Transportation Research Part C 66 (2016) 61–78

Contents lists available at ScienceDirect

Transportation Research Part C

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

Short-term traffic flow rate forecasting based on identifying similar traffic patterns

Filmon G. Habtemichael [a,*], Mecit Cetin [b,1]

[a] Civil & Environmental Engineering, Transportation Research Institute at Old Dominion University (ODU), 132 Kufman Hall, Norfolk, VA 23529, United States
[b] Director of Transportation Research Institute at ODU, Civil & Environmental Engineering, Transportation Research Institute at Old Dominion University (ODU), 135 Kufman Hall, Norfolk, VA 23529, United States

ARTICLE INFO

Article history:
Received 2 February 2015
Received in revised form 8 July 2015
Accepted 25 August 2015
Available online 12 September 2015

Keywords:
Short-term traffic forecasting
K-nearest neighbor
Traffic patterns
Weighted Euclidean distance
Traffic management
Non-parametric modeling

ABSTRACT

The ability to timely and accurately forecast the evolution of traffic is very important in traffic management and control applications. This paper proposes a non-parametric and data-driven methodology for short-term traffic forecasting based on identifying similar traffic patterns using an enhanced K-nearest neighbor (K-NN) algorithm. Weighted Euclidean distance, which gives more weight to recent measurements, is used as a similarity measure for K-NN. Moreover, winsorization of the neighbors is implemented to dampen the effects of dominant candidates, and rank exponent is used to aggregate the candidate values. Robustness of the proposed method is demonstrated by implementing it on large datasets collected from different regions and by comparing it with advanced time series models, such as SARIMA and adaptive Kalman Filter models proposed by others. It is demonstrated that the proposed method reduces the mean absolute percent error by more than 25%. In addition, the effectiveness of the proposed enhanced K-NN algorithm is evaluated for multiple forecast steps and also its performance is tested under data with missing values. This research provides strong evidence suggesting that the proposed non-parametric and data-driven approach for short-term traffic forecasting provides promising results. Given the simplicity, accuracy, and robustness of the proposed approach, it can be easily incorporated with real-time traffic control for proactive freeway traffic management.

1. Introduction

The prime objective of traffic management strategies is to handle road traffic operations up to the highest level of service possible to provide reliable, safer, and greener transportation. In past decades, traffic management has been limited to responsive schemes which react to prevailing traffic conditions. However, with the advancement in technology and the wide deployment of intelligent transportation systems, traffic operators are deploying active traffic management strategies which can dynamically apply alternative strategies proactively in response to predicted traffic conditions. Therefore, the ability to timely, reliably, and accurately forecast the dynamics of traffic over short-term horizons is becoming very important. Short-term traffic forecasting models, therefore, are an integral element of the toolset needed for real-time traffic control

* Corresponding author. Tel.: +1 757 652 1320.
E-mail addresses: fhabtemi@odu.edu (F.G. Habtemichael), mcetin@odu.edu (M. Cetin).
[1] Tel.: +1 757 683 6700.

http://dx.doi.org/10.1016/j.trc.2015.08.017
0968-090X/

and management. Moreover, such tools are important in providing travelers with reliable travel time information, optimizing traffic signals, and deployment of emergency management systems.

Given the importance of predicting the expected volume of traffic ahead of time, considerable amount of research has been focused on the topic (see Van Lint and Van Hinsbergen (2012) and Vlahogianni et al. (2014)). The availability of a vast amount of spatial and temporal traffic data coupled with advancements in statistics and data analysis techniques have created an opportunity to perform short-term traffic forecast with a reasonable prediction accuracy and short processing time. Short-term traffic forecast aims at predicting the evolution of traffic over time horizons ranging from few seconds to few hours (Van Lint and Van Hinsbergen, 2012; Vlahogianni et al., 2014).

In this paper, we present a methodology for short-term traffic forecast based on learning similar traffic patterns as a reference for providing the predictions on future traffic. Similar traffic profiles are identified using an enhanced K-nearest neighbor algorithm based on measurements of a sequence of volume of traffic at 15-min intervals. For a given prevailing volume profile, K similar profiles (nearest neighbors or candidates) are identified from a large collection of a historic traffic database. The neighboring candidates drawn from the historic database corresponding to the desired forecast time or horizon are aggregated to provide predictions of future flow rate measurements. In this paper, two similarity measures are considered to determine similar traffic patterns, namely: (1) Correlation distance, and (2) Weighted Euclidean distance. To reduce noise while computing the distance measures, the lagging volume profiles are slightly filtered using locally estimated scatter-plot smoothing (loess) technique. Moreover, to dampen the effect of dominant or extreme values of candidate neighbors Winsorization is applied.

As opposed to fitting and optimizing parametric prediction models (e.g., ARIMA), the approach proposed in this paper is fundamentally a data-driven one. The main advantage of this approach is that the predictions are generated based on the observed historical traffic patterns that are discovered from the historical datasets. Moreover, this approach is capable of providing predictions over multiple time steps or a trace forecast over a specified prediction horizon. In this study, predictions are provided over a 15-min step with a forecast horizon of 6 steps (i.e., predictions were made for one hour and thirty minutes at 15-min intervals). Considering the simplicity of the proposed approach and a relatively short computation time, it can be easily incorporated with online traffic management strategies to provide short-term traffic forecasts in real-time. The performance of the proposed approach is tested using a wide variety of freeway flow rate datasets collected from different regions and is compared with the works of Guo et al. (2014) which developed several advanced parametric models based on time series analysis reinforced with an adaptive Kalman filter. The proposed methodology is found to outperform the works of Guo et al. (2014) in terms of forecast accuracy; provided that enough samples are available in the archived datasets as explained in the paper. Moreover, comparing the works of this paper with others which applied K-NN approach of forecast, the level of details presented in terms of optimizing the parameters of the enhanced K-NN and testing the robustness of the model under different traffic scenarios corresponding to different datasets collected from different regions and different proportion of missing values is unprecedented.

The remainder of this paper is organized as follows. Following this introductory section, literature review on short-term traffic forecast is presented. This is followed by an in-depth discussion of the proposed methodology for short-term traffic forecast and the corresponding results. Finally, conclusions are drawn and future works are discussed.

2. Literature review

In the literature, short-time traffic forecast covers prediction of traffic over the time period of a few seconds to few hours in to the future using current and historic measurements of traffic variables (Vlahogianni et al., 2014). According to Van Hinsbergen and Sanders (2007) as well as Van Lint and Van Hinsbergen (2012), the approaches used in short-term traffic forecast can be broadly classified into four categories: Naïve, parametric, non-parametric, and hybrid. Naïve approaches refer to models that provide simple estimate of traffic in the future, e.g., historic averages. Parametric approaches refer to models-based techniques which require a set of fixed parameter values as part of the mathematical or statistical equations they utilize, e.g., analytical models, macroscopic models and models based on time series analysis (e.g., Wang et al., 2006). The majority of these approaches suffer from the assumptions they consider to parameterize the models and were proven to perform relatively poorly under unstable traffic conditions and complex road settings (Vlahogianni et al., 2014). On the other hand, non-parametric approaches are mostly data-driven and apply empirical algorithms to provide the predictions, e.g., approaches based on data analysis and neural network techniques. Such approaches are advantageous as they are free of any assumptions regarding the underlying model formulation and the uncertainty involved in estimating the model parameters. Other short-term traffic models have implemented a hybrid of the above-mentioned approaches (e.g., Szeto et al., 2009). Smith et al. (2002), Lin et al. (2013a) and Lippi et al. (2013) have provide a comparative analysis of a few models selected amid many.

The majority of the studies on short-term traffic forecast were conducted using standard statistical techniques such as simple smoothing, complex time series analysis and filtering methods. Application of smoothing for traffic forecast include: kernel smoothing (ElFaouzi, 1996), simple exponential smoothing (Ross, 1982), and hybrid exponential smoothing and neural networks (Chan et al., 2012). Others used time series analysis such as Autoregressive Integrated Moving Average (ARIMA) models (Cetin and Comert, 2006; Cools et al., 2009; Hamed et al., 1995; Lee and Fambro, 1999; Moorthy and Ratcliffe, 1988). A variation of the ARIMA model, which is Seasonal ARIMA (SARIMA) models, has also been implemented in many studies