Generating network-wide travel diaries using smartcard data

Oded Cats · Alex Vermeulen · Matej Cebecauer · Erik Jenelius · Yusak Susilo

Problem description

There is a rapid increase in the deployment, acquisition and analysis of automated fare collection (AFC) systems, enabling a profound change in the ability to analyze high-volume data that relate to observed passenger travel behavior and recurrent patterns. The analysis of such passively collected data offers direct access to a continuous flow of observed passenger behavior at a large scale, saving expensive data collection efforts. For a review of the spectrum of applications – from strategic demand estimation to operational service performance measurements - see Pelletier et al. (2010).

The partial information contained in smartcard data has led to a vibrant development of destination inference (Munizaga and Palma 2012), transfer inference (Nassir et al. 2015) and train inference (Zhu et al. 2017). These studies were often performed for a given fare scheme where passengers either tap in at gates or upon boarding a vehicle. Large metropolitan public transport systems often consist of different modes, operators and fare schemes giving rise to a hybrid fare collection regime where a combination of inferences is required. The objective of this study is to combine destination, transfer and train inference algorithms in order to obtain a network-wide OD matrix estimation in the context of a mixed fare scheme system such as in the case of Stockholm County (described below). Our approach is to generate a travel diary database for the entire population of smartcard holders based on the inferred alighting locations and times as described in the subsequent section.

The case of Stockholm County

The public transport system in Stockholm extends across the greater Stockholm area, covering ca 6,500 km^2 and 2.3 million inhabitants. The system includes 21 commuter train, metro, light rail and tram lines spanning ca 470 km, around 490 bus lines spanning ca 9,100 km, and a number of ferry lines (SLL 2016). The main ticketing system is the Access system, which uses electronic tickets that are loaded onto contactless cards. The system was introduced in limited scale in 2008 and the average number of ticket validations per day has since grown to 1.9 million in 2018. Each card has a unique number, which allows it to be traced.

Tickets are validated upon access to stations or boarding of vehicles but not on egress or alighting. In other words, the Access system is “tap-in only”. In the commuter train and metro networks, tickets are validated at the station entrances. In the light rail and tram network, there has been a transition from validations made by on-board staff to validations at the ungated platforms. In both networks, the travel direction and the line (in cases where multiple lines serve the station) are not stored in the Access data. In the bus network, validations are generally done upon boarding the vehicle, and the data contain information about the line, bus stop, and specific vehicle. Moreover, depending on the transfer locations, passengers may or may not have to tap in and tap out when switching between lines and modes.

Method

We propose a method to estimate the alighting station in a multimodal public transport system, where tap-in transactions are observed in a complex network. Similar to previous literature it is assumed that the alighting occurs within a certain distance of the next transaction. Trip elements are assessed individually resulting in individual travel diaries for users that can be aggregated at a later stage. This involves three sequential inference algorithms:

(1) *Alighting inference* is based on data gathered from the current transaction and the assumption of a maximum walking distance to the next transaction. This is summarised in Figure 1. The result of this step is an individual travel diary where the boarding and alighting locations are known.
(2) *Vehicle inference* is done using AVL data in case of station or gate validations by matching the passenger trip with a vehicle trips given time window constraints and the inferred alighting location. The process infers both the vehicle/trip as well as the alighting time.
(3) *Journey inference* is performed by comparing the time gap between the inferred alighting time and the next tap-in transaction time with a given threshold for transfers. If the time is less than the threshold we assume that a transfer was made, otherwise we assume that an activity was carried out.

This process results in the personal travel diary for each single card holder, which are used to build the journey Origin-Destination (OD) database.

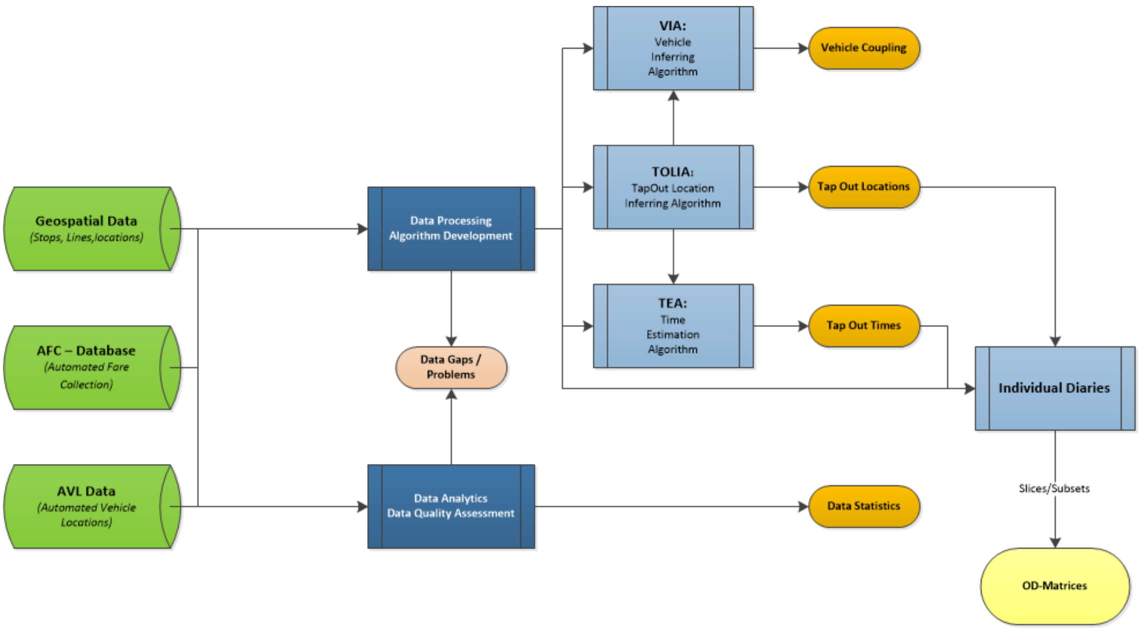

Figure 1: Schematic analysis process. Input data is represented by the three databases on the left, Processes are denoted in the middle, outputs of algorithms or processes are denoted in orange bubbles and the final output is the OD matrices formulated from Individual Travel Diaries.

Preliminary results

The raw smartcard data consist of approximately 680 million tap-in records in 2017. The majority of tap-ins are recorded at metro gates (45%) and upon boarding buses (41%) while the remaining consists of commuter trains, trams and ferries. The framework described above enables the generation of travel diaries from raw smartcard data. The current implementation of the inference algorithms allows performing the inference for 80% of the metro trips and 70% of the bus trips.

Figure 2 shows a snapshot of observed tap-ins and inferred tap-outs occurring within a given time slice for the central part of the case study area. During the morning peak period, passenger volumes directed towards the inner city are significantly larger than outbound trips.

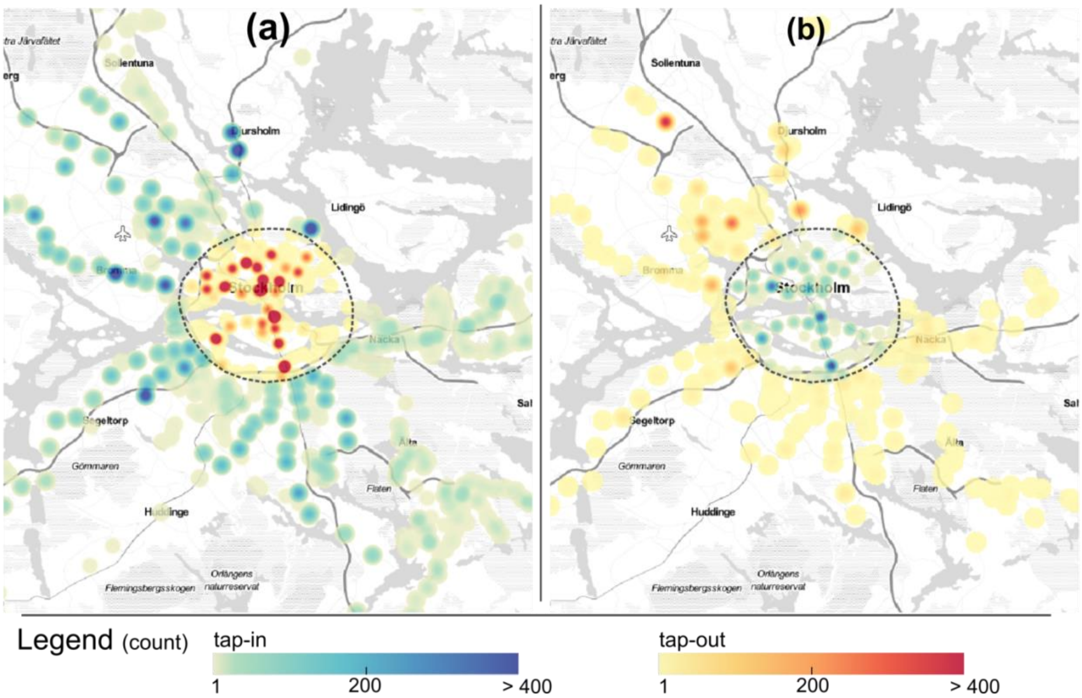

Figure 2: Density heatmap visualization of all smart card data trips (a) into the city and (b) out of the city on 2017-10-09 at 08:00-08:15.

Outlook

The implemented inference algorithms and the derived travel diaries facilitate the construction of OD matrices that are essential input for services planning. Furthermore, the estimated OD matrices can be used for measuring the impacts of various interventions such as fare policy and service design changes. The inferred travel diaries also allow for extracting passenger loads for each vehicle trip segment across the network at the same resolution as the flow outputs of schedule-based transit assignment models. These can be used for service performance and monitoring purposes. Moreover, a partial calibration and validation can be performed by contrasting the obtained flows with observed passenger counts available for a large sample of bus trips.

References

Munizaga M.A. and Palma C. (2012). Estimation of a disaggregate multimodal public transport origin-destination matric from passive smartcard data from Santiago, Chile. *Transportation Research Part C*, 24, 9-18.

Nassir N., Hickman M. and Ma Z-L. (2015). Activity detection and transfer identification for public transit fare card data. *Transportation*, 42(4), 683-705.

Pelletier M.P., Trepanier M. and Morency C. (2011). Smart card data use in public transit: A literature review. *Transportation Research Part C*, 19 (4), 557-568.

SLL (2016) ”Fakta om SL och länet 2016.” Trafikförvaltningen Stockholms läns landsting, Report S-2016-0762.

Yap M., Cats O. and van Arem B. (2018). Crowding valuation in urban tram and bus transportation based on smart card data. *Transportmerica A*. In press.

Zhu Y., Koutsopoulos H.N. and Wilson N.H.M. (2017). A probabilistic passenger-to-train assignment model based on automated data. *Transportation Research Part B*, 104, 522-542.