ACADEMIC
World Academic Union

ISSN 1746-7659, England, UK
Journal of Information and Computing Science
Vol. 16, No. 1, 2021, *pp*.016-023

A short-time traffic flow prediction model based on TCN-LSTM with causal convolutional layer

Chenmin Ma
Nanjing University of Information Science and Technology, Nanjing-210044, China
(Received January 04, 2021, accepted February 18, 2021)

Abstract. Short term traffic flow prediction is of great significance for traffic management, early guidance and dispersion, avoiding congestion and improving safety. Traffic flow is affected by multiple factors and exists strong time dependency. Therefore, it is very meaningful to establish a prediction model with multiple features, such as time-day-week-biweekly, holiday and weather situation, etc. In this paper, a TCN- LSTM model with causal convolution block is proposed, which is composed of two subnets: LSTM subnetwork is used to extract feature from original traffic flow data sequence, three TCN+LSTM subnetworks are used to extract features from traffic flow data with day-week-biweekly, holiday and weather. TCN is embedded to maintain causation of the input traffic flow data. Finally, features extracted from the two sub networks are merged and imported into top-level full connection network. The prediction sequence of the future short-term traffic flow is obtained at the output layer of FCN. Experimental results show that the proposed TCN-LSTM model has high accuracy and stability in short-term traffic flow prediction.

Keywords: Short-term traffic flow forecast; Convolutional neural network; Long and short term memory neural network; Causal convolutional layer

1. Introduction

In recent years, with the continuous growth of various types of vehicles in urban and rural areas, highway traffic flow is also growing rapidly, traffic congestion is serious and major casualties occur frequently. Scientific and reasonable scheduling, especially during holidays, the traffic flow problem has been the call of the whole society. In order to relieve the traffic pressure, the construction of traffic infrastructure can only alleviate the traffic pressure in a period of time. While, improving the efficiency of traffic management, making rational use of the road network and dredging ahead of time are the most effective ways to reduce traffic congestion and improve traffic safety [1]. Short-term traffic forecasting is a process of forecasting the future short-term traffic flow condition directly according to the continuous feedback of traffic information. In order to solve the problem of short-term traffic prediction, people have done a lot of different modeling work.

Traditional analysis models based on time series include: autoregressive moving average (ARMA) model and autoregressive integrated moving average (ARIMA) model. ARIMA model can eliminate the short-term fluctuations in time series, so as to better capture the long-term characteristics of time series. However, ARIMA model needs to assume that the time series are linearly correlated, and the traffic flow data is a complex and changeable non simple linear series. When the traffic conditions change dramatically, the model has obvious shortcomings in forecasting delay. In order to solve the shortcomings of ARIMA, Okutani and Stephanedes applied Kalman filtering theory to the dynamic prediction of short-term traffic flow [3], Kalman filtering model (KFM) is a state space model based on the Kalman filter theory. The equations are updated by constantly adding new sample data to make real-time prediction. KFM model can deal with both stationary and non-stationary data, and has the advantages of more flexible prediction factors and high prediction accuracy [4]. In addition, k-nearest neighbor algorithm (k-NN), wavelet decomposition and reconstruction (WDR) are popular traffic flow prediction models. Due to the randomness of traffic flow and the highly nonlinear characteristics of short-term prediction, artificial intelligence technology as an alternative method of traffic flow prediction model has been widely concerned in recent years. Smith et al. used BP neural network to predict traffic flow earlier [5], Hu et al. input historical traffic flow data into BP neural network after difference, and then predict short-term traffic flow at the next moment [6], Chan et al. (2012) proposed a new neural network training method, which used mixed exponential smoothing method

(EXP) and Levenberg method Marquardt (LM) algorithm aims to improve the generalization ability of previous methods used to train NNs for short-term traffic flow prediction [7].

In recent years, deep learning (DL) has been proved to be an effective method to extract data features, and its performance has been confirmed in many fields such as image recognition, video processing and natural language processing [13,14]. Different from the traditional ML algorithm, DL model can accept the input data in the original format and automatically discover the required features step by step. This technology is called end-to-end learning. DL greatly simplifies the process of feature engineering and improves the quality of feature [15]. Long and short term memory (LSTM) is a special deep neural network (RNN), which is suitable for modeling the dynamic time dependence in time series. Therefore, it has been proved that the prediction accuracy of LSTM model is much higher than that of traditional prediction methods [16,17]. Tian et al. discussed the performance of LSTM recurrent neural network in predicting short-term traffic flow, and compared it with several other commonly used models [18]. Fu et al. used LSTM and GRU neural network models to predict short-term traffic flow [19]. Convolutional neural networks (CNNs) are specially designed for data domain with conventional grids. They can directly identify the spatial dependencies between meshes, utilize various localized filters or kernels, and automatically learn these shift invariant kernels from the data. Based on CNN, Zhang et al. used multi-channel deep convolution neural network to classify multivariate time series [20]. Wu et al. combined CNN and LSTM, proposed a hybrid CLTFP model that can capture spatiotemporal correlation [21].

However, by studying the existing short-term traffic flow prediction models, it was found that they seldom consider the external conditions, such as weather, holidays and so on. These factors are also the main that people consider before going out and naturally needed to be considered in the road traffic. In this paper, we propose a TCN-LSTM model embedded with causal convolution layer module. The main intention is: 1) TCN can better capture the causality of time series data, and better solve the defect that CNN can't maintain the time sequence after feature extraction; 2) Combined with the influence of multi-timelines, holiday, weather and other features on traffic flow, a multi-subnetwork is established to extract features. Then, the fused multiple features are used as the input the FCN to generate the traffic flow. Experiments show that the proposed TCN-LSTM model greatly improves the accuracy of the model.

2. Related work

2.1 Long and short term memory neural network (LSTM)

Long and short term memory model (LSTM) is a special RNN structure, was proposed by Hochreiter and Schmidhuber in 1997. The main innovation of LSTM is its storage cell, which is essentially a calculator for entering information. The cell is controlled by several automatic parameters to access, write and clear. When a new input x_t appears, if the input gate z^i is activated, its information will be accumulated into the cell. At the same time, if the forget gate z^f is opened, the past cell state c_{t-1} may be forgotten in the process. Whether the latest cell output $c^{'}$ will propagate to the output state needs to further controlled by the output gate z^o. One of advantages of the LSTM model using memory cells and gates to control information flow is to prevent the gradient disappearing too quickly. Figure 1 shows the internal structure of LSTM model. The main calculation formulas of LSTM model are as follows (1) - (5):

$$f_t = \sigma\left(W_{x_f}x_t + W_{h_f}h_{t-1} + W_{c_f} \cdot C_{t-1} + b_f\right) \quad (1)$$

$$i_t = \sigma\left(W_{x_i}x_t + W_{h_i}h_{t-1} + W_{c_i} \cdot C_{t-1} + b_i\right) \quad (2)$$

$$C^{'} = f_t \cdot C_{t-1} + i_t \cdot g_t \quad (3)$$

$$O_t = \sigma\left(W_{x_o}x_t + W_{h_o}h_{t-1} + W_{c_o} \cdot C_t + b_o\right) \quad (4)$$

$$h_t = O_t \cdot tanh(C^{'}) \quad (5)$$

Where $\sigma(\cdot)$ is the sigmoid activation function, W_{x_f}, W_{h_f} and W_{c_f} represents the gate weight of each unit.