Transportation Research Part C 54 (2015) 187–197

Contents lists available at ScienceDirect

Transportation Research Part C

journal homepage: www.elsevier.com/locate/trc

Long short-term memory neural network for traffic speed prediction using remote microwave sensor data

Xiaolei Ma [a,b,*], Zhimin Tao [a,c], Yinhai Wang [d,1], Haiyang Yu [a], Yunpeng Wang [a]

[a] School of Transportation Science and Engineering, Beijing Key Laboratory for Cooperative Vehicle Infrastructure, Systems, and Safety Control, Beihang University, Beijing 100191, China
[b] Jiangsu Province Collaborative Innovation Center of Modern Urban Traffic Technologies, SiPaiLou #2, Nanjing 210096, China
[c] Department of Science and Technology, Beijing Traffic Management Bureau, Beijing 100037, China
[d] Department of Civil and Environmental Engineering, University of Washington, Seattle, WA 98195-2700, United States

ARTICLE INFO

Article history:
Received 31 October 2014
Received in revised form 1 February 2015
Accepted 9 March 2015
Available online 7 April 2015

Keywords:
Neural networks
Long short-term neural network
Traffic speed prediction
Remote microwave detector data

ABSTRACT

Neural networks have been extensively applied to short-term traffic prediction in the past years. This study proposes a novel architecture of neural networks, Long Short-Term Neural Network (LSTM NN), to capture nonlinear traffic dynamic in an effective manner. The LSTM NN can overcome the issue of back-propagated error decay through memory blocks, and thus exhibits the superior capability for time series prediction with long temporal dependency. In addition, the LSTM NN can automatically determine the optimal time lags. To validate the effectiveness of LSTM NN, travel speed data from traffic microwave detectors in Beijing are used for model training and testing. A comparison with different topologies of dynamic neural networks as well as other prevailing parametric and nonparametric algorithms suggests that LSTM NN can achieve the best prediction performance in terms of both accuracy and stability.

1. Introduction

The success of Intelligent Transportation Systems (ITS) applications relies on the quality of traffic information. This is especially true for Advanced Traffic Management Systems (ATMS) and Advanced Traveler Information Systems (ATIS), where accurate and reliable traffic information is highly desired for both transportation agencies and travelers. One of the critical needs for transportation community is to understand and forecast future traffic condition (e.g. travel time, travel flow and travel speed). A successful implementation of traffic prediction application not only can benefit travelers' route preplanning and rescheduling, but also can provide insightful information for transportation professionals to reduce congestion and improve traffic safety.

Due to the stochastic characteristics of traffic flow, accurately predicting traffic state is not a straightforward task. However, the widely deployed traffic sensors quickly increase data availability and coverage, and trigger a large number of traffic prediction studies based on various data sources. Most of these studies rely on inductive loop detector data to

* Corresponding author at: School of Transportation Science and Engineering, Beijing Key Laboratory for Cooperative Vehicle Infrastructure, Systems, and Safety Control, Beihang University, Beijing 100191, China.

E-mail addresses: xiaolei@buaa.edu.cn (X. Ma), msetao@126.com (Z. Tao), yinhai@uw.edu (Y. Wang), hyyu@buaa.edu.cn (H. Yu), ypwang@buaa.edu.cn (Y. Wang).

[1] Tel.: +1 (206) 616 2696; fax: +1 (206) 543 1543.

http://dx.doi.org/10.1016/j.trc.2015.03.014
0968-090X/

forecast either travel time or traffic volume in a short-term range, and utilize the data from video detectors or toll collection records as ground-truth to train the prediction model (Huang and Sadek, 2009; Rice and Zwet, 2004; Zou et al., 2014; Zhang and Ge, 2013). Little attention is put on travel speed prediction from other data sources (Hamad et al., 2009). Compared with the travel time data that are difficult to be directly measured in a large-scale network, speed data can be readily collected by loop detectors, Global Positioning Systems (GPS) devices and Remote Traffic Microwave Sensors (RTMS). As the most popular non-intrusive traffic detectors (Coifman, 2005, 2006), RTMS are mounted on the side of the roadway, and do not cause temporary lane closures for installation or traffic flow interruption. Therefore, more and more transportation agencies favor this non-intrusive sensor as an automatic transportation data collection approach. The principle of RTMS is to transmit microwave beams to both moving and stationary objects (i.e. vehicles, pavement, barriers, trees, etc.), and receive the reflected signals as the background signals (Junger et al., 2009). If a vehicle enters the detection zone, the reflected signal will be strengthened to exceed the background signal threshold. Consequently, the vehicle will be detected (EIS, 2003). Therefore, RTMS can detect traffic volume, occupancy and speed in multiple lanes without causing interference. The study conducted by Yu and Prevedouros (2013) indicated that the speed measurement accuracy of RTMS can achieve up to 95%, and is higher than that of single loop detector. Due to the high accuracy, the travel speed captured by RTMS is used as the predictor in this study.

As pointed by Vlahogianni et al. (2014), traffic forecasting methods have been gradually shifting from traditional statistical models to Computational Intelligence (CI) approaches. Compared with those classical statistical models such as Autoregressive Integrated Moving Average (ARIMA), the CI approaches are more flexible with no or little prior assumptions for input variables. In addition, the CI approaches are more capable of processing outliers, missing and noisy data (Karlaftis and Vlahogianni, 2011). Therefore, the CI approaches are often used to depict high dimensional and non-linear relationship. As the most representative model among the CI approaches, Artificial Neural Network (ANN) receives numerous successes in the domain of transportation (Vlahogianni et al., 2004). The precursory study of utilizing neural network into traffic prediction can be traced back to the 1990s, when Hua and Faghri (1994) introduced the concept of neural network into freeway traffic time estimation. Since then, more and more neural network variants begun to emerge for improving traffic forecasting performance of ANN. Typical examples include the widely employed Feed Forward Neural Network (FFNN) (Park and Rilett, 1999), Modular Neural Network (MNN) (Park and Rilett, 1998), Radial Basis Frequency Neural Network (RBFNN) (Park et al., 1999), Spectral-basis Neural Network (SNN) (Park et al., 1999), Neuro-Fuzzy Neural Network (NFNN) (Yin et al., 2002), and Recurrent Neural Network (RNN) (Lingras et al., 2002; Van Lint et al., 2002). Due to the dynamic nature of transportation system, RNN is especially suitable to capture the temporal and spatial evolution of traffic flow, volume and speed. This is because RNNs can use internal memory units to process arbitrary sequences of inputs, and thus grants the RNNs the capability of learning temporal sequence. Various topologies of RNNs were proposed to predict freeway traffic in the existing literatures, such as Time-Delay Neural Network (TDNN) (Ishak et al., 2003), Jordan–Elman Neural Network (Ishak et al., 2003), and State-Space Neural Network (SSNN) (Van Lint et al., 2002, 2005; Liu et al., 2006). There are two issues associated with the traditional RNN models: (1) the number of time steps ahead has to be predetermined for most RNNs. To achieve a better accuracy, finding the optimal time lag setting largely relies on the trial-and-error method; (2) previous studies have confirmed that the traditional RNNs fail to capture the long temporal dependency for the input sequence. Training the RNN with 5–10 time lags is proven difficult due to the vanishing gradient and exploding gradient problems (Hochreiter and Schmidhuber, 1997). Nevertheless, it is common to see that the strong correlation exists between two traffic events with a relative long time window. A typical example is that a traffic incident that happened 1 h ago may still cause severe congestion in the following 2 or 3 h. To address these drawbacks, a special RNN architecture named Long Short-Term Memory Neural Network (LSTM NN) (Hochreiter and Schmidhuber, 1997) is developed to predict travel speed in this study. Unlike traditional RNNs, LSTM NN is able to learn the time series with long time spans and automatically determine the optimal time lags for prediction. In the past decade, the LSTM NN has been successfully applied into robot control, speed recognition, handwriting recognition, human action recognition, etc. To the best our knowledge, there is no application of LSTM NN in the domain of transportation. This study aims to test the effectiveness of LSTM NN for short-term travel speed prediction.

Based on the aforementioned discussion, the contributions of this paper are 3-fold: (1) a novel recurrent neural network architecture: Long Short-Term Memory Neural Network, is developed to capture the long-term temporal dependency for short-term travel speed prediction; (2) the proposed algorithm can determine the optimal time window for time series in an automatic manner; and (3) a comparative study is conducted to provide a general guideline for selecting different RNN structures for short-term traffic prediction problems.

The remainder of this paper is organized as follows: A general overview of existing literatures on traffic forecasting is provided in the first section, and then, the Long Short-Term Memory Neural Network architecture is introduced, followed by a case study using RTMS detector speed data from a major expressway in Beijing, China. To further evaluate the performance of the LSTM NN algorithm, a comparison between other RNN structures (Time-delayed NN, Elman NN, and Nonlinear Autoregressive NN), Support Vector Machine (SVM) regression, Autoregressive Integrated Moving Average (ARIMA) model and Kalman Filter approach is made. Finally, conclusion and future envisions are discussed at the end of this paper.

2. Literature review

Short-term traffic forecasting has attracted numerous attentions from worldwide researchers in the past decades. Considerable research efforts have been made to enrich the traffic prediction approaches. In general, traffic forecasting