Transportation Research Part C 96 (2018) 96–121

Contents lists available at ScienceDirect

Transportation Research Part C

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

Estimating multi-year 24/7 origin-destination demand using high-granular multi-source traffic data

Wei Ma[a], Zhen (Sean) Qian[a,b,*]

[a] Department of Civil and Environmental Engineering, Carnegie Mellon University, Pittsburgh, PA 15213, United States
[b] H. John Heinz III Heinz College, Carnegie Mellon University, Pittsburgh, PA 15213, United States

A B S T R A C T

Dynamic origin-destination (OD) demand is central to transportation system modeling and analysis. The dynamic OD demand estimation problem (DODE) has been studied for decades, most of which solve the DODE problem on a typical day or several typical hours. There is a lack of methods that estimate high-resolution dynamic OD demand for a sequence of many consecutive days over several years (referred to as 24/7 OD in this research). Having multi-year 24/7 OD demand would allow a better understanding of characteristics of dynamic OD demands and their evolution/trends over the past few years, a critical input for modeling transportation system evolution and reliability. This paper presents a data-driven framework that estimates day-to-day dynamic OD using high-granular traffic counts and speed data collected over many years. The proposed framework statistically clusters daily traffic data into typical traffic patterns using t-Distributed Stochastic Neighbor Embedding (t-SNE) and k-means methods. A GPU-based stochastic projected gradient descent method is proposed to efficiently solve the multi-year 24/7 DODE problem. It is demonstrated that the new method efficiently estimates the 5-min dynamic OD demand for every single day from 2014 to 2016 on I-5 and SR-99 in the Sacramento region. The resultant multi-year 24/7 dynamic OD demand reveals the daily, weekly, monthly, seasonal and yearly change in travel demand in a region, implying intriguing demand characteristics over the years.

1. Introduction

The increasing complexity and inter-connectivity of mobility systems call for large-scale deployment of dynamic network models that encapsulate traffic flow evolution for system-wide decision making. As an indispensable component of dynamic network models, time-dependent Origin-Destination (OD) demand plays a key role in transportation planning and management. Obtaining accurate and high-resolution time-dependent OD demand is notoriously difficult, though the dynamic OD estimation (DODE) problem has been intensively studied for decades. A number of DODE methods have been proposed, most of which aim at estimating dynamic OD demand for a typical day or even several hours on a typical day. To our best knowledge, there is a lack of research estimating dynamic OD demand for a long time period over the years. The OD demand and its behavior, though are generally repetitive in an aggregated view, can vary from day to day. The day-to-day variation of OD demand would need to be considered in estimate OD demand for a long period of many consecutive days. For example, estimating the dynamic OD demand for every 5-min in an entire year is computationally implausible using most of the existing DODE methods. In view of this, this paper presents an efficient data-driven approach to estimate time-dependent OD demand using high-granular traffic flow counts and traffic speed data collected over many years.

Dynamic OD demand represents the number of travelers departing from an origin at a particular time interval heading for a destination. It reveals traffic demand level, and is critical input for estimating and predicting network level congestion in a region. In addition, policymakers can understand the travelers' departure patterns and daily routines through the day-to-day OD demand. As a result, many Advanced Traveler Information Systems/Advanced Traffic Management Systems (ATIS/ATMS) require accurate time-

* Corresponding author at: Department of Civil and Environmental Engineering, Carnegie Mellon University, Pittsburgh, PA 15213, United States.
E-mail addresses: weima@cmu.edu (W. Ma), seanqian@cmu.edu (Z.S. Qian).

https://doi.org/10.1016/j.trc.2018.09.002
Received 6 March 2018; Received in revised form 2 September 2018; Accepted 2 September 2018
0968-090X/

dependent OD demand as an input. A tremendous number of studies estimate time-dependent OD demand using observed traffic data which includes traffic counts, probe vehicle data and Bluetooth data. Oftentimes those data collected over multiple days are taken daily average before being input to dynamic network models, which represent the average traffic pattern and OD demand on a typical day.

With the development of cutting-edge sensing technologies, many traffic data can be collected in high spatial and temporal granularity at a low cost. For example, traffic count and traffic speed for a road segment of 0.1 mile can be sensed and updated every 5 min throughout the year. This is a $12 \times 24 = 288$ dimension of counts/speed data for a single road segment on one day. Most of existing DODE methods become computationally inefficient or even implausible when dealing with large-scale networks with thousands of observed road segments and thousands of days of high dimensional data. How to efficiently obtain high-resolution OD demand on a daily basis over many years remains technically challenging. In this research, we estimate high-resolution dynamic OD demand for a sequence of many consecutive days over several years, referred to 24/7 OD demand throughout this paper.

Dynamic OD estimation (DODE) was formulated as either a least square problem or a state-space model. Cascetta et al. (1993) extended the concepts of static OD estimation problem and formulated a generalized least square (GLS) based framework for estimating dynamic OD demands. Tavana (2001) proposed a bi-level optimization framework which solves for a GLS problem in the upper level with a dynamic traffic assignment (DTA) problem in the lower level. The bi-level formulations for OD estimation problem were also discussed by Nguyen (1977), LeBlanc and Farhangian (1982), Fisk (1989), Yang et al. (1992), Florian and Chen (1995), Jha et al. (2004) for static OD demand. Zhou et al. (2003) extended the bi-level formulation to incorporate multi-day traffic data. To implement efficient estimation algorithms on real-time traffic management systems, Bierlaire and Crittin (2004) proposed a least square based real-time OD estimation/prediction framework for large-scale networks. Zhou and Mahmassani (2007), Ashok and Ben-Akiva (2000) established a state-space model for real-time OD estimation based on on-line traffic data feeds. Hazelton (2008) built a statistical inference framework using Markov chain Monte Carlo algorithm for generating posterior OD demand.

The bi-level OD estimation framework can be solved using heuristically computed gradient, convex approximation or gradient free algorithms. Yang (1995) proposed two heuristic approaches for the bi-level OD estimation problem, the iterative estimation-assignment (IEA) algorithms and sensibility-analysis based algorithm (SAB). Josefsson and Patriksson (2007) further improved the sensitivity analysis procedures adopted in SAB process. A Dynamic Traffic Assignment (DTA) simulator is also used to determine the numerical derivatives of link flows. Balakrishna et al. (2008), Cipriani et al. (2011) fitted such an estimation process into a stochastic perturbation simultaneous approximation (SPSA) framework. Lee and Ozbay (2009), Vaze et al. (2009), Ben-Akiva et al. (2012), Lu et al. (2015), Tympakianaki et al. (2015), Antoniou et al. (2015) further enhanced the SPSA based methods. Verbas et al. (2011) compared different gradient based methods to solve the bi-level formulation of DODE problem. Flötteröd et al. (2011) proposed a Bayesian framework that calibrates the dynamic OD using agent-based simulators. In addition to numerical solutions, research has been looking into computing the analytical derivatives for the lower-level formulations (Ghali and Smith, 1995; Frederix et al., 2011; Qian et al., 2012; Qian and Zhang, 2011). Other machine learning and computational technologies are also employed to enhance the efficiency of OD estimation methods (Kim et al., 2001; Kattan and Abdulhai, 2006; Huang et al., 2012; Xu et al., 2014).

The general bi-level formulation for OD estimation is proved to be non-continuous and non-convex, and thus its scalability is limited. Nie and Zhang (2008, 2010) formulated a single-level static and dynamic OD estimation framework that incorporates User Equilibrium (UE) path flows solved by the variational inequality, which is further improved by Shen and Wynter (2012) under the static cases. Recently, Lu et al. (2013) formulated a Lagrangian relaxation-based single-level non-linear optimization to estimate dynamic OD demand.

A large number of data sources are feeding to DODE methods. Zhang et al. (2008) evaluated the roles of count data, speed data and history OD data in the effectiveness of DODE. Van Der Zijpp (1997), Antoniou et al. (2004), Zhou and Mahmassani (2006), Rao et al. (2018) used automated vehicle identification (AVI) data together with flow counts to estimate dynamic OD demand. Emerging technologies such as Bluetooth (Barceló et al., 2010), mobile phone location (Calabrese et al., 2011; Iqbal et al., 2014), probe vehicles (Antoniou et al., 2006) data were also employed to estimate dynamic OD demands.

Two important issues are yet to be addressed. Firstly, many existing DODE methods (Ashok and Ben-Akiva, 2000; Josefsson and Patriksson, 2007; Nie and Zhang, 2008; Lu et al., 2013; Lu et al., 2015) require a dynamic traffic loading (DNL) process (either microscopic or mesoscopic) to endogenously encapsulate the traffic flow evolution and congestion spillover. As the DNL process requires relatively high computational budget, it can take hours to estimate dynamic OD demand on a network of thousands of links/nodes for a single day. Not only does it have hard time converging under the data fitting optimization problem, but estimating the 24/7 OD demand for several years becomes computationally impractical. The other issue is that most studies estimate OD demand for a few hours or a single day. OD demand varies from day to day, but is also repetitive to some extent. The day-to-day features of OD demand has not be taken into consideration of the DODE methods. For this reason, demand patterns that evolve daily, weekly, monthly, seasonally and yearly have not been explored, despite of high-granular data collected over many years.

In this paper, we develop a data-driven framework that estimates multi-year 24/7 dynamic OD demand using traffic counts and speed data collected over the years. The framework builds the relationship between dynamic OD demand and traffic observations using link/path indices matrix, dynamic assignment ratio (DAR) matrix, and route choice matrix. These three matrices enable the estimate framework to circumvent the bi-level formulation, since each of the matrices can be directly calibrated using high-granular real-world data rather than from complex simulation. The proposed framework utilizes data-driven approaches to explore the daily, weekly, monthly and yearly traffic patterns, and group traffic data into different patterns. The proposed estimation framework is computational efficient: 5-min dynamic OD demand for three years can be estimated within hours on an inexpensive personal computer.

In order to address computation issues, this paper uses a Graphics Processing Unit (GPU) which is currently attracting tremendous