Transportation Research Part C 46 (2014) 151–178

Contents lists available at ScienceDirect

Transportation Research Part C

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

Local online kernel ridge regression for forecasting of urban travel times

James Haworth [a,*], John Shawe-Taylor [b], Tao Cheng [a], Jiaqiu Wang [c]

[a] SpaceTimeLab, Department of Civil, Environmental and Geomatic Engineering, University College London, Gower Street, London WC1E 6BT, United Kingdom
[b] Centre for Computational Statistics and Machine Learning, University College London, Gower Street, London WC1E 6BT, United Kingdom
[c] Centre for Advanced Spatial Analysis, University College London, 1st Floor, 90 Tottenham Court Road, London W1T 4TJ, United Kingdom

ARTICLE INFO

Article history:
Received 13 January 2014
Received in revised form 23 May 2014
Accepted 24 May 2014

Keywords:
Forecasting
Travel time
Prediction
Time series
Kernel method
Machine learning

ABSTRACT

Accurate and reliable forecasting of traffic variables is one of the primary functions of Intelligent Transportation Systems. Reliable systems that are able to forecast traffic conditions accurately, multiple time steps into the future, are required for advanced traveller information systems. However, traffic forecasting is a difficult task because of the nonlinear and nonstationary properties of traffic series. Traditional linear models are incapable of modelling such properties, and typically perform poorly, particularly when conditions differ from the norm. Machine learning approaches such as artificial neural networks, nonparametric regression and kernel methods (KMs) have often been shown to outperform linear models in the literature. A bottleneck of the latter approach is that the information pertaining to all previous traffic states must be contained within the kernel, but the computational complexity of KMs usually scales cubically with the number of data points in the kernel. In this paper, a novel kernel-based machine learning (ML) algorithm is developed, namely the local online kernel ridge regression (LOKRR) model. Exploiting the observation that traffic data exhibits strong cyclic patterns characterised by rush hour traffic, LOKRR makes use of local kernels with varying parameters that are defined around each time point. This approach has 3 advantages over the standard single kernel approach: (1) It allows parameters to vary by time of day, capturing the time varying distribution of traffic data; (2) It allows smaller kernels to be defined that contain only the relevant traffic patterns, and; (3) It is online, allowing new traffic data to be incorporated as it arrives. The model is applied to the forecasting of travel times on London's road network, and is found to outperform three benchmark models in forecasting up to 1 h ahead.
© 2014 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/3.0/).

1. Introduction

The short term forecasting of traffic variables such as speeds, flows and densities is one of the primary goals of Intelligent Transportation Systems (ITSs), with applications in dynamic signal control, advanced traffic management systems (ATMSs) and advanced traveller information systems (ATISs) (Vlahogianni et al., 2004). To date, a wide range of methods have been

* Corresponding author. Tel.: +44 7816076958.
E-mail addresses: j.haworth@ucl.ac.uk (J. Haworth), j.shawe-taylor@ucl.ac.uk (J. Shawe-Taylor), tao.cheng@ucl.ac.uk (T. Cheng), w.jiaqiu@ucl.ac.uk (J. Wang).

http://dx.doi.org/10.1016/j.trc.2014.05.015
0968-090X/© 2014 The Authors. Published by Elsevier Ltd.
This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/3.0/).

used for short term traffic forecasting, which can be broadly separated into two categories: (1) parametric methods, and; (2) machine learning (ML) methods.[1] The former type includes statistical (space) time series methods such as the (space–time) auto-regressive integrated moving average (ST)ARIMA model family (Billings and Yang, 2006; Cheng et al., 2010; Kamarianakis and Prastacos, 2005; Williams and Hoel, 2003), state space models (Okutani and Stephanedes, 1984; Stathopoulos and Karlaftis, 2003), and Bayesian networks (Anacleto et al., 2013a; Fei et al., 2011; Sun et al., 2004, 2005, 2006; Zheng et al., 2006). Comprehensive reviews can be found in Vlahogianni et al. (2004) and, more recently Vlahogianni et al. (2014). These methods typically assume that the data being described are stationary. That is, they must have constant mean and variance. If this assumption is not satisfied, then the data must be transformed through differencing, or some other transformation (Kendall and Ord, 1990). It is often found that these assumptions are difficult to satisfy, leading standard parametric methods to perform poorly. This has led to the recent development of local parametric model specifications that attempt to model the local characteristics of traffic data in time and/or space (Ding et al., 2010; Kamarianakis et al., 2012; Min and Wynter, 2011; Min et al., 2009, 2010).

An alternative approach is to model the data directly in a nonlinear machine learning (ML) framework. ML methods typically make minimal explicit assumptions about the data generating process, and instead try to *learn* the characteristics of the data through exposure to examples (Mitchell, 1997). ML methods have often been shown to outperform parametric methods in the literature, although a recent comparison study by Chen et al. (2012) suggests that data preprocessing is just as important as model choice. The most widespread ML method in the short term traffic forecasting literature is the artificial neural network (ANN). ANNs have a long history of successful implementation in traffic forecasting, and readers are directed to Dougherty (1995) for a review of the early work. As the power of computers has increased, researchers have developed increasingly sophisticated ANNs for forecasting traffic variables both on highways (see, e.g., Chen and Grant-Muller, 2001; van Lint et al., 2005; van Hinsbergen et al., 2009; Huang and Sadek, 2009; Li and Rose, 2011) and on urban networks (see, e.g. Ledoux, 1997; Yin et al., 2002; Vlahogianni et al., 2005, 2007). The most successful implementations circumvent the traditional black box problem of neural networks by explicitly incorporating domain knowledge into the model structure. For example, the topological structure of the road network can be explicitly represented in the nodes of the hidden layer of an ANN, making the internal function of the model more transparent (van Lint et al., 2005; van Lint, 2006). In a similar vein, Vlahogianni et al. (2005) frame their genetically optimised modular ANN as a multivariate non-linear time series model, fed with spatially and temporally lagged data. The recent work of (Chan et al., 2013a,b; Chan et al., 2012a,b) has further enhanced the position of ANNs at the forefront of the traffic forecasting literature. Other ML methods that have been applied to traffic forecasting include fuzzy rule based systems (Dimitriou et al., 2008) and hybrid models (Van Der Voort et al., 1996; Hofleitner et al., 2012). Karlaftis and Vlahogianni (2011) summarise the main differences (and similarities) between ML methods and parametric methods.

Recent developments have seen the application of kernel methods (KMs) to traffic forecasting. The term *kernel method* is an umbrella term for a broad set of techniques that share a common characteristic. They comprise two components: (1) a *function* that maps the input data into a high (possibly infinite) dimensional space, known as a feature space, and; (2) a *learning algorithm* capable of discovering linear patterns in that space (Shawe-Taylor and Cristianini, 2004). Mapping to the feature space is accomplished efficiently using a *kernel function*, hence the term KM. Because linear relations are sought in the feature space, a broad range of theoretically well founded and efficient linear algorithms can be used. To date, many linear algorithms have been *kernelised* including ridge regression (Hoerl and Kennard, 1970; Saunders et al., 1998), the generalised portrait (Vapnik and Lerner, 1963; Boser et al., 1992), principal components analysis (Schölkopf et al., 1997) and canonical correlation analysis (Hotelling, 1936; Hardoon et al., 2004) amongst many others. KMs are modular in nature, meaning that any kernel algorithm can be applied using a particular kernel, and vice versa (Shawe-Taylor and Cristianini, 2004). This gives them great flexibility as a tool for solving a wide range of practical problems. KMs are an attractive approach for modelling nonlinear and nonstationary data because they combine the advantages of principled, linear learning algorithms such as ordinary least squares (OLS) with nonlinear solutions.

The most widely used KM in traffic forecasting is Support Vector Regression (SVR). Wu et al. (2004) first used SVR for the forecasting of travel times on Taipei's freeway system. Travel times over three distances for 28 consecutive days are used to train a model to forecast the following 7 days' travel times in a one-step-ahead scenario. The results are compared to a current–time predictor and a historical mean predictor and are found to be superior in all cases. The performance of SVR compares favourably with that of ANNs. Vanajakshi and Rilett (2007, 2004) compare the performance of SVR and ANNs in forecasting travel times on San Antonio's freeway system using a forecast horizon varying from 2 min to 1 h ahead. It is found that SVR performs better than ANNs when the size of the training set is small. Zhang and Xie (2008) used ν-SVR for highway traffic volume forecasting, and found the method to outperform a multi-layer feedforward neural network (MLFNN). Other KMs have been applied to traffic forecasting with similar success. Xie et al. (2010) apply Gaussian processes regression (GPR) to highway traffic volume forecasting, and the results are compared with the ν-SVR model of Zhang and Xie. The results are found to be similar, but the GPR model has the advantage of providing error bounds on the forecasts.

One of the main challenges in applying KMs lies in deciding what information to include in the kernel. In KMs a kernel induced feature space is constructed from a database of historical data patterns to store the relevant information about a

[1] ML methods are often referred to as nonparametric methods (Vlahogianni et al., 2004) or computational intelligence (CI) methods (Karlaftis and Vlahogianni, 2011).

Download English Version:

https://daneshyari.com/en/article/6937103

Download Persian Version:

https://daneshyari.com/article/6937103

Daneshyari.com