ARTICLE IN PRESS

Transportation Research Part C xxx (2017) xxx–xxx

Contents lists available at ScienceDirect

Transportation Research Part C

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

A big data approach for clustering and calibration of link fundamental diagrams for large-scale network simulation applications

Ziyuan Gu [a], Meead Saberi [a,*], Majid Sarvi [b], Zhiyuan Liu [c]

[a] Institute of Transport Studies, Department of Civil Engineering, Monash University, VIC 3800, Australia

[b] Department of Infrastructure Engineering, University of Melbourne, VIC 3010, Australia

[c] Jiangsu Key Laboratory of Urban ITS, Jiangsu Province Collaborative Innovation Center of Modern Urban Traffic Technologies, School of Transportation, Southeast University, Nanjing 210096, China

ARTICLE INFO

Article history:
Received 9 August 2017
Received in revised form 15 August 2017
Accepted 16 August 2017
Available online xxxx

Keywords:
Link fundamental diagram
Calibration
Big traffic data
Clustering
Fréchet distance
Traffic dynamics

ABSTRACT

Existing methods for calibrating link fundamental diagrams (FDs) often focus on a limited number of links and use grouping strategies that are largely dependent on roadway physical attributes alone. In this study, we propose a big data-driven two-stage clustering framework to calibrate link FDs for freeway networks. The first stage captures, under normal traffic state, the variations of link FDs over multiple days based on which links are clustered in the second stage. Two methods, i.e. the standard k-means algorithm combined with hierarchical clustering and a modified hierarchical clustering based on the Fréchet distance, are applied in the first stage to obtain the FD parameter matrix for each link. The calibrated matrices are input into the second stage where the modified hierarchical clustering is re-employed as a static approach resulting in multiple clusters of links. To further consider the variations of link FDs, the static approach is extended by modifying the similarity measure through the principle component analysis (PCA). The resulting multivariate time-series clustering models the distributions of the FD parameters as a dynamic approach. The proposed framework is applied on the Melbourne freeway network using one-year worth of loop detector data. Results have shown that (a) similar roadway physical attributes do not necessarily result in similar link FDs, (b) the connectivity-based approach performs better in clustering link FDs as compared with the centroid-based approach, and (c) the proposed framework helps achieving a better understanding of the spatial distribution of links with similar FDs and the associated variations and distributions of the FD parameters.

1. Introduction

Urban traffic management typically requires an understanding of traffic dynamics at the network level (Zheng et al., 2016), leading to extensive investigation into characterizing spatial and temporal travel patterns for network-level traffic flow analysis. To capture the spatial-temporal traffic dynamics in large-scale networks, simulation-based dynamic traffic assignment (DTA) models are deployed which require, however, accurately calibrated demand and supply inputs (Ngoduy

* Corresponding author.
E-mail address: meead.saberi@monash.edu (M. Saberi).

http://dx.doi.org/10.1016/j.trc.2017.08.012
0968-090X/

Please cite this article in press as: Gu, Z., et al. A big data approach for clustering and calibration of link fundamental diagrams for large-scale network simulation applications. Transport. Res. Part C (2017), http://dx.doi.org/10.1016/j.trc.2017.08.012

and Maher, 2012; Hale et al., 2015; Lu et al., 2015; Paz et al., 2015; Shafiei et al., 2016, 2017). As one of the major supply inputs, link fundamental diagrams (FDs) relate primary traffic flow variables with one another. Fig. 1 gives an example of off-line calibration for the dual-regime modified Greenshields traffic flow model (Mahmassani et al., 2009) using loop detector data from a freeway section on Western Ring Road, Melbourne in April 2011. Since transportation network models are critical in understanding network-wide traffic flow dynamics over time and space, calibration of link FDs needs to further consider the spatial-temporal features of traffic flow for achieving a better simulation performance. Also with the growing availability of big traffic data from mobile and infrastructure-based sources, a unique opportunity exists to improve the existing calibration and validation methods (Ozbay et al., 2014). Despite the rapid growth in the size and number of traffic data sets, traditional traffic data management tools and mining algorithms have not been sufficiently exploited. Limited research progress has been made both theoretically (Fahad et al., 2014; Zerhari et al., 2015) and empirically (Mudigonda and Ozbay, 2014; Ozbay et al., 2014) to provide valuable insights into clustering big data for different applications.

The majority of the existing research on FD calibration focuses on a limited number of links and involves grouping strategies assuming that similar roadway physical attributes lead to similar link FDs (refer to Section 1.1 for a comprehensive literature review). A few recent studies have explored a clustering-based framework for calibrating link FDs at either the section level (Jiang and Huang, 2009; Jiang et al., 2012) or the network level (Gu et al., 2016a, 2016b). Nevertheless, as a supplementary approach to the traditional calibration methods, the big data-driven perspective has not been fully investigated at the network level particularly with regard to the spatial distribution of links with similar FDs and the associated variations and distributions of the FD parameters. Hence this study aims to address this concern and to extend the knowledge on FD calibration methods for freeway networks (refer to Section 1.2 for details of objectives and contributions).

1.1. Related literature

A vast body of literature has been devoted to the off-line calibration of link FDs for network simulation applications. Previous studies mainly focused on curve fitting using field data from a few number of days, either in normal situations (Del Castillo and Benitez, 1995; Smith et al., 1996; Leclercq, 2005) or under adverse weather conditions (Mahmassani et al., 2012; Hou et al., 2013; Kim et al., 2013). As a widely used technique for regression analysis, the least squares method (LSM) was typically employed to solve the curve fitting problem for dual- or multiple-regime traffic flow models (Dervisoglu et al., 2009; Li and Zhang, 2011). Because the LSM does not control for the sample selection bias, Qu et al. (2015) proposed a weighted least squares method (WLSM) for calibration of single-regime traffic flow models that better represents the congested regime. To improve network simulation applications, Chiu et al. (2010) further introduced the speed influencing region (SIR) for calibration of an anisotropic mesoscopic simulation (AMS) model.

We identify two limitations in these studies: (a) a steady state analysis of aggregated traffic data was applied without considering traffic flow dynamics, and (b) link FDs were assumed deterministic rather than stochastic. To overcome the first limitation, Zhong et al. (2015) proposed an automatic calibration method where a bi-level optimization problem was formulated to address the issue of data variability. The upper level aimed to minimize a merit function, i.e. the discrepancy between simulated and observed data, while the lower level was a cell transmission model (CTM). The studied freeway section was separated into different cells for simulation and calibration rather than treated as a whole, suggesting that spatial

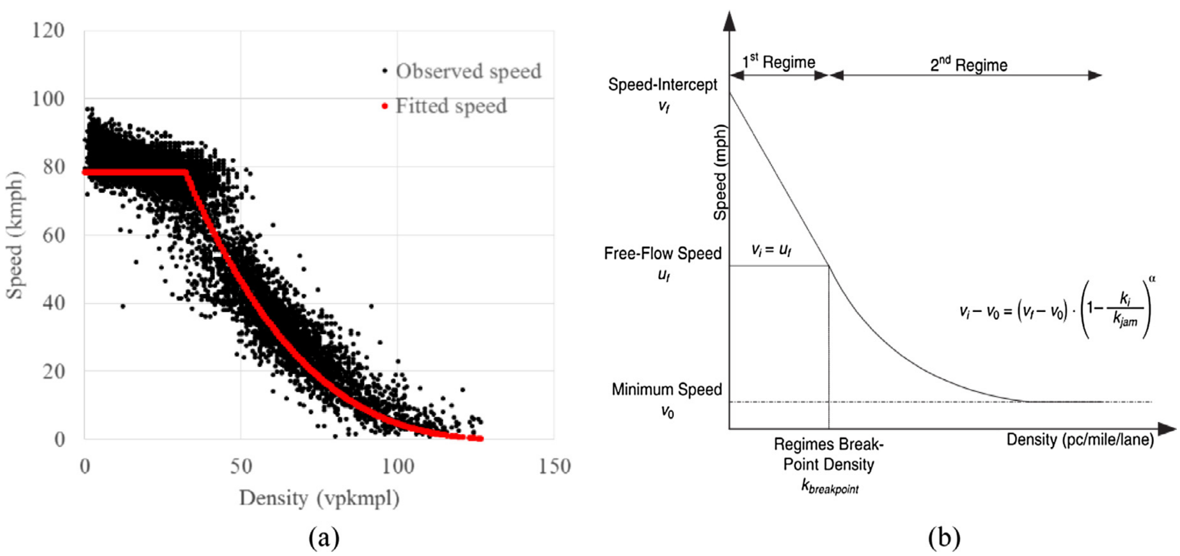

Fig. 1. The off-line calibration for a freeway section on Western Ring Road, Melbourne in April 2011: (a) fitted speed vs. observed speed; (b) the dual-regime modified Greenshields traffic flow model. Source: Hou et al. (2013).

Please cite this article in press as: Gu, Z., et al. A big data approach for clustering and calibration of link fundamental diagrams for large-scale network simulation applications. Transport. Res. Part C (2017), http://dx.doi.org/10.1016/j.trc.2017.08.012