ARTICLE IN PRESS

Transportation Research Part B xxx (2014) xxx–xxx

Contents lists available at ScienceDirect

Transportation Research Part B

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

A functional approach to monitor and recognize patterns of daily traffic profiles

I.G. Guardiola *, T. Leon, F. Mallor

Dept. of Eng. Management and Systems Eng., Missouri University of Science and Technology, 600 W. 14th Street, Rolla, MO 65409, USA
Dept. of Statistics and Operations Research, Universidad de Valencia, C/Doctor Moliner, 50, 46100 Burjassot, Spain
Dept. of Statistics and Operations Research, Public University of Navarre, Pamplona 31006, Spain

ARTICLE INFO

Article history:
Received 26 April 2013
Received in revised form 11 April 2014
Accepted 12 April 2014
Available online xxxx

Keywords:
Clustering
Functional data analysis
Temporal traffic flow
Multivariate control charts

ABSTRACT

Functional Data Analysis (FDA) is a collection of statistical techniques for the analysis of information on curves or functions. This paper presents a new methodology for analyzing the daily traffic flow profiles based on the employment of FDA. A daily traffic profile corresponds to a single datum rather than a large set of traffic counts. This insight provides ideal information for strategic decision-making regarding road expansion, control, and other long-term decisions. Using Functional Principal Component Analysis the data are projected into a low dimensional space: the space of the first functional principal components. Each curve is represented by their vector of scores on this basis. The principal component scores are used for clustering and also to identify outliers (meaning that there was a bad performance in the recording of data or special circumstances affected the traffic) and to monitor the traffic profile by multivariate control charts. This paper introduces this new methodology and illustrates good results by using 1-min traffic data from the I-94 Freeway in the Twin Cities, Minnesota (U.S.) metroplex ranging from 2004 to 2011.

1. Introduction

Historical traffic pattern analysis can be a great asset in the operation of transportation management systems and assessing the performance of a transportation system. Due to this relevance, historical traffic pattern analysis related to the flow, speed, and capacity variations over time have been extensively studied for the purpose of developing effective traffic prediction models, control and monitoring methodologies (Tan et al., 2009).

The analysis of the data is an antecedent step to the development of any prediction model, control or monitoring methodology. Currently, a common approach is the basic graphical interpretation of the temporal traffic patterns collected through detectors (Chung, 2003; Song and Miller, 2012); however, a far more analytical approach to determine the qualitative characteristics of the data is to undertake clustering analysis (Chung and Rosalion, 2001). The objective of clustering techniques is to identify relatively homogenous groups within a sample of entities based on the similarities between individuals (Yildirimoglu and Geroliminis, 2013). The use of clustering has been primarily focused in the field of short-term prediction models (Chiou, 2012; Caceres et al., 2012). In (Wang et al., 2006), it is argued that it is through this classification and clustering that real insights into traffic management policy can be determined for intelligent transportation systems. In

* Corresponding author at: Dept. of Eng. Management and Systems Eng., Missouri University of Science and Technology, 600 W. 14th Street, Rolla, MO 65406, USA. Tel.: +1 5733416153.

E-mail addresses: guardiolai@mst.edu (I.G. Guardiola), teresa.leon@uv.es (T. Leon), mallor@unavarra.es (F. Mallor).

http://dx.doi.org/10.1016/j.trb.2014.04.006
0191-2615/

Please cite this article in press as: Guardiola, I.G., et al. A functional approach to monitor and recognize patterns of daily traffic profiles. Transportation Research Part B (2014), http://dx.doi.org/10.1016/j.trb.2014.04.006

this paper, clustering is performed from a functional data analysis point of view, which aims to yield good discrimination of the data in combination with a reduction of dimension, and all the while maintaining the functional integrity of the data. The functional characteristics can be attributed to the behavior of traffic. In (Daganzo, 2002), a thorough review of traffic behavior and related works is supplied. An example of functional behavior can be observed during a work week day as peaks can be expected in the early daylight and evening hours as commuters set off to and head home for a day's work respectively. Thus, the traffic counts contain functional characteristics, which makes the application of functional data analysis a natural one.

The use of historical data has been used in the development of short-term prediction models and real-time monitoring and control systems. First, prediction modeling varies a great deal within the literature due to the type of data and the purpose of the model. It can be summarized that the most common traffic flow modeling techniques are time series analysis (Min and Wynter, 2011), neural networks (Kit et al., 2012; Zhang et al., 1997), nonparametric regression (Wei et al., 2011), intelligent computation (Ming-Wei et al., 2013), stochastic models (Jabari and Liu, 2013; Wei et al., 2014) and dynamic network analysis (Du et al., 2012). In (Vlahogianni et al., 2004), modeling is generally categorized into nonparametric and parametric methods. However, there is a consensus within the literature that there is no all-embracing model or methodology. Smith and Demetsky (1994), Cassidy and Bertini (1999), Smith et al. (2002), Chrobok et al. (2004), Tan et al. (2009), and Duret et al. (2012) supply a historical view of the literature regarding short-term traffic modeling as well as the references there in. In all approaches the understanding of the data leads to the appropriate method of modeling.

In contrast, monitoring and control methods seek to achieve real-time and long-term traffic management through the utilization of synergistic technologies. Specifically, monitoring and control research has been primarily focused with and on methodologies for *Intelligent Transportation Systems* (ITS). In (Zhang et al., 2011), a comprehensive and relevant survey of the literature regarding data based developments in ITS is presented. In the literature a clear distinction can be made regarding real-time (on-line) and long-term traffic monitoring. Blandin et al. (2013) summarize that observability of quantities required for model calibration, the ability to take advantage of measurements of various traffic quantities, computational tractability and model accuracy are essential properties for traffic monitoring. In (Daganzo, 2007), a real-time model for adaptive control in an effort to improve urban mobility and congestion reduction is presented. Similarly, Ramezani and Geroliminis (2012), provide a method that seeks to understand how traffic progresses and travel time correlation behave over time, which results in their presentation of a heuristic clustering approach to determine the states of traffic within distinct neighborhoods in real-time. The primary goal of on-line monitoring of traffic flow for reasons of safety and efficiency, is to limit the traffic input into sections of high traffic density, which results in better performance measures such as travel time, emission reduction and others (Gazis and Knapp, 1971).

In contrast, the long-term traffic monitoring seeks to develop methodologies for the strategic level in which the calculated traffic state is monitored. This results in achieving a consistency of objectives for different control systems (i.e. interurban, urban, public transportation, etc.) (Friedrich and Schutte, 2000). The integration of these control systems is of high importance for the implementation of common transport policy in an efficient way for holistically optimizing the capacity of all these transportation systems. Long-term analysis continues with an effort to develop models that describe the day-to-day evolution of traffic flows. A recent example of such a model is (Parry and Hazelton, 2013), where two problems are presented that hinder the design of a computationally feasible Markov Chain Monte Carlo (MCMC) for flow prediction, which are: (1) They seek to avoid an exponential rise in complexity, which results from the growth in data, as the number of days for which data is available increases. (2) They seek to develop an efficient sampler for route flows that are constrained to match the link counts, which they proposed remains a challenging problem. In addition, Parry and Hazelton (2013) state that, these day-to-day models show considerable potential, but they tend to be heavily parameterized and hence require careful fitting for practical implementation. The development of effective traffic management requires knowledge regarding the typical daily traffic patterns, which in turn supply insight into the daily travel demand profiles from day-to-day (Weijermars and Berkum, 2005).

To this end, this paper seeks to address data dimensionality reduction and the long-term monitoring of traffic flow patterns from day-to-day to capture their evolution over time. The need for long-term monitoring is necessary to review and examine traffic evolution due to the ever-changing dynamics of growing cities. These changes can be produced by a number of reasons such as but not limited to: physical alterations to the roadway network, new roadways, geometric changes to existing roadways, new businesses or additional bus stops. Changes such as these are common in growing communities, which can experience many of these transformations over a period of three to five years (Peter et al., 2008). An example of this can be found in the U.S. Department of Transportations Traffic Signal Timing Manual which states, Retiming traffic signals every three to five years is generally good engineering practice. This frequency of signal retiming is particularly important for jurisdictions or localized areas within a jurisdiction in which land use changes lead to rapid changes in traffic patterns.

In this paper two problems are considered. The first problem is to identify different patterns in the daily traffic data. The second problem is to be able to monitor the daily traffic profile to detect changes in the traffic pattern. In this paper a traffic profile corresponds to the representation of the traffic flow over one day or 24 h period. Although the data is discretely collected, the true nature of the data is continuous. Hence, these two problems are addressed from the Functional Data Analysis (FDA) point of view. The term FDA refers to the set of statistical methods for analyzing continuous data as curves or images. The functional extension of the Principle Component Analysis (PCA) to stochastic processes can be referred to as the starting