Transportation Research Part C 58 (2015) 224–239

Contents lists available at ScienceDirect

Transportation Research Part C

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

Quantifying transit travel experiences from the users' perspective with high-resolution smartphone and vehicle location data: Methodologies, validation, and example analyses

Andre Carrel [a,*], Peter S.C. Lau [b], Rabi G. Mishalani [c], Raja Sengupta [a], Joan L. Walker [a]

[a] University of California at Berkeley, Department of Civil and Environmental Engineering, United States
[b] San Francisco, CA, United States
[c] The Ohio State University, Department of Civil, Environmental and Geodetic Engineering, United States

ARTICLE INFO

Article history:
Received 9 June 2014
Received in revised form 19 January 2015
Accepted 13 March 2015
Available online 16 April 2015

Keywords:
Public transportation
Automatic data collection
Automatic vehicle location systems
Smartphone location tracking
GPS
Spatial data matching
Data mining
Reliability
Travel time variability
User-centric performance metrics
Origin–destination matrices
Mode detection
Transit travel diaries
Dynamic time warping
Travel time decomposition
Passenger trajectories
Underground trip detection

ABSTRACT

While transit agencies have increasingly adopted systems for collecting data on passengers and vehicles, the ability to derive high-resolution passenger trajectories and directly associate them with transit vehicles in a general and transferable manner remains a challenge. In this paper, a system of integrated methods is presented to reconstruct and track travelers usage of transit at a detailed level by matching location data from smartphones to automatic transit vehicle location (AVL) data and by identifying all out-of-vehicle and in-vehicle portions of the passengers trips. High-resolution travel times and their relationships with the timetable are then derived. Approaches are presented for processing relatively sparse smartphone location data in dense transit networks with many overlapping bus routes, distinguishing waits and transfers from non-travel related activities, and tracking underground travel in a Metro network. The derived information enables a range of analyses and applications, including the development of user-centric performance measures. Results are presented from an implementation and deployment of the system on San Francisco's Muni network. Based on 103 ground-truth passenger trips, the detection accuracy is found to be approximately 93%. A set of example applications and findings presented in this paper underscore the value of the previously unattainable high-resolution traveler-vehicle coupled movements on a large-scale basis.

1. Motivation and background

Tracking transit passengers as they travel through a transit network can generate data that are useful for numerous applications. In particular, one of the main inputs into long-range planning is the origin–destination (OD) matrix, which is often created using combinations of survey data and automatically collected data. Furthermore, researchers and practitioners have become interested in calculating passenger travel times from automatically collected data with the goal of deriving user-based reliability and performance metrics to complement supply-side metrics that are currently in use. This ties in with

* Corresponding author at: 109 McLaughlin Hall, Berkeley, CA 94720, United States. Tel.: +1 510 642 3585.
E-mail address: acarrel@alum.mit.edu (A. Carrel).

http://dx.doi.org/10.1016/j.trc.2015.03.021

research efforts in travel demand modeling, where it has been recognized that travel time reliability, and in particular an individual's *personal experience* with unreliability in the past, plays a major role in travelers' decisions (Fosgerau and Karlström, 2010; Hensher et al., 2003; Li et al., 2010; Benezech and Coulombel, 2013). So far, the data that have been available for these applications were from automatic data collection systems operated by the agency, most notably automatic vehicle location (AVL), fare collection (AFC) and passenger count (APC) data.

The collection of high-resolution, individual travel time data can be challenging. Trips on transit tend to be complex, involving access and egress, wait times and transfers in addition to in-vehicle segments. Passenger-centric measures of travel time distribution are a combination of the travel times experienced on these segments; the overall distribution resulting from the convolution of these individual distributions can be quite complex (Bates et al., 2001). For the aforementioned applications, being able to capture all these segments of the passenger's trip would be highly beneficial: it would afford a better picture of the users' typical *overall* travel experience and of how the in-vehicle and out-of vehicle travel times compare with the automobile. In the long run, sampling and quantifying the overall travel times of transit users on a large scale can lead to the development of new reliability metrics that better integrate the passengers' perspective.

In fully gated systems, where the fare card needs to be tapped both upon entry and exit from the system, the time spent in the system can be derived, but typically there is limited to no information on out-of-vehicle trip segments. In systems that are not fully gated and do not require passengers to tap their fare cards upon exiting, information on alighting and transfer stops is missing. In open systems without fare gates, such as most bus systems, even the information on boarding stops may not be exact. To further complicate matters, in some systems pass holders are not required to tap their fare card. The incompleteness of stop information for trips in such systems makes determining transit OD matrices and deriving travel time distributions challenging, as one has to rely on inferences and limit oneself to the trip components that are observed.

To derive either passenger-focused reliability metrics or transit OD matrices, passenger trips have to be assigned to transit routes, stops and, if possible, vehicle runs. There is a spectrum of methods published in the literature that make use of AVL, AFC and APC data, in various combinations. As is summarized by Zhao et al. (2007), who uses AFC data with entry tags only, a common assumption that is made when only entry data are available is that the stop where passengers board on one trip is the stop where they alighted on the previous trip. A selection of further related work with AFC data from gated rail systems is by Cui (2006), Chan (2007) and Rahbee (2008). Chapleau et al. (2008) and Chu and Chapleau (2010) used AFC data augmented with a geographical information system to derive transit origin–destination matrices. Yuan et al. (2013) works on a similar problem, but from the perspective of tracking individual mobility behavior through smart card transactions. Using smart card data from a fully gated system, (Sun and Xu, 2012) develop a model to infer the various travel time components on an underground network based on travel time distributions. Several authors (including Farzin et al., 2008; Nassir et al., 2011; Wang et al., 2011; Munizaga et al., 2012; Gordon et al., 2013) have also focused on connecting passenger trips from bus AFC data to vehicle locations observed via AVL data in an effort to better infer boarding locations and times. These were cases where the fare card reader was on the transit vehicle and did not directly record the boarding stop. Frumin and Zhao (2012) used AFC and AVL data in a gated system to infer rail platform wait times and (Seaborn et al., 2009) examined distributions of transfer times between a fully gated rail system and a bus system to distinguish pure transfer times from activities carried out at the transfer location.

While these aforementioned contributions have been very valuable, they have in common that due to the coarse resolution of the data, researchers could not obtain exact measurements of every travel time component, including out-of-vehicle travel times, especially in the case of bus travel. Work that attempted to disaggregate total travel time into its individual components did so primarily based distributions of total travel times. However, thanks to smartphones and other location-enabled devices, it is becoming increasingly feasible to collect individual-level location data over long periods of time and with low respondent burden; with the help of these data, the shortcomings noted above can in many cases be overcome, which is most valuable in ungated systems. Smartphone location data allow a high resolution view of individual trips as long as the traveler remains above ground, including out-of-vehicle segments and exact information on stops. As location-aware devices become ubiquitous, and given the scalability of such systems, planners can find themselves in possession of very large amounts of location data from which user-based performance metrics, personal travel experiences for demand modeling and transit OD matrices can be generated. In addition, such data allow the observation of the *true* origin and destination of a trip.

There has been previous work that utilized passenger smartphone location data, but so far it has mainly been focused on determining the travel mode from data collected through location and other sensors (e.g., accelerometer, microphone). This has been performed by map-matching the location points in a GIS system (Chung and Amer, 2005; Gong et al., 2012; Jariyasunant, 2012), by extracting features from location and accelerometer data related to velocity, acceleration or distance traveled and using those as inputs for mode classification algorithms (Zheng, 2008; Gonzalez, 2008; Parlak et al., 2012) or a combination of the two (Biagioni et al., 2009; Thiagarajan, 2010. Stenneth (2011) included AVL data and extracted the proximity of the phone to a transit vehicle and to transit stops as a classification feature (on a point-by-point basis). Other authors have worked on crowd-sourcing transit arrival times where no AVL data are available. These applications rely on real-time detection of when a smartphone owner is on board transit. Notably, this has been done by Zhou et al. (2012) based on microphone, accelerometer and coarse location, and by Kostakos et al. (2013) based on Bluetooth sensors. Lastly, Barbeau et al. (2010) developed a personal transit travel assistant for cognitively disabled riders and note that AVL data were used in their system, but they do not specify the exact role of those data.