Link travel time inference using entry/exit information of trips on a network

CrossMark

Kai Yin [a], Wen Wang [b], Xiubin Bruce Wang [b,*], Teresa M. Adams [c]

[a] Nomis Solutions, Inc., United States
[b] Zachry Department of Civil Engineering, Texas A&M University, United States
[c] Department of Civil and Environmental Engineering, University of Wisconsin-Madison, United States

A R T I C L E   I N F O

A B S T R A C T

This paper studies link travel time estimation using entry/exit time stamps of trips on a steady-state transportation network. We propose two inference methods based on the likelihood principle, assuming each link associates with a random travel time. The first method considers independent and Gaussian distributed link travel times, using the additive property that trip time has a closed-form distribution as the summation of link travel times. We particularly analyze the mean estimates when the variances of trip time estimates are known with a high degree of precision and examine the uniqueness of solutions. Two cases are discussed in detail: one with known paths of all trips and the other with unknown paths of some trips. We apply the Gaussian mixture model and the Expectation–Maximization (EM) algorithm to deal with the latter. The second method splits trip time proportionally among links traversed to deal with more general link travel time distributions such as log-normal. This approach builds upon an expected log-likelihood function which naturally leads to an iterative procedure analogous to the EM algorithm for solutions. Simulation tests on a simple nine-link network and on the Sioux Falls network respectively indicate that the two methods both perform well. The second method (i.e., trip splitting approximation) generally runs faster but with larger errors of estimated standard deviations of link travel times.

© 2015 Elsevier Ltd. All rights reserved.

1. Introduction

Travel time is one of the most important factors when a traveler plans a route from an origin to a destination, and is also critical to transportation planners and operators as a performance measure. Accurate travel time estimation on a transportation network is therefore becoming an essential task and is made possible now by widely available traffic data.

Travel time data on a network is regularly obtained from traffic tracking such as through probing phones (Bar-Gera, 2007; Ygnace et al., 2000), global positioning system (GPS) devices (Bertini and Tantiyanugulchai, 2004), and vehicle ID readers (through Bluetooth or vehicle plate identification and matching, e.g., Haghani et al. (2010) and Chang et al. (2004)). While the associated methods to estimate roadway travel time range from regression models (Chan et al., 2009), machine learning approaches (Zheng and Zuylen, 2013), and to analytical models dealing with traffic conditions (Hellinga et al., 2008), many

required parameters limit their applicability in practice and there is a lack of general model approach to the network-wide travel time estimation problem.

Valid statistical analysis becomes increasingly important as the data becomes widely available (Fan et al., 2014). In what follows, we will briefly review two categories of statistical approaches in relevant literature: the traditional maximum-likelihood method and the Bayesian approach.

Among the scant literature that focuses directly on this topic, Hunter et al. (2009) formulate a maximum-likelihood problem to estimate link travel time distributions on an arterial network. Their model considers the observations of unknown trajectories. They present an Expectation–Maximization (EM) algorithm to simultaneously learn the likely paths of probe vehicles as well as the travel time distributions on the network. They assume that travel times on different links are independent and briefly conduct numerical tests using San Francisco taxi data. Instead of the assumption of independent link travel times, Jenelius and Koutsopoulos (2013) present a statistical model for travel time estimation on an urban road network considering the correlation between travel times on different links. They capture the correlation using a moving average specification for link travel times. The specific information of link attributes (such as speed limit and roadway functional class) and trip conditions (such as day-of-week, time-of-day, and weather condition) are incorporated as explanatory variables. The model is estimated using maximum-likelihood method, and is applied to a particular route on the Stockholm network in Sweden.

In contrast to the traditional maximum-likelihood method, some studies apply the Bayesian approach to travel time distribution prediction. Hofleitner et al. (2012a,b) propose a dynamic Bayesian network for unobserved traffic conditions and model link travel time distributions conditional on traffic state. Their method is from the conventional traffic flow perspectives, and is applied to a San Francisco road network to predict travel times using taxi data. Westgate et al. (2013) also propose a Bayesian model to estimate the distribution of ambulance travel times on road segments in Toronto. They apply a multinomial Logit model to formulate the path choices for ambulance trips, and perform the path inference and travel time estimation simultaneously using a Bayesian approach. They also assume that link travel times are independent and log-normally distributed. The parameters are estimated using Markov Chain Monte Carlo (MCMC) methods. Instead of modeling travel times at the link level in the previous work, Westgate et al. (2013) model ambulance travel times at the trip level. They propose a regression approach for estimating the ambulance travel time distribution on an arbitrary route, and use a Bayesian formulation to estimate the model parameters. The advantage of applying the Bayesian approach is that it utilizes expert knowledge as prior information, and tackles many complicated problems that traditional statistical approaches find difficult to analyze. However, its implementation relies on computationally expensive methods such as the MCMC.

This research aims to develop inference methods for link travel time estimation on a steady state network, assuming that each link is associated with a random travel time due to different traveling vehicles and variation of day time. We believe the statistical characteristics of the link travel time as well as the distribution can be measured approximately. We estimate network-wide link travel times by only using vehicle start and end locations and time of trips, referred to as traveler entry/exit time stamps in this paper. This type of data is representative of the data available when discrete points of a trip are recorded. Sparse vehicle trajectories reported by GPS-equipped probe vehicles or smart phones (Wang et al., 2014) can also be regarded as a particular case of traveler entry/exit trip information on a network. By trip we mean sequential records that denote the time stamps for start and end nodes on a roadway network. For example, two consecutive GPS records are considered as a trip in this study, though they essentially represent a trip segment. Specially, this research is motivated by a practical application on a toll road network, in which traveler entry/exit time stamps are recorded at tollbooths and the toll road authority has a practical need to use the travel time inference results to evaluate the toll systems. Other potential applications include using public transit data for network performance analysis when passenger entry/exit information is recorded at fare boxes (Ma et al., 2013).

We start with the assumption of independent and normally distributed link travel times, and present the EM algorithm to address the trips with unknown paths, as in Hunter et al. (2009) and Siripirote et al. (2013). However, our study differs from this earlier work by focusing on exploring the analytical properties of the proposed methods. We examine the impact of errors in trip variance estimates on mean link travel time estimates, and investigate the uniqueness of solutions in the algorithm. We also provide confidence intervals for mean link times.

Furthermore, we propose a statistical method of trip splitting approximation, as Method II in this paper, to mainly address a technical situation in which the summation of random link travel times for a path does not have a closed-form probability distribution. We assume that trips on the same path under similar traffic conditions (e.g., different vehicles are measured within a time interval of 30 min or so) have approximately a constant proportion of the trip time for a traversed link. Similar idea of decomposing trip travel time has already been proposed in practical applications (Hellinga et al., 2008), but without appropriate justification and investigation. The proposed trip splitting approximation method may be applied to arbitrary distributions, and is statistically justified for the network estimation problem. Our numerical tests indicate that this method is computationally cheaper, though it may lead to relatively larger errors for link time standard deviation estimates. Its potential application would be more promising if more traffic information is available.

The remainder of this paper is organized as follows. Section 2 proposes the first method assuming that the trip time has a closed-form distribution, using Gaussian distribution for link travel time as an example. Two cases are discussed in detail: one with known paths of all trips and the other with unknown paths of some trips. Section 3 develops a statistical framework of the trip splitting method. Section 4 tests the proposed methods with simulated data on a simple nine-link test network