ISSN: 2414-1895 DOI: 10.6919/ICJE.202203_8(3).0057

Traffic Flow Prediction based on Bi LSTM and Attention

Wei Dong*, Wei Sun

School of Information Engineering, Shanghai Maritime University, Shanghai 201306, China

*1181985292@qq.com

Abstract

In order to improve the accuracy of traffic time prediction, a combined model based on bilstm and attention is established. That is to add attention mechanism to the bilstm model, so that the model can give different weights to different time steps. Experiments show that the bilstm-att model with attention mechanism has better prediction ability than the bilstm model.

Keywords

Traffic Forecast; Spatiotemporal Correlation; Self Attention Mechanism; Graph Convolution Network.

1. Introduction

Traffic flow forecasting is an important application of computational intelligence and an active research topic in Intelligent Transportation Systems (ITS). This is of great significance to congestion prediction, real-time dynamic traffic signal optimization, real-time dynamic route planning, and traffic management decision. However, accurate and real-time forecasting of traffic flow in urban road network is a huge challenge. On the one hand,there is complex spatiotemporal dependence among traffic flows of the whole road network. On the other hand, traffic flows are nonlinear and dynamic, which could be influenced by many factors, such as weather, holidays, large conference or performances, traffic accidents and so on.

In order to solve this limitation, we adopt bilstm (bilstm ATT) based on attention mechanism, which dynamically learns the directional time dependence by considering real-time traffic conditions, connectivity and distance between sensors and traffic flow direction. We first extract and learn the temporal characteristics of traffic flow through bilstm, and then learn through the attention mechanism to redistribute the correlation size of each time, so as to optimize the traffic flow.

2. Relate Work

Traditional single-link traffic flow forecasting usually predict one link's unidirectional traffic flow at a time, which do not take the relevance of adjacent links into account. In time series community, the most popular models are Auto-Regressive Integrated Moving Average (ARIMA) and Calman filtering [1], [2], [3]. In recent years, several multi- link traffic forecasting methods have been proposed to forecast traffic flows or speeds of the whole road network simultaneously. It requires us to consider the complex spatiotemporal dependence among traffic flows of different parts of road network. To address this challenge, we could use deep neural networks. In [4], [5], [6], the authors tried to use some neural network models to solve the problem, but spatial structure was not taken into account. In [7], [8], the authors modelled the spatial correlation with Convolutional Neural Networks (CNNs), in other word, the spatial structure is in the Euclidean space (e.g., 2D images). In [9], the authors modelled the sensor network as a undirected graph and applied ChebNet and convolutional sequence model [10] for forecasting. One limitation of the mentioned spectral based convolutions is

ISSN: 2414-1895 DOI: 10.6919/ICJE.202203_8(3).0057

that they generally require the graph to be undirected to calculate meaningful spectral decomposition. At present, the state-of-the-art model is Diffusion Convolutional Recurrent Neural Network (DCRNN) [11], which models the traffic flow as a diffusion process on a directed graph, but this model needs to calculate the accurate weights of edges in graphs in advance, which is hard to realize in many cases.

3. Problem Definition

In this study, traffic prediction is carried out to predict urban roads according to historical traffic conditions and future traffic conditions. Generally speaking, traffic state can refer to traffic flow. Speed and density. In this study, traffic state only refers to traffic speed.

Definition 1: Road network G: urban road network with topological structure is described as G=(V,E), $V = \{v_1, v_2, \dots v_N\}$ is a segment set, N Is the number of sections. E is a collection of edges, which reflects the connection between various sections. The entire connectivity information is stored in adjacent matrices $A \in \mathbb{R}^{N\times N}$Rows and columns arrange segments according to the index of the road, and the value of each entry represents the connectivity between the corresponding road segments. The entry value is 0 if there is no connection between roads; If there is no connection between roads, it is 1 (unweighted graph), otherwise it is non negative (weighted graph).

Definition 2: Characteristic matrix $X \in \mathbb{R}^{N\times P}$: The speed of traffic in an area. The link is regarded as the attribute of the network node. Xi Represents the speed of the ith time in the whole time series feature.

Therefore, the temporal and spatial dependence of traffic prediction model can be regarded as learning mapping function f, based on road network G and ground features. Matrix X of road network. The traffic speed moment of T in the future is calculated as follows:

$$[X_{t+1}, \cdots, X_{t+T}] = f\big(G;\ (X_{t-n}, \cdots, X_{t-1}, X_t)\big) \tag{1}$$

Where, f is the model you want to learn, n is the length of a given historical time series, and t is the length of the time series to be predicted.

3.1 Sub-section Headings

The section headings are in boldface capital and lowercase letters. Second level headings are typed as part of the succeeding paragraph (like the subsection heading of this paragraph). All manuscripts must be in English, also the table and figure texts, otherwise we cannot publish your paper. Please keep a second copy of your manuscript in your office. When receiving the paper, we assume that the corresponding authors grant us the copyright to use the paper for the book or journal in question. When receiving the paper, we assume that the corresponding authors grant us the copyright to use the paper for the book or journal in question. When receiving the paper, we assume that the corresponding authors grant us the copyright to use.

4. Model-introduction

The model in this paper is mainly composed of input layer, Bi LSTM, attention module and output layer. Each sub module is introduced in detail below.

Volume 8 Issue 3, 2022
DOI: 10.6919/ICJE.202203_8(3).0057

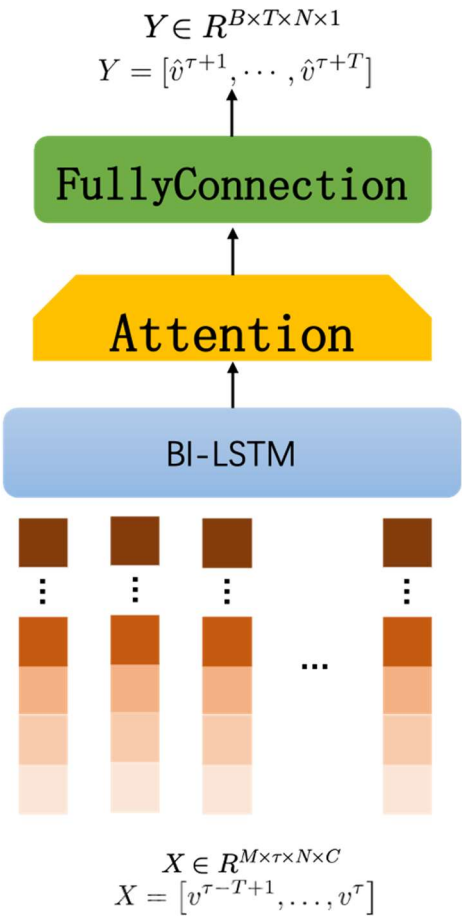

Fig. 1 Bi-LSTM-Att model

4.1 BI-LSTM Model

Recurrent neural network (RNN) is often used in the network with serialization characteristics. Data processing, each hidden layer neuron h is composed of the current input and the previous one The hidden layer is composed of the output of neurons, which enables the data to be transmitted backward. But it also brings a big problem. This long-term dependence will cause the network to remember a large amount of redundant information, the weight update is slow, and some important information will be forgotten with the increase of nodes. Short term memory network (LSTM) is an improvement of RNN, which can solve the above problems of RNN. LSTM is composed of input gate, output gate, forgetting gate and internal memory unit, as shown in equation (2) - equation (7). After setting the time step, we can pay adaptive attention to the characteristics in these time steps, so as to solve the traffic prediction problem of time series.

$$f_i = \sigma\left(W_f x_i + U_f h_{i-1} + b_f\right) \tag{2}$$

$$i_t = \sigma(W_i x_t + U_i h_{t-1} + b_i) \tag{3}$$

$$c_t' = \tan h\,(W_c x_t + U_c h_{t-1} + b_c) \tag{4}$$

$$f_i = f_c h_{t-1} + i_t c_t' \tag{5}$$

$$o_t = \sigma(W_o x_t + U_o h_{t-1} + b_o) \tag{6}$$

ISSN: 2414-1895 DOI: 10.6919/ICJE.202203_8(3).0057

$$h_t = o_t \tanh(c_t) \tag{7}$$

Equation (2) represents the forgetting gate, which can control the discarding of some unimportant information and reduce a large amount of redundant information in network transmission; Equation (3) represents the input gate, which can be used to control the degree of updating the input XT and the current state HT - 1 to the memory unit; Equations (4) and (5) are internal memory units that combine the information of the forgetting gate and the input gate to determine which information can be updated; Equations (6) and (7) are output gates to calculate the current hidden layer state so that the network can continue to pass back. Compared with RNN sharing the same set of weights and offsets will lead to gradient explosion and disappearance. LSTM shares a set of weights and offsets for each gate, which can control the inflow and outflow of some information, so that the whole network can better grasp the relationship between sequence information.

Bidirectional long short term memory (bilstm) is a combination of forward LSTM and backward LSTM. It mainly processes the stream data with time series relationship, and fully excavates and uses the context information while retaining the sequential characteristics of the data. The result of each unit in the bilstm is affected by the front and rear units. In the forward LSTM, the unit state at is affected by the result of the previous unit at - 1, and in the backward LSTM, it is affected by the state of at + 1. The two states work together to obtain the result at this time [7]. The combination of forward LSTM and backward LSTM makes full use of the context information of time series, making the prediction results more accurate.

In the task of traffic time prediction, the traffic time of a time point is related to the adjacent front and rear time points, so the two-way prediction model has strong interpretability. For example, if a road is congested at 9:00, the traffic time at 9:15 will also be affected. LSTM uses the characteristics of one-way transmission to predict the current time through the traffic time of the first five time steps. However, this transmission direction is not only one-way. If you get the traffic time of 9:30 and find that the road is blocked at this time, it is reasonable to believe that the road condition at 9:15 is the same, so as to predict the corresponding traffic time. When the transmission is bidirectional, considering the influence of the front and rear time adjacent to the prediction time on the time will make the result more convincing.

4.2 Attention Mechanism

Attention model improves the ability of attention to features through the combination of neural network model and attention mechanism. Attention mechanism in view of human visual attention mechanism, human beings will focus on the most important part in the whole visual range, so as to effectively and quickly screen the most meaningful information for themselves, and greatly improve the accuracy of information processing. The attention mechanism added to the neural network is also to achieve this effect, quickly pay attention to important information and reduce the impact of irrelevant information on the results.

Specifically, in the traffic time prediction task, the addition of attention mechanism is to give different weights to each time step, reduce attention to some distant time steps and give more attention to adjacent time steps, so as to make the prediction results more accurate. Equation (8) - equation (10) is a neural network describing the attention mechanism, in which h is the output of the bilstm network, and W and W are the weights in the attention mechanism neural network, which are randomly initialized at the beginning and continuously updated through model training, α Is the attention weight vector, R represents the weighted sum of the output of bil - STM network and the attention mechanism, and represents the final output of the network.

$$M = \tanh(W * H) \tag{8}$$

Volume 8 Issue 3, 2022
DOI: 10.6919/ICJE.202203_8(3).0057

$$\alpha = \text{softmax}(W^T * H) \tag{9}$$

$$r = \tanh(H * \alpha^T) \tag{10}$$

5. Experiment and Analysis

5.1 Dataset Description

This paper uses a real-world traffic data set, that is, taxi trajectory. Ring detector data set. (Los loop), used in Los Angeles. The data set is related to traffic speed. Therefore, the traffic speed is regarded as the traffic in the experiment.

5.2 Data Processing

The flow is summarized every 15 minutes and normalized with Z-score as input. The road topology information is represented by the adjacent graph matrix.

5.3 Experimental Setting

5.3.1 Evaluation Index

In order to evaluate the prediction performance of the model, the error between the actual traffic speed and the prediction results is used as the basis for evaluation according to the following indicators:

(1) RMSE(Root Mean Squared Error)

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_{true} - y_{pred})^2} \tag{11}$$

The error between the actual speed and the predicted speed calculated by formula (11) is the smallest. ytrue is the real value at a certain time, and ypred is the predicted value of the model at a certain time.

(2) MAE(Mean Absolute Error)

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|y_{true} - y_{pred}\right| \tag{12}$$

The error between the actual speed and the predicted speed calculated by formula (12) is the smallest. ytrue is the real value at a certain time, and ypred is the predicted value of the model at a certain time.

5.3.2 Experimental Comparison and Result Analysis

The model proposed in this paper is used to predict traffic flow on the LOS datasets, and compared with the four models proposed above. Table 1 shows the prediction results of various models for traffic flow in the next 15 minutes. Fig 2 shows the traffic flow prediction results of a station.

Table 1. Three Scheme comparing

MODEL	MAE	RMSE
HA	3.7815	4.2951
ARIMA	4.9824	7.2406
BILSTM	3.3341	5.4379
BILSTM-Att	3.0156	3.3451

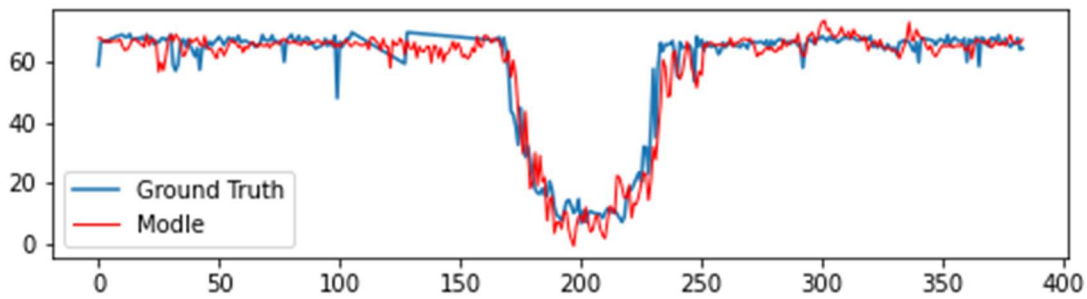

Fig. 2 Bi-LSTM-Att traffic flow prediction results of a station

We can get from the experiment that compared with Bi LSTM, our model MAE is improved by 9.55% and RMSE is improved by 38%, which is the result of the allocation optimization of each time.

6. Conclusion

In this paper, we propose a novel model for spatiotemporal traffic data prediction. Our model uses the attention mechanism to dynamically and efficiently capture the hidden dependencies between time, and the experimental results of two real data sets show the effectiveness of the model. In addition, the proposed deep learning model based on attention mode is extended to various application dynamic graph feature learning, which will be left for future research.

References

[1] M. Lippi, M. Bertini, and P. Frasconi, "Short-term traffic flow forecasting: An experimental comparison of time-series analysis and supervised learning," IEEE Transactions on Intelligent Transportation Systems, vol. 14, no. 2, pp. 871–882, 2013.

[2] I. Okutani and Y. J. Stephanedes, "Dynamic prediction of traffic volume through kalman filtering theory," Transportation Research Part B, vol. 18, no. 1, pp. 1–11, 1984.

[3] B. M. Williams and L. A. Hoel, "Modeling and forecasting vehicular traffic flow as a seasonal arima process: Theoret- ical basis and empirical results," Journal of Transportation Engineering, vol. 129, no. 6, pp. 664–672, 2003.

[4] Y. Gao and S. Sun, "Multi-link traffic flow forecasting using neural networks," in International Conference on Natural Computation, 2010, pp. 398–401.

[5] Y. Lv, Y. Duan, W. Kang, Z. Li, and F. Y. Wang, "Traffic flow prediction with big data: A deep learning approach," IEEE Transactions on Intelligent Transportation Systems, vol. 16, no. 2, pp. 865–873, 2015.

[6] R. Yu, Y. Li, C. Shahabi, U. Demiryurek, and Y. Liu, Deep Learning: A Generic Approach for Extreme Condition Traffic Forecasting, 2017.

[7] H. Yu, Z. Wu, S. Wang, Y. Wang, and X. Ma, "Spatiotemporal recurrent convolutional networks for traffic prediction in transportation networks," Sensors, vol. 17, no. 7, 2017.

[8] X. Ma, Z. Dai, Z. He, J. Ma, Y. Wang, and Y. Wang, "Learning traffic as images: A deep convolutional neural network for large-scale transportation network speed prediction," Sensors, vol. 17, no. 4, 2017.

[9] B. Yu, H. Yin, and Z. Zhu, "Spatio-temporal graph convolu- tional neural network: A deep learning framework for traffic forecasting," 2017.

[10] J.Gehring,M.Auli,D.Grangier,D.Yarats,andY.N.Dauphin, "Convolutional sequence to sequence learning," 2017.

[11] Y. Li, R. Yu, C. Shahabi, and Y. Liu, "Diffusion convolutional recurrent neural network: Data-driven traffic forecasting," 2017.

[12] P. Frasconi, M. Gori, and A. Sperduti, "A general framework for adaptive processing of data structures," IEEE Transactions on Neural Networks, vol. 9, no. 5, p. 768, 1998.