Transportation Research Part C 92 (2018) 323–348

Contents lists available at ScienceDirect

Transportation Research Part C

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

Quantifying uncertainty in short-term traffic prediction and its application to optimal staffing plan development

Lei Lin[a], John C. Handley[b], Yiming Gu[c], Lei Zhu[d], Xuejin Wen[e], Adel W. Sadek[f,*]

[a] NEXTRANS Center, Purdue University, West Lafayette, IN 47906, USA
[b] Goergen Institute for Data Science, University of Rochester, Rochester, NY 14627, USA
[c] Department of Civil and Environmental Engineering, Carnegie Mellon University, Pittsburgh, PA 15213, USA
[d] Department of Civil Engineering and Engineering Mechanics, The University of Arizona, AZ 85721, USA
[e] Department of Electrical and Computer Engineering, Ohio State University, OH 43210, USA
[f] Department of Civil, Structural and Environmental Engineering, University at Buffalo, the State University of New York, Buffalo, NY 14260, USA

ARTICLE INFO

Keywords:
PSO-ELM
GARCH
Kalman filter
Prediction interval
Reliability
Sharpness
Staffing plan
Waiting time
Uncertainty

ABSTRACT

In this paper, we aim to quantify uncertainty in short-term traffic volume prediction by enhancing a hybrid machine learning model based on Particle Swarm Optimization (PSO) and Extreme Learning Machine (ELM) neural network. Different from the previous studies, the PSO-ELM models require no statistical inference nor distribution assumption of the model parameters, but rather focus on generating the prediction intervals (PIs) that can minimize a multi-objective function which considers two criteria, reliability and interval sharpness. The improved PSO-ELM models are developed for an hourly border crossing traffic dataset and compared to: (1) the original PSO-ELMs; (2) two state of the art models proposed by Zhang et al. (2014) and Guo et al. (2014) separately; and (3) the traditional ARMA and Kalman filter models. The results show that the improved PSO-ELM can always keep the mean PI length the lowest, and guarantee that the PI coverage probability is higher than the corresponding PI nominal confidence, regardless of the confidence level assumed. The study also probes the reasons that led to a few points being not covered by the PIs of PSO-ELMs. Finally, the study proposes a comprehensive optimization framework to make staffing plans for border crossing authority based on bounds of PIs and point predictions. The results show that for holidays, the staffing plans based on PI upper bounds generated much lower total system costs, and that those plans derived from PI upper bounds of the improved PSO-ELM models, are capable of producing the lowest average waiting times at the border. For a weekday or a typical Monday, the workforce plans based on point predictions from Zhang et al. (2014) and Guo et al. (2014) models generated the smallest system costs with low border crossing delays. Moreover, for both holiday and normal Monday scenarios, if the border crossing authority lacked the required staff to implement the plans based on PI upper bounds or point predictions, the staffing plans based on PI lower bounds from the improved PSO-ELMs performed the best, with an acceptable level of service and total system costs close to the point prediction plans.

1. Introduction

In the last few decades, short-term traffic volume prediction models, an essential component for efficient traffic management, have been extensively studied. Short-term traffic volume prediction, which usually focuses on forecasting traffic changes in the near

* Corresponding author.
E-mail address: asadek@buffalo.edu (A.W. Sadek).

https://doi.org/10.1016/j.trc.2018.05.012
Received 26 January 2017; Received in revised form 26 February 2018; Accepted 12 May 2018

future (i.e., ranging from 5 min to 1 h), is very challenging because of the uncertain and chaotic nature of transportation systems. The traffic volume is the result of the interaction among many and diverse factors, such as travelers, road network, traffic accidents, weather conditions, and special events (Zhu et al., 2017a, 2017b; Zhu and Chiu, 2015). Not surprisingly, this interesting problem has drawn the attention of researchers, and various models were studied, ranging from classical statistical models to relatively novel machine learning models. With respect to the first group, the Box and Jenkins techniques (e.g., Autoregressive Integrated Moving Average (ARIMA) models) were firstly applied to the field of traffic forecasting by Ahmed and Cook (1979). Since then more and more advanced techniques from that family have been applied to traffic volume prediction, such as the seasonal ARIMA models (SARIMA) (Smith et al., 2002; Williams and Hoel, 2003), the ARIMA models with intervention X-variables (ARIMAX) (Williams, 2001), and the combination of Kohonen self-organizing map with ARIMA models (Van Der Voort et al., 1996). In addition to the Box and Jenkins models, other multivariate time series techniques were exploited to increase prediction accuracy, including the state space model (Stathopoulos and Karlaftis, 2003) and the multivariate structural time series model (MST) (Ghosh et al., 2009). Min and Wynter (2011) adopted a multivariate spatial-temporal autoregressive model (MSTAR) to predict the network-wide speed and volume in real time. Kalman filtering theory was also utilized for short-term traffic forecasting. Examples include its initial examination by Okutani and Stephanedes (1984), the state space approach by Stathopoulos and Karlaftis (2003), and the work by Xie et al. (2007) which used a Kalman filter with discrete wavelet decomposition for short term traffic prediction. Extended Kalman filter techniques that assume nonlinear state space equations to model dynamic traffic networks were also applied for traffic state estimation and prediction (Abdi et al., 2010; Hinsbergen et al., 2012; Wang et al., 2006, 2008). Usually for these Kalman filter models, macroscopic traffic flow theories are used to model the traffic states in the traffic network first, then recursive Bayesian methods such as Extended Kalman Filter, Unscented Kalman Filter and Particle Filter, are applied to infer the unknown traffic states (Wang et al., 2007, 2009).

On the machine learning side, among the most widely used methods are Neural Networks (NNs) (Vlahogianni et al., 2014). Several NN topologies have been utilized in previous studies including the multilayer perceptron networks (MLP) (Smith and Demetsky, 1994), wavelet networks (Xie and Zhang, 2006) and local linear neural networks (Chen et al., 2006; Lin et al., 2013a). NNs were also sometimes combined with other methods, such as fuzzy sets and genetic algorithms, to develop hybrid and more powerful predictions methods (Vlahogianni et al., 2005; Wei and Chen, 2012; Yin et al., 2002). More recently, with deep learning NNs demonstrating significant advantages in dealing with big data problems such as computer vision and speech recognition, these advanced techniques have also been applied to traffic state prediction (Lin et al., 2017; Lv et al., 2015; Ma et al., 2015). Besides NNs, other machine learning methods were recently proposed for short-term traffic prediction. Dimitriou et al. (2008), for example, proposed an adaptive hybrid fuzzy rule-based system approach to predict traffic flow in urban arterial networks. Support Vector Machine (SVM) has also been exploited for short-term traffic flow prediction. Specifically, Zhang and Xie (2008) compared a v-support vector machine (v-SVM) model to a NN model and concluded that the former performed better.

To further understand the strengths and weaknesses of these models, and to provide insight into choosing the most appropriate model when facing a specific traffic flow prediction task, Lin et al. (2013b) diagnosed four traffic volume datasets on the basis of various statistical measures and correlated these measures to the performance results of the three prediction models Autoregressive and Moving Average (ARMA), k nearest neighbor (k-NN) and support vector machine (SVM). Also, Karlaftis and Vlahogianni (2011) reviewed the previous studies and explained the differences and similarities of statistical models and neural networks in detail. Generally speaking, data mining models such as NNs and SVMs are often regarded as more flexible than statistical models when dealing with complex datasets with nonlinearities or missing data (Lin et al., 2013b; Zhu et al., 2017a, 2017b). However, data mining models have their limitations such as lacking explanatory power, and being computationally expensive (Karlaftis and Vlahogianni, 2011; Lin et al., 2013b).

It is worth noting, however, that most previous studies have focused on a single-value prediction of the short-term traffic volume, and relied almost exclusively on the prediction error when assessing the effectiveness of a modeling approach (Karlaftis and Vlahogianni, 2011). Given the nonlinearity of traffic flow, traditional single-value prediction approaches are unfortunately almost guaranteed to result in high prediction errors, which could have significant negative impact on the effectiveness of traffic management schemes. For example, an underestimation of traffic flow during a sports game can pose a heavy burden on the whole road network and result in heavy traffic congestion and in traffic accidents. In such a case, an accurate and reliable prediction interval (PI) with upper bound and lower bound would be more useful to quantify the uncertainty and make robust plans for traffic operators.

For forecasting applications in various domains, the use of PIs is quite useful because PIs try to capture the uncertainty associated with predicting the next observation, by asserting that the next observation will be contained within a given interval with a given probability. PIs are particularly useful in operational contexts where it is desired to make staffing plans. Jongbloed and Koole (2001) showed that point prediction of the call volume to a call center cannot guarantee the desired service quality at peak hours (calls need to be answered quickly on average in between 10 and 20 s). To address this, the researchers computed the PIs for the arrival rates and adapted the workforce for the call center based on the results. Similarly, Kortbeek et al. (2015) introduced PIs to develop flexible staffing policies that would allow hospitals to dynamically respond to their fluctuating patient population by employing float nurses. PIs also have applications in the energy industry, especially in regard to wind-generated electricity. For example, owing to the variability of wind production, PIs can be used to construct contracts for supply in an auction market (Pinson et al., 2007). Within the transportation domain, PIs have been used for bus and freeway travel times prediction (Khosravi et al., 2011), who argued that PIs of travel times are more meaningful because of the underlying complex traffic processes, and given the data quality used to infer travel time. There are also a few studies that have generated PIs for short-term traffic volume forecasting (Guo et al., 2014; Kamarianakis et al., 2005; Zhang et al., 2014), and for real-time traffic speed uncertainty quantification (Guo and Williams, 2010). More recently, Zhu and Laptev (2017) applied a Bayesian deep learning model for Uber time series demand prediction and uncertainty quantification.