Transportation Research Part C 65 (2016) 16–30

Contents lists available at ScienceDirect

Transportation Research Part C

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

Reconstructing maximum likelihood trajectory of probe vehicles between sparse updates

Nianfeng Wan [a,*], Ardalan Vahidi [a], Andre Luckow [b]

[a] Department of Mechanical Engineering, Clemson University, Clemson, SC 29634-0921, United States
[b] BMW Group Information Technology Research Center, Greenville, SC, United States

ARTICLE INFO

Article history:
Received 20 February 2015
Received in revised form 17 December 2015
Accepted 2 January 2016

Keywords:
Probe vehicular data
Expectation Maximization
Maximum likelihood
Trajectory estimation
Transit bus

ABSTRACT

Data from connected probe vehicles can be critical in estimating road traffic conditions. Unfortunately, current available data is usually sparse due to the low reporting frequency and the low penetration rate of probe vehicles. To help fill the gaps in data, this paper presents an approach for estimating the maximum likelihood trajectory (MLT) of a probe vehicle in between two data updates on arterial roads. A public data feed from transit buses in the city of San Francisco is used as an example data source. Low frequency updates (at every 200 m or 90 s) leaves much to be inferred. We first estimate travel time statistics along the road and queue patterns at intersections from historical probe data. The path is divided into short segments, and an Expectation Maximization (EM) algorithm is proposed for allocating travel time statistics to each segment. Then the trajectory with the maximum likelihood is generated based on segment travel time statistics. The results are compared with high frequency ground truth data in multiple scenarios, which demonstrate the effectiveness of the proposed approach, in estimating both the trajectory while moving and the stop positions and durations at intersections.

1. Introduction

Traditionally costly infrastructure and dedicated sensors have been utilized to collect vehicular traffic data (Soriguera and Robusté, 2011; Singh and Li, 2012). But in recent years data from probe vehicles have provided another viable method for estimating traffic conditions (Herrera et al., 2010; Hofleitner et al., 2012a; Goodall et al., 2012). Probe vehicles are those equipped with Global Positioning System (GPS) devices on board that periodically report their coordinates and other features such as velocity and heading. Traditional loop detector sensors, for instance, can only provide flow or occupancy at fixed locations; probe vehicles on the other hand, can provide traffic information samples at varying locations.

Nowadays the technologies of wireless communication and cloud storage enable collection of probe data more efficiently. But existing probe data sets are spatiotemporally sparse. Higher penetration rate of probe vehicles and higher reporting frequencies are needed for accurate traffic estimation or vehicle control purposes, but are unlikely in the near future. High reporting frequency also raises privacy concerns (Hoh et al., 2007, 2012). Moreover, current available high frequency vehicle trajectory data usually contains large measurement errors and other errors as well. Researchers propose several methods to refine trajectories (Montanino and Punzo, 2013; Punzo et al., 2011). On the other hand, due to their low reporting frequency, current probe data, in raw form, can provide only a very incomplete picture of traffic on the road. But if vehicle trajectories between two sparse updates could be effectively reconstructed, additional virtual data points are generated that perhaps

* Corresponding author.

http://dx.doi.org/10.1016/j.trc.2016.01.010

help more accurate evaluation of road traffic. One application is shown in Lu and Skabardonis (2007) where the authors use high frequency data to conduct traffic shockwave analysis. This paper proposes a method for reconstruction of vehicle trajectories between probe updates based on historical data and current traffic information such as signal timing.

There are a number of papers that address vehicular trajectory reconstruction (Huang and Tan, 2006; Mehran et al., 2012; Ni and Wang, 2008; Sun and Ban, 2013; Hao et al., 2014). For example in Huang and Tan (2006), the authors proposed a method for short term prediction of a vehicle trajectory based on the usage of nearby vehicles information. In Sun and Ban (2013) trajectories are reconstructed based on the variational formulation of kinematic waves, and results have been tested with NGSIM data and microsimulation data. In Hao et al. (2014), the authors focus on finding the most likely driving mode sequences (deceleration, idle, acceleration, and cruise), to estimate a vehicle's trajectory. The approach in Sun and Ban (2013) is macroscopic while (Hao et al., 2014) employ a microscopic-based approach.

In this paper we employ a different probabilistic microscopic-based approach, and use segment travel time statistics to reconstruct vehicle trajectories. Using historic probe data, we first estimate the travel time statistics across short road segments and also queue patterns at intersections. We then employ a maximum likelihood approach to generate the most likely trajectory of a probe vehicle between consecutive GPS updates.

Because travel time is an important measure of traffic conditions, a number of papers have focused on travel time estimation or prediction for freeways (Wan et al., 2014) and arterials (Hofleitner et al., 2012b; Sun et al., 2008; Hellinga et al., 2008; Zheng and Van Zuylen, 2013). An example of a probabilistic approach to travel time estimation can be found in Hofleitner et al. (2012a). In Coifman (2002) it is shown that travel time can be used in estimating vehicle trajectories.

Most of these existing papers estimate link travel times; the limitation is the implicit assumption of uniform distribution of travel time along an entire link. However, travel time along an arterial road may not be uniformly distributed, and for instance is higher near signalized intersections. To capture this variability, in Wan and Vahidi (2014), we proposed to divide each link to short segments of equal length and estimated statistics of travel time for each segment by using an Expectation Maximization algorithm.

In this paper we build on our earlier results reported in Wan and Vahidi (2014, 2015) to estimate the most likely trajectory of a probe vehicle between two consecutive updates. Presence of signalized intersections between the two updates complicates the problem and is influenced by factors such as queue size, signal timing and phase (Liu et al., 2009; Hao and Sun, 2011). In this paper we present solutions for when the updates span a single or multiple intersections. We evaluate the proposed algorithms using sparse updates from transit buses in the city of San Francisco. We demonstrate the effectiveness of the proposed method in comparison with high frequency data obtained from ground truth measurements of those same buses.

The rest of the paper is organized as follows: Section 2 describes the bus data feed. Section 3 explains estimation of segmental travel time statistics. Section 4 outlines our proposed method for estimating the most likely trajectory of a vehicle including the cases where the vehicle comes to a stop at an intersection queue. Section 5 presents the results as compared to ground truth measurements followed by conclusions in Section 6.

2. Description of the dataset

In this paper, we use a public data feed of transit buses in the city of San Francisco. The data contains GPS time stamp, longitude and latitude, velocity, heading and several other attributes of transit buses and is provided by NextBus. NextBus provides real-time passenger information for over 135 transit agencies and organizations in North America (NextBus, 1-15-2015). The data can be queried in almost real-time in eXtensible Markup Language (XML) interface using URLs with parameters specified in the query string.

Fig. 1 shows aggregated GPS updates from all buses in the city of San Francisco within a twenty-four hour period. As shown in the figure the coverage includes most major streets. Also the fact that buses traverse each route regularly is an advantage of using them as probe vehicles. However, these updates are sparse; at every 200 m or 90 s whichever comes first. Moreover, buses stop not only at intersections but also at bus stops which complicates the trajectory estimation problem. The bus data was among the very few publicly available data feeds that we could find and therefore was used to verify the effectiveness of our proposed algorithms.

3. Estimation of segment travel time statistics

In order to reconstruct the most likely path of a vehicle between its two updates, we first estimate the statistics of travel time for each road segment relying on historical probe data. Two successive updates of a probe vehicle provides a travel time observation. As mentioned before, most existing papers offer methods for estimating "link travel time", which is the travel time between two adjacent intersections. Their implicit assumption is that travel time is uniformly distributed along a link, which is not true in most cases. Moreover, probe vehicle updates normally occur at random positions and times and not necessarily at the two ends of a link. To address these aforementioned issues, we propose to divide a whole path into short segments, and to allocate a travel time to each segment based on probe data.

Using the haversine formula, the reported longitude and latitude coordinates are converted to a linear distance measured from an arbitrary reference point at the upstream end. Let's denote each of such segments by $x_i,\ i = 1, 2, \ldots, N$. All segments