Transportation Research Part C 46 (2014) 132–150

Contents lists available at ScienceDirect

Transportation Research Part C

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

A trade-off analysis between penetration rate and sampling frequency of mobile sensors in traffic state estimation

Christopher Bucknell [a,*,1], Juan C. Herrera [b,1]

[a] Center for Sustainable Urban Development (CEDEUS), Pontificia Universidad Católica de Chile, Chile
[b] Department of Transport Engineering and Logistics, Pontificia Universidad Católica de Chile, Chile

ARTICLE INFO

Article history:
Received 9 January 2014
Received in revised form 15 May 2014
Accepted 15 May 2014

Keywords:
Mobile sensors
Traffic monitoring
Data collection

ABSTRACT

The rapid-growth of smartphones with embedded navigation systems such as GPS modules provides new ways of monitoring traffic. These devices can register and send a great amount of traffic related data, which can be used for traffic state estimation. In such a case, the amount of data collected depends on two variables: the penetration rate of devices in traffic flow (P) and their data sampling frequency (z). Referring to data composition as the way certain number of observations is collected, in terms of P and z, we need to understand the relation between the amount and composition of data collected, and the accuracy achieved in traffic state estimation. This was accomplished through an in-depth analysis of two datasets of vehicle trajectories on freeways. The first dataset consists of trajectories over a real freeway, while the second dataset is obtained through microsimulation. Hypothetical scenarios of data sent by equipped vehicles were created, based on the composition of data collected. Different values of P and z were used, and each unique combination defined a specific scenario. Traffic states were estimated through two simple methods, and a more advanced one that incorporates traffic flow theory. A measure to quantify data to be collected was proposed, based on travel time, number of vehicles, penetration rate and sampling frequency. The error was below 6% for every scenario in each dataset. Also, increasing data reduced variability in data count estimation. The performance of the different estimation methods varied through each dataset and scenario. Since the same number of observations can be gathered with different combinations of P and z, the effect of data composition was analyzed (a trade-off between penetration rate and sampling frequency). Different situations were found. In some, an increase in penetration rate is more effective to reduce estimation error than an increase in sampling frequency, considering an equal increase in observations. In other areas, the opposite relationship was found. Between these areas, an indifference curve was found. In fact, this curve is the solution to the optimization problem of minimizing the error given any fixed number of observations. As a general result, increasing sampling frequency (penetration rate) is more beneficial when the current sampling frequency (penetration rate) is low, independent of the penetration rate (sampling frequency).

* Corresponding author.
E-mail addresses: cpbuckne@uc.cl (C. Bucknell), jch@ing.puc.cl (J.C. Herrera).
[1] Address: Vicuña Mackenna 4860, Macul, Santiago, Chile. Tel.: +562 2 354 4270; fax: +562 2 553 0281.

http://dx.doi.org/10.1016/j.trc.2014.05.007
0968-090X/

1. Introduction

The arrival of mobile internet and the incorporation of satellite navigation systems (such as GPS[2] and GLONASS[3]) into smartphones has enabled a new approach for monitoring and estimating traffic, since these devices are able to collect and send traffic data. Unlike infrastructure-based technology, the use of smartphones as traffic sensors provides a good spatial and temporal coverage of the transportation network at a relatively low cost, since the cost of the devices are assumed by the user and information systems are of little cost (Sanwal and Walrand, 1995). For these reasons, traffic monitoring and traffic state estimation based on data provided by probe vehicles has been analyzed in the context of freeways (Herrera and Bayen, 2007, 2010; Herrera et al., 2010; Work et al., 2008; Nanthawichit et al., 2003) and urban networks (Sun and Ban, 2013; Herring et al., 2010; Hofleitner et al., 2012; Ban et al., 2011; Jenelius and Koutsopoulos, 2013; Feng et al., 2014).

In spite of the clear benefits this technology provides, there are some inherent downsides in using it to monitor or to estimate traffic (Herrera et al., 2010). First, not every vehicle on the road acts as a probe vehicle because not all of them are equipped with this technology and able to send data. Second, energy consumption rises as the navigation system is turned on all the time. Moreover, the frequency with which data are sent (sampling frequency) also affects the energy consumption, as analyzed by Frank et al. (2012). The sampling frequency also relates to internet data consumption, since this could be of cost for the user, or could be within a monthly data limit. Finally, because of the nature of this technology, it could be privacy-invasive. This is an important topic to address in order to achieve a high participation rate in the system (Rose, 2006).

Because of the previous downsides, the amount of data received from probe vehicles is limited, and essentially depends on:

(i) The penetration rate of devices in the traffic flow: that is, how many vehicles act as probe vehicles and send data. This rate not only depends on the penetration of this technology in the population, but also in the people's willingness to share their data. Smartphones in the population have been increasing steadily across the globe, with an estimate of 63% of total new handset sales in the USA and 37% globally for 2014 (GSMA, 2011). In Chile, smartphones connections have increased by 50% (from 2.6 to 4 MM) during 2012, achieving a penetration rate of 22.8 smartphones per 100 inhabitants (Subsecretaría de Telecomunicaciones, 2013). At the same time, the popular crowdsourcing application and social network Waze has 1.9 MM users in Chile (Ministerio de Transportes y Telecomunicaciones, 2013).

(ii) The strategy and frequency of data sampling: that is, the way in which data are sampled and how frequently each probe vehicle sends data. For the reasons stated before, probe vehicles do not send data continuously. The way in which they send data (i.e. data are sampled) is called sampling strategy, and there are at least three different strategies (Herrera, 2009; Mohan et al., 2008; Nanthawichit et al., 2003): temporal, spatial and event-based sampling strategy. The temporal sampling strategy sends data frequently in time. The time elapsed between two consecutive packages of data is called the *sampling interval*. Spatial sampling sends data at specific geographical locations, which might be of interest. Examples of this are the Virtual Trip Lines or VTLs (Hoh et al., 2012), which also help to preserve privacy. Finally, in the event-based or triggered sampling data are sent after a certain action is performed, such as braking (detected by an accelerometer) or a horn sound (detected by a microphone).

This article will focus on the effects of the penetration rate and sampling frequency on traffic state estimation, specifically to reconstruct the velocity field. Assuming a temporal sampling, penetration rate and sampling intervals can take different values, yielding different numbers of total data sent. It is of interest to thoroughly understand the effect of different penetration rates and sampling intervals in the accuracy of traffic state estimates. Herrera (2009) presents preliminary evidence suggesting that sporadically sampling more vehicles is preferred to a more frequently sampling of fewer vehicles. However, as stated by the author, further analysis needs to be conducted in order to be conclusive in this matter.

Trade-offs between inductive loops and probe vehicles has been addressed previously for the purpose of travel time estimation (Mazaré et al., 2012), but no analyses on the trade-off between penetration rate and sampling frequency (the inverse of the sampling interval) was found in the literature. Studies have used different penetration and sampling frequencies (Piccoli et al., 2012), but using relatively small sampling intervals (from 1 to 3 s) and focusing on error degradation and model calibration, rather than the trade-off between variables.

The method used to estimate traffic states is of relevance. Its performance depends on the amount of data sent by probe vehicles and also how these data are incorporated in the model. To the best of our knowledge, only one study investigates the benefits of using advanced traffic models to estimate traffic states in contrast to simpler models in relation to penetration rate, but not sampling frequency (Work et al. (2008) uses ten VTLs and only varies penetration rate). It is expected that above certain amounts of data provided by probe vehicles, advanced models would not be needed to produce accurate estimates. It would also be useful to quantify the performance of the method for scenarios in which the amount of data is below the previous threshold.

[2] Global Positioning System.

[3] *Globalnaya Navigatsionnaya Sputnikovaya Sistema* or Global Navigation Satellite System.