Transportation Research Part C 72 (2016) 202–224

Contents lists available at ScienceDirect

Transportation Research Part C

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

Adaptive traffic parameter prediction: Effect of number of states and transferability of models

Gurcan Comert [a,*], Anton Bezuglov [b], Mecit Cetin [c]

[a] Physics and Engineering Department, Benedict College, 1600 Harden St., Columbia, SC 29204, USA
[b] Math and Computer Science Department, Benedict College, 1600 Harden St., Columbia, SC 29204, USA
[c] Civil and Environmental Engineering Department, Old Dominion University, 135 Kaufman Hall, Norfolk, VA 23529, USA

A R T I C L E   I N F O

Article history:
Received 11 November 2015
Received in revised form 22 September 2016
Accepted 25 September 2016

Keywords:
Hidden Markov model
Speed prediction
Time series ARIMA forecasting
Change point models
Recursive least squares filtering

A B S T R A C T

Traffic parameters can show shifts due to factors such as weather, accidents, and driving characteristics. This study develops a model for predicting traffic speeds under these abrupt changes within regime switching framework. The proposed approach utilizes Hidden Markov, Expectation Maximization, Recursive Least Squares Filtering, and ARIMA methods for an adaptive forecasting method. The method is compared with naive and mean updating linear and nonlinear time series models. The model is fitted and tested extensively using 1993 I-880 loop data from California and January 2014 INRIX data from Virginia. Analysis for number of states, impact of number of states on forecasting, prediction scope, and transferability of the model to different locations are investigated. A 5-state model is found to be providing best results. Developed model is tested for 1-step to 45-step forecasts. The accuracy of predictions are improved until 15-step over nonadaptive and mean adaptive models. Except 1-step predictions, the model is found to be transferable to different locations. Even if the developed model is not retrained on different datasets, it is able to provide better or close results with nonadaptive and adaptive models that are retrained on the corresponding dataset.

1. Problem definition

Intelligent Transportation Systems (ITS) applications such as Advanced Traveler Information Systems (ATIS), Freight Advanced Traveler Information Systems (FRATIS), real-time route guidance, and emergency response systems planning seek to improve the efficiency of transportation networks (i.e., safety, environment, economic competitiveness, via enhancing decision making capabilities of the system users). Accurate and reliable traffic parameter (e.g., speed, travel time) prediction can be used in real-time routing to find the fastest way to a destination and to avoid congestion for on-time delivery which can lead to efficient transportation networks where people and goods are mobilized at minimum travel time, fuel consumption, and emissions. These systems, however, demand for highly robust traffic parameter predictions. The accuracy and reliability become more important for large scale network-level applications. In real life, traffic may exhibit unexpected or expected changes due to various factors such as driver behaviors, incidents, inclement weather, special events, and demand surges which adversely affect the performance of forecasting models. This paper presents an online adaptive traffic parameter prediction model based on regime switching and numerically tests if it can be transferred to different locations (transferability), be used for prediction scopes, and effectively handle nonlinearities in the process by monitoring and switching the candidate forecasting models (Clements et al., 2004).

* Corresponding author.
E-mail address: comertg@benedict.edu (G. Comert).

http://dx.doi.org/10.1016/j.trc.2016.09.014
0968-090X/

Regime switching concept simply consists of changing the parameters of a certain model fit. This framework originally presented by Goldfeld and Quandt (1973), Cosslett and Lee (1985), and Hamilton (1989). Based on thorough review studies, regime-switching models can be divided into two categories of threshold and Markov-switching models (Potter, 1999; Kuan, 2002; Piger, 2009). The primary differences are number of changes and how they evolve overtime. Threshold models treat the shifts as triggered by the level of observed variables (usually the same lagged response variable) through smooth transitioning threshold parameters. These models, originally introduced by Tong (1990), are used to model long lasting structural breaks. Main disadvantages can be listed as fitting procedure can be challenging with possible local optimums in parameter optimization and unable to adapt to different change patterns (Kuan, 2002). More towards modeling correlated data with distinct patterns during different time periods (repetitive temporary shifts), Markov-switching models are introduced within econometrics. In Markov switching framework, possibly multiple regime shifts are treated as stochastic process which are abrupt and evolve according to a Markov chain-an unobserved discrete random variable. Thus, changes are determined through statistical inference rather than an observed specific data. In traffic flow, data generation processes (DGP) that contain stable segments with unknown break points indicating the instances when the dynamic behavior switches from one regime to another due to discrete shifts are common. They can be mainly caused by accidents, roadwork, and sudden demand surges. Certainly, fitted model parameters need to be updated in relation to these break points. However, updating parameters continuously may not be necessary while they are not changing statistically. Making unnecessary updates which entails online optimization problems, is not a computationally efficient strategy. As in regime switching models, it would be appropriate to devise a statistical test to monitor changes and update parameters conditional on the occurrence of changes. Calculating these updates and effectively detecting abrupt changes also require careful identification of observations belonging to different regimes (states).

In this research, the main objective is to treat the traffic data with unknown breaks and develop an adaptive prediction model for traffic parameters through incorporating machine learning algorithms as detection and adapting tools. The approach falls into above-mentioned Markov switching category. Technically, Hidden Markov model (HMM) is allocated to calculate the conditional probability of being in a certain traffic speed state given past observations. It provides the detection of possible shifts in the DGP and switch of forecasting models. Number of states (i.e., mixture distributions) are optimized using parts of the datasets. Time series autoregressive integrated moving average (ARIMA) models are fit to each speed condition and utilized as the main forecasting tool. Expectation Maximization (EM) is employed to update mean levels (intercepts) of ARIMA models. It is able to quickly detects a change from a normal traffic speed level alleviating the problem identified in Piger (2009). Recursive Least Squares (RLS) algorithm is used to filter the parameters of ARIMA models in order to replicate the recalibration of forecasting model parameters. The impact of different geographical locations in terms of freeways in two States (California and Virginia) and multiple detector sets on a freeway is tested. This is referred as transferability of the methods (HMM, ARIMA, EM, and RLS parameters) to other locations without retraining. A through recalibration can be done for changed factors such as varying facility types, speed limits, number of lanes, or anything that can permanently alter the state-speed representation and/or the transitional probabilities of the HMM. Practically, the recalibration or updating can be carried out given the low HMM training time (i.e., less than 20 s PC with 8 GB of memory, Pentium I3 Quad-Core CPU). However, such recalibration of HMM or RLS parameters are not dealt within this paper. Major assumption for the proposed model is an underlying Normal distribution with constant variance for the data generation process within each state.

1.1. Literature review

In the literature, many researchers have proposed a number of parametric and nonparametric methods for short-term traffic prediction. Parametric models mainly include time series, their variations, and Kalman Filter (KF) models. Nonparametric methods consist of nonparametric regression and neural networks (NN). For detailed reviews of the various short-term forecasting models and their critical aspects, the reader is referred to the work by Smith and Demetsky (1997), Smith et al. (2002), Vlahogianni et al. (2004, 2014), and Vlahogianni and Karlaftis (2011). In sum, these studies report that the non-parametric techniques have superior performance to simple time series models. However, it is acknowledged that higher computational power and volume of data are demanded. Major advantages of the parametric approach are well-developed theory and ability to interpret the parameters. Generally, parametric prediction methods first train a model using historical data to estimate parameters and then test it on different data. Main assumption in this parametric modeling approach is unchanged process characteristics (e.g., mean and variance) which may affect the prediction accuracy. Since traffic is subject to occasional abrupt disturbances (e.g., incidents and weather) or congestion level (see Sun and Zhou, 2005; Kamarianakis et al., 2010) that can potentially change the underlying dynamics of the data generation process. Therefore, robust forecasting models are needed that incorporate the changes when they occur.

Hybrid methods have been used by researchers to obtain adaptive models. Cetin and Comert (2007) combine an ARIMA model with Expectation Maximization (EM) and Cumulative Summation (CUSUM) algorithms to update the intercept. Qi and Ishak (2014) and Noroozi and Hellinga (2014) apply Hidden Markov models (HMM) for traffic speed predictions. Elhenawy and Rakha (2014) identify the congestion from mixtures of skewed speed distributions. Adaptive fuzzy logic and Bayesian credit assignment algorithms in Dimitriou et al. (2008) and Min and Wynter (2011) are applied to select the best NN flow predictor at a given period. Min and Wynter (2011) present an extended time series model that includes spatial and temporal correlations for full time range network level flow forecast in urban networks. Tan et al. (2009) offer an aggregate traffic flow prediction model that combines Exponential Smoothing, Moving Average, and ARIMA (Autoregressive Integrated