Transportation Research Part C 95 (2018) 731–747

Contents lists available at ScienceDirect

Transportation Research Part C

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

A robust method for estimating transit passenger trajectories using automated data

Pramesh Kumar[a], Alireza Khani[a,*], Qing He[b]

[a] Department of Civil, Environmental and Geo-Engineering, University of Minnesota, Twin Cities, United States
[b] Department of Civil, Structural and Environmental Engineering, University at Buffalo, The State University of New York, United States

ARTICLE INFO

Keywords:
Automatic Fare Collection (AFC)
General Transit Feed Specification (GTFS)
Transit origin-destination (O-D) Matrix
Transit
Trip chaining algorithm
Smart card data

ABSTRACT

Development of an origin-destination demand matrix is crucial for transit planning. The development process is facilitated by automated transit smart card data, making it possible to mine boarding and alighting patterns on an individual basis. This research proposes a novel trip chaining method which uses Automatic Fare Collection (AFC) and General Transit Feed Specification (GTFS) data to infer the most likely trajectory of individual transit passengers. The method relaxes the assumptions on various parameters used in the existing trip chaining algorithms such as transfer walking distance threshold, buffer distance for selecting the boarding location, time window for selecting the vehicle trip, etc. The method also resolves issues related to errors in GPS location recorded by AFC systems or selection of incorrect sub-route from GTFS data. The proposed trip chaining method generates a set of candidate trajectories for each AFC tag to reach the next tag, calculates the probability of each trajectory, and selects the most likely trajectory to infer the boarding and alighting stops. The method is applied to transit data from the Twin Cities, MN, which has an open transit system where passengers tap smart cards only once when boarding (or when alighting on pay-exit buses). Based on the consecutive tags of the passenger, the proposed algorithm is also modified for pay-exit cases. The method is compared to previous methods developed by the researchers and shows improvement in the number of inferred cases. Finally, results are visualized to understand the route ridership and geographical pattern of trips.

1. Introduction

For better service and planning, transit agencies need to understand passengers' travel behavior. For this purpose, they conduct on-board surveys which collect data about passengers' boarding and alighting location, purpose of travel, etc., and then use expansion factors to expand the survey data for the whole population. There are various limitations associated with these surveys, such as cost, small sample size, bias, and other general reporting errors (Attanucci and Wilson, 1981). Conversely, automated data collection systems (ADCS), which are designed for administrative purposes such as revenue management, provide a rich source of information about passengers travel pattern on an individual basis. The automated data offers several advantages (Wang et al., 2011) over traditional surveys by:

1. providing a link to passenger's trips over a longer period of time

* Corresponding author.
E-mail address: akhani@umn.edu (A. Khani).

https://doi.org/10.1016/j.trc.2018.08.006
Received 9 February 2018; Received in revised form 13 August 2018; Accepted 14 August 2018
0968-090X/

2. providing information about the share of different transit commuters (e.g. students, workers, etc.)
3. storing the information in SQL database systems and using it efficiently
4. providing various research opportunities for analyzing passengers' travel pattern

In recent years, there has been growing interest in using automated smart card data for travel behavior research in transit systems. Automatic Fare Collection (AFC) systems collect information about on-board transaction of passengers such as boarding stop/station, date and time of the transaction, route information, etc. The data is useful not only for improving day-to-day transit operations but also for long-term strategic planning of transit network (Pelletier et al., 2011). It has been used for a variety of purposes such as:

1. stop-level origin-destination matrix estimation (Barry et al., 2007; Trépanier et al., 2007; Zhao et al., 2007; Alfred Chu and Chapleau, 2008; Barry et al., 2009; Chu and Chapleau, 2010; Wang et al., 2011; Nassir et al., 2011; Munizaga and Palma, 2012; Gordon et al., 2013).
2. trip purpose inference (Lee and Hickman, 2014; Kusakabe and Asakura, 2014; Alsger et al., 2018)
3. route choice modeling (Kim et al., 2017; Zhao et al., 2017)
4. passenger trip prediction (Zhao et al., 2018)
5. mining spatial and temporal clusters of similar travel patterns (Ma et al., 2013; Briand et al., 2017; Khani, 2018)
6. passenger waiting time estimation (Ingvardson et al., 2018)

This study focuses on one of the important input for analyzing a public transit system, which is the flow of passengers between different stations/stops known as an origin-destination (O-D) matrix. O-D estimation using automated smart card data has attracted attention of many researchers over the last decade (Barry et al., 2007; Trépanier et al., 2007; Zhao et al., 2007; Alfred Chu and Chapleau, 2008; Farzin, 2008; Barry et al., 2009; Chu and Chapleau, 2010; Nassir et al., 2011; Wang et al., 2011; Ma et al., 2012; Munizaga and Palma, 2012; Gordon et al., 2013; He and Trépanier, 2015). The estimation requires a sequence of trips made by the passenger throughout the day recorded using AFC system. But the information available with this data is limited and the full sequence of trips is usually not available. This is because of the type of the fare collection system (open or closed) employed by a transit agency. In closed transit systems (Alsger et al., 2016), origin and destination is known for the trips as passengers tap their card both when boarding as well as when alighting, whereas in open transit systems (Barry et al., 2007; Trépanier et al., 2007; Zhao et al., 2007; Alfred Chu and Chapleau, 2008; Barry et al., 2009; Chu and Chapleau, 2010; Nassir et al., 2011; Wang et al., 2011; Munizaga and Palma, 2012; Gordon et al., 2013), the boarding of passengers is usually known, and the alighting is unknown as passengers only tap their card when boarding a transit vehicle. Passengers' alighting location can then be inferred based on the next boarding location using a trip chaining algorithm (Barry et al., 2007; Trépanier et al., 2007; Zhao et al., 2007; Alfred Chu and Chapleau, 2008; Farzin, 2008; Barry et al., 2009; Chu and Chapleau, 2010; Nassir et al., 2011; Wang et al., 2011; Munizaga and Palma, 2012; Ma et al., 2012; Gordon et al., 2013; He and Trépanier, 2015; Kumar et al., 2018).

Trip chaining algorithms developed so far use assumptions on various parameters, e.g. buffer radius to find the closest stop to the boarding location, walking distance threshold after alighting to board the next route, time threshold to distinguish between boarding and transfer, etc. These parameters can vary among different transit systems and can affect the trip chaining results and therefore the origin-destination matrix. The current research tries to relax the assumptions related to these parameters by proposing a robust trip chaining algorithm.

The algorithm is applied to the AFC data from Twin Cities, Minnesota which has an open transit system (Nassir et al., 2011), where transit passengers use (tap) their card only once. The system is more complex than other systems described in previous research because sometimes passengers tap their card while entering the bus (when they board a "regular route" or "non pay-exit" bus) or sometimes while exiting the bus (when they alight a "pay-exit" bus). The pay exit buses are generally outbound trips from central areas such as Downtown Minneapolis or the University of Minnesota campus to sub-urban areas. The existing trip chaining algorithm changes significantly when the combination of such tags are observed for a card number. The proposed method creates a set of possible trips for a given card tag, calculates the probability that the passenger has used each trip, and then infers the boarding and alighting on the basis of the most likely trip.

The rest of the paper is organized as follows: Section 2 presents a summary of related work done in this research area, followed by motivation behind this research in Section 3. Then, the proposed trip chaining algorithm is described in Section 4, which is followed by the analysis of the results in Section 5. Finally, conclusions and recommendations for future research are provided in Section 6.

2. Related work

As most of the fare collection systems record passengers' boarding information only, alighting information must be inferred using the sequence of taps (or tags) made by the passenger throughout the day. Thus, a significant amount of research has been done to develop algorithms to determine the alighting location (Li et al., 2018). Navick and Furth (2002) used location-stamped fare box data of Los Angeles area bus routes to determine alighting location using an assumption that boarding pattern of current trip and alighting pattern of opposite trip are symmetric for the entire day which means passengers board the bus again from the same stop where they alighted during the previous trip. Building on that assumption, Zhao et al. (2007), Barry et al. (2007), Barry et al. (2009), Gordon et al. (2013) developed a method of trip chaining for origin and destination inference with the following assumptions:

1. passengers return to the same location to board the bus where they alighted during the previous trip,