Transportation Research Part C 80 (2017) 216–238

Contents lists available at ScienceDirect

Transportation Research Part C

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

A real time forecasting tool for dynamic travel time from clustered time series

A. Ladino *, A.Y. Kibangou, C. Canudas de Wit, H. Fourati

Univ. Grenoble Alpes, CNRS, Inria, Gipsa-Lab, F-38000 Grenoble, France

ARTICLE INFO

Article history:
Received 2 August 2016
Received in revised form 4 May 2017
Accepted 6 May 2017

Keywords:
Traffic forecasting
Travel time forecasting
Sensor fusion
Clustering
Kalman filter

ABSTRACT

This paper addresses the problem of dynamic travel time (*DTT*) forecasting within highway traffic networks using speed measurements. Definitions, computational details and properties in the construction of *DTT* are provided. *DTT* is dynamically clustered using a K-means algorithm and then information on the level and the trend of the centroid of the clusters is used to devise a predictor computationally simple to be implemented. To take into account the lack of information in the cluster assignment for the new predicted values, a weighted average fusion based on a similarity measurement is proposed to combine the predictions of each model. The algorithm is deployed in a real time application and the performance is evaluated using real traffic data from the South Ring of the Grenoble city in France.

1. Introduction

Short-term travel time forecasting is one of the most important tools for traffic management and it has received a lot of attention during the last decade. This kind of tool has been widely requested by traffic management operators and drivers using traffic infrastructures. In the actual era of data deluge, measurements collected by multiple sensors are important sources of information that require analysis, classification and processing in order to detect patterns and behaviors that can be exploited in traffic prediction. Technological and analytic solutions have been the focus of several research papers along the last years. For the sake of completeness, Vlahogianni et al. (2004) presented a collection of methods for traffic forecasting, and a more recent updated study is summarized in Vlahogianni et al. (2014). For data driven approaches, the collection of information can be classified in order to facilitate the data analysis. The separation of the information into multiple groups can be achieved through unsupervised learning techniques such as K-means, where each cluster is created in an automatic way from traffic data patterns, those can characterize in some cases typical regimes such as congestion. This information can be obtained from different variables such as travel time, queue length, density, delay, among others. These indicators qualify and provide the status of the network and multiple tools have emerged to extend this processed information to the users.

This work is motivated by the forecasting problem originally proposed in Ojeda et al. (2013) based on the existence of traffic regimes identified from a clustering process. The main idea to reconstruct the forecast relies on the existence of travel time patterns identified along the history as described in Weijermars and van Berkum (2005) and Chrobok et al. (2004) when large historical data sets are observed. The problem has been already studied in the literature from different approaches, here we raise the more relevant recent works related to this one. Ojeda et al. (2013) proposes a travel time forecast based on an

* Corresponding author.
E-mail addresses: andres.ladino-lopez@gipsa-lab.fr (A. Ladino), alain.kibangou@gipsa-lab.fr (A.Y. Kibangou), carlos.canudas-de-wit@gipsa-lab.fr (C. Canudas de Wit), hassen.fourati@gipsa-lab.fr (H. Fourati).

http://dx.doi.org/10.1016/j.trc.2017.05.002
0968-090X/

Adaptive Kalman Filter (AKF) strategy in which observations are built from historical data set of speed and flow. Ojeda et al. (2013) considers a similar strategy in flow prediction and presents the problem of multi-step ahead forecasting based on clustered time series by applying several predictors such as Gaussian maximum likelihood (GML) and AKF. Wu et al. (2014) presents an approach for short-term flow forecasting using multiple ARMAX based predictors obtained from clustered data. The ARMAX model is adapted independently to different groups of flow time series and a single prediction is selected based on one criteria that considers minimum error estimation for the predicted signals. Wan et al. (2014) uses Link Node Cell Transmission Model calibrated via Monte Carlo methods in order to generate a prediction using the expectation maximization algorithm. Recently Habtemichael and Cetin (2016) has proposed data driven strategies based on KNN ($K-$nearest neighbors) selection in order to produce the forecast. All the aforementioned methods consider the selection of single forecast among multiple candidates in order to provide an outcome for the prediction phase. Although different criteria can be established for the selection, regimes described by clustered data are not totally separable and studies like Zhang and Liu (2011) have shown the improvement of performance with combined forecasts. New methods have been emerging to combine information from these models inspired from data fusion algorithms such as Sun (2004). For instance Zhang (2011) proposes a forecast strategy based on interactive multiple models by combining different individual forecasting methods. Du et al. (2012) proposes an adaptive fusion method combining historical information and current day data. Lately Cai et al. (2016) proposes a fusion of the $K-$nearest neighbors information where the terms are weighted according to a spatio-temporal correlation and Ladino et al. (2016) proposes the combination of multiple forecasts through a weighted least squares optimization problem in which the weights are determined by the covariance error matrix.

In this study we consider the dynamic travel time (*DTT*), this indicator provides useful information for multiple users and it is common in actual traffic simulators and applications such as Canudas de Wit et al. (2015) and Papageorgiou et al. (2010). The approach in this work constructs observations for the *DTT* as explained in Yildirimoglu and Geroliminis (2013) and Elhenawy et al. (2014) (also referred as the experienced travel time for a driver) for historical data and current day data. Generally, forecast algorithms in the literature are designed to satisfy a set of constraints given by the forecast problem. Most of the algorithms take into account availability of a full set of measurements for all possible locations and time instants within the traffic network and they overcome the problems of missing data, low penetration ratio or unbalanced spatial coverage by introducing additional steps such as imputation algorithms. First, we explain the data collection and aggregation process, then, we consider a simple imputation algorithm for a real time application in order to address this issue and we present results about the performance of the applied data imputation. From the imputed data, a clustering approach is applied defining then different clusters characterized by a centroid containing the mean of the data and a given dispersion around it. The evolution of the centroid can be used as future observation that can feed a Kalman filter. Therefore the prediction problem can be viewed as a filtering one. Nevertheless the assignment of the observation during the current day to a specific cluster remains an open issue since we don't know its future. To overcome this issue, we run a Kalman filter for each cluster and then we make the fusion of the obtained forecasts. The main contribution of this paper is to apply a fusion method based on a similarity distance between the known information of current day and clusters in the history. The full mechanism is deployed in a real time application and the behavior of this strategy over the *DTT* is explored. The performance of the proposed method is evaluated using a real traffic data from the Grenoble Traffic Lab (GTL) (Canudas de Wit et al., 2015).

The remainder of this work is structured as follows. Next section will present the data workflow and the main definitions for dynamic and instantaneous travel times, considering real data scenarios. In Section 3 we present the clustering method used to identify and separate traffic regimes. Section 4 describes the forecasting algorithm and the interpretation of data. Finally we present the main results of the methodology over the real time application (Canudas de Wit et al., 2015). For the sake of completeness, we have included an appendix dedicated to the data imputation problem and its effect on the computation of the *DTT*.

1.1. Notations

n	Total number of links
$v(x,t)$	Velocity space/time continuous field.
$v(x_i,s_j)$	Discrete velocity space/time field
$DTT(x_0,x_n,t_0)$	Dynamic travel time (DTT) from x_0 to x_n at t_0
$ITT(x_0,x_n,t_0)$	Instantaneous travel time (ITT) from x_0 to x_n at t_0
$\tilde{v}_l(x_i,t_\tau)$	Imputed velocity from temporal coherence
$\tilde{v}_s(x_i,t_\tau)$	Imputed velocity from spatial coherence
$\tilde{v}_d(x_i,t_\tau)$	Imputed velocity from historical coherence
$\mathcal{K}$	Number of clusters
$D_{\mathcal{K}}$	Total distortion for a clustered set with $\mathcal{K}$ clusters
$\mu_q(k)$	Cluster's centroid q at time t_k
$\Delta\mu_q(k)$	Cluster derivative's centroid q at time t_k