream and downstream data were added to input data too. The number of the input data unchanged, since the upstream and downstream data were added to the dimension of the input data. The upstream and downstream of the VDS 500010092 are 500010121 and 500010072. The results of considering the spatial correlation are shown in Table 3. We can see that the results obtained by adding the upstream and downstream constraints were better than those using the single VDS data.

Table 3 Performance of considering the upstream and downstream

Input data	MAE	RMSE	MRE
Added the day of the week data	13.1	19.6	0.241
Considered upstream and dowmstream	12.0	17.7	0.237

3.2.3 Using data collected from the same small network of VDSs

The closer the spatial correlation between the sections is, the higher the correlation is, and the effect of the longer sections on the main section is negligible. Therefore, 110 VDSs were divided into 10 smaller networks, and each smaller network was clustered together by system clustering with higher spatial correlation. Fig.3 shows the clustering results of the road network system where the main section of the study is located.

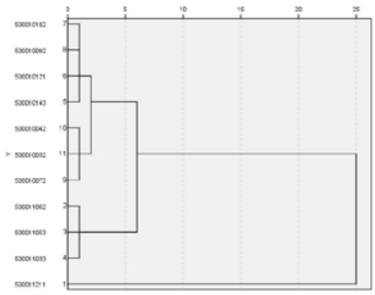

Fig. 3 The result of system clustering

We use all the data of the smaller networks and the data with the highest correlation with the main section to implement the model. The experimental results of the model are shown in Table 4, it can be seen that using all the VDS data training model was better than the single VDS data prediction effect, and using the VDS data with the highest correlation with the main section training effect was better than all the VDS data training effect.

Table 4 Performance of considering all smaller network VDSs and Clustering VDSs

Input data	MAE	RMSE	MRE
All VDSs	11.1	16.6	0.240
Clustering VDSs	11.0	15.8	0.232

3.3 Model comparison

Fig.3 shows the result of the model on predicting short-term traffic flow, and we show the one-week prediction in the test data. In Fig. 4, it can be seen that the predicted traffic flow can fit the actual data well. Therefore, the considering spatiotemporal correlation traffic flow prediction method proposed is practical and effective, which is based on the deep architecture.

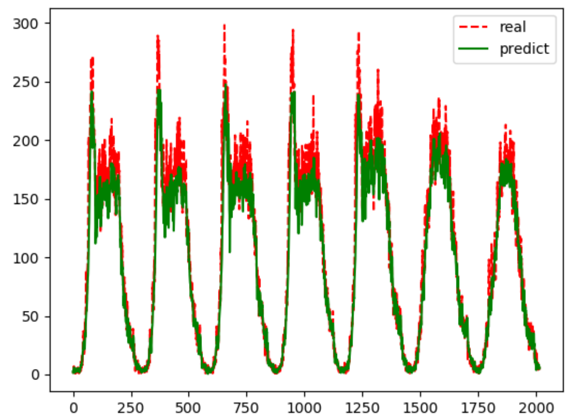

Fig. 4 The result of predicting traffic flow

Compared the performance of our prediction model with ARIMA, shallow neural network (SNN), regression tree, and wavelet neural network. Among them, ARIMA is the most basic method in time series. SNN model has good accuracy for traffic flow prediction. Regression tree is the application of non-parametric regression in traffic flow forecasting. Wavelet neural network is a more advanced technique. Table 5 shows the results of the model performance comparison. The comparison demonstrates the efficiency of our deep architecture model for short-term traffic flow prediction.

Table 5 The comparison Performance

Model	MAE	RMSE	MRE
ARIMA	11.7	16.4	0.323
SNN	13.5	18.3	0.662
Regressing tree	12.2	17.4	0.271
Wavelet NN	15.5	24.0	0.361
ST-DNN	11.0	15.8	0.232

4. Conclusion

In this paper, we consider short-term traffic flow prediction as our research problem and propose a deep neural network model considering spatiotemporal correlation to solve this problem. Through a series of experiments using data from PEMS, we evaluate the performance of proposed method and compared with ARIMA, SNN, regression tree and wavelet neural network, and the results show that the proposed method is superior to comparing methods.

Acknowledgements

This research was supported by the National Natural Science Foundation of China General Projects Grant No. 61672002 and 61272029.

References

[1] Yang H F, Dillon T S, Chen Y P, Optimized Structure of the Traffic Flow Forecasting Model With a Deep Learning Approach[J]. IEEE Transactions on Neural Networks & Learning Systems, 2016, PP (99):1-11.

[2] Fusco G, Colombaroni C, Isaenko N, Short-term speed predictions exploiting big data on large urban road networks [J]. Transportation Research Part C, 2016, 73:183-201.

[3] Younes M B, Boukerche A, A performance evaluation of an efficient traffic congestion detection protocol (ECODE) for intelligent transportation systems [J]. Ad Hoc Networks, 2015, 24(2):317-336.

[4] Hou Y, Edara P, Sun C, Traffic Flow Forecasting for Urban Work Zones [J]. IEEE Transactions on Intelligent Transportation Systems, 2015, 16(4):1761-1770.

[5] Huang W, Song G, Hong H, et al, Deep Architecture for Traffic Flow Prediction: Deep Belief Networks With Multitask Learning [J]. IEEE Transactions on Intelligent Transportation Systems, 2014, 15(5):2191-2201.

[6] J Zhang， AE Kamel, Virtual traffic simulation with neural network learned mobility model [J]. Advances in Engineering Software, 2017:103-111.

[7] Orlandi A, Multiple Objectives Optimization for an EBG Common Mode Filter by Using an Artificial Neural Network [J]. IEEE Transactions on Electromagnetic Compatibility, 2017, PP (99):1-6.

[8] Zeiler M D, ADADELTA: An Adaptive Learning Rate Method [J]. Computer Science, 2012.

[9] 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.

[10] http://pems.dot.ca.gov