Reconstructing bus vehicle trajectories from transit smart-card data

Pieter J. Fourie, *Future Cities Laboratory, Singapore ETH Centre for Global Environmental Sustainability (SEC), Singapore, fourie@ivt.baug.ethz.ch, +65 9116 8440*

Abstract

Transit smart card automated fare collection systems are becoming increasingly prevalent in cities across the world. Trajectory reconstruction from smart card data includes constructing passenger trajectories or flows (Itoh et al., 2013; Yang et al., 2013; Yuan et al., 2013) or train trajectory reconstruction through regression analysis (Sun et al., 2012). It appears that no work has been done so far to reconstruct the detail trajectories of bus vehicles from these data. In this work, bus vehicle trajectories have been (approximately) reconstructed from a full day of smart card transactions in Singapore, from a mid-April Wednesday, 2013. The trajectories were recorded in a format compliant with the output from the agent-based simulation system, MATSim (www.matsim.org), using the specifications developed by (Rieser, 2010). This makes visualisation and analysis possible in compliant software, such as Senozon Via (http://senozon.com/products/via). A significant application of the resulting trajectory data is for the construction of realistic meta-models of travel time between transit stops during the course of the day in a fast pseudo-simulation (Fourie et al., 2013), as well as to route agents using realistic travel times in a dynamic transit router (Ordóñez Medina and Erath, 2013).

The Singapore Contactless ePurse Application System (CEPAS)

The CEPAS system operates with a contactless smart card recording a user's pre-loaded cash balance. When a user taps the card on a system sensor, a transit transaction is recorded and relevant information calculated and the user's balance adjusted. The Singapore public transportation system uses distance-based pricing, and operates differently depending on transportation mode (rail or bus). If they need to use mass rapid transit (MRT) train service, users tap in and out at station entrances. The exact service and vehicle they use, as well as their boarding/alighting time are therefore not directly recorded, but can be derived, e.g. (Sun et al., 2012).

In the case of bus services, users have to tap in and out when they enter and leave the bus. For each transaction, the system records an array of information, including the vehicle's registration number, the user's card identifier, the service and direction of the bus, the boarding or alighting time, as well as the boarding or alighting stop locations, inferred from the vehicle's GPS tracker.

If we are to reconstruct the trajectories of buses from these transactions, we need to know when the first users enter or leave a bus at each stop the bus visits in its route, assuming that the vehicle had arrived or departed shortly before or after these events. We also require an electronic transit schedule for each route, identifying the stops in the route and the sequence of links traversed between stops. For the case of Singapore, this information has been gathered and converted into the MATSim format in a previous exercise (Erath et al., 2012).

Method

Vehicles are processed by transit line, direction and vehicle identification. Vehicles not appearing in the electronic transit schedule are ignored. All transactions for a line, direction and vehicle are drawn from the database. The relevant transit route from the schedule is identified based on the route scoring the highest number of transactions across its set of stops. Some GPS errors are eliminated by removing transactions that have stop identifiers not occurring in this assigned route. Then, the 'speed' between consecutive transactions is calculated: the network distance between the transaction stop identifiers is divided by the transactions' time difference. Transactions with a high 'speed' value (arbitrarily taken as > 80 km/h) are removed from the set (see Figure 1).

Figure 1 'Speed' classification of transactions to eliminate GPS errors. The plots show the total ridership for a bus, with dots on lines denoting transactions over time. In the top plot, transactions are coloured by stop ID. The following two plots show transaction ‘speeds‘ before and after errant transaction removal. Acceptable speeds (< 80km/h) are shown in blue to grey, higher speeds tend towards pink. Note the correlation between pink dots and sudden changes in stop colour in the first plot.

Transactions are then processed by stop ID, and closely timed transactions are clustered to produce a vehicle dwell event (arrival at and departure from a bus stop). For each dwell event, the vehicle arrival and departure time is then taken from the largest sub-cluster of closely timed transactions, in order to eliminate false dwell event timing from early tap-outs and late tap-ins (Figure 2).

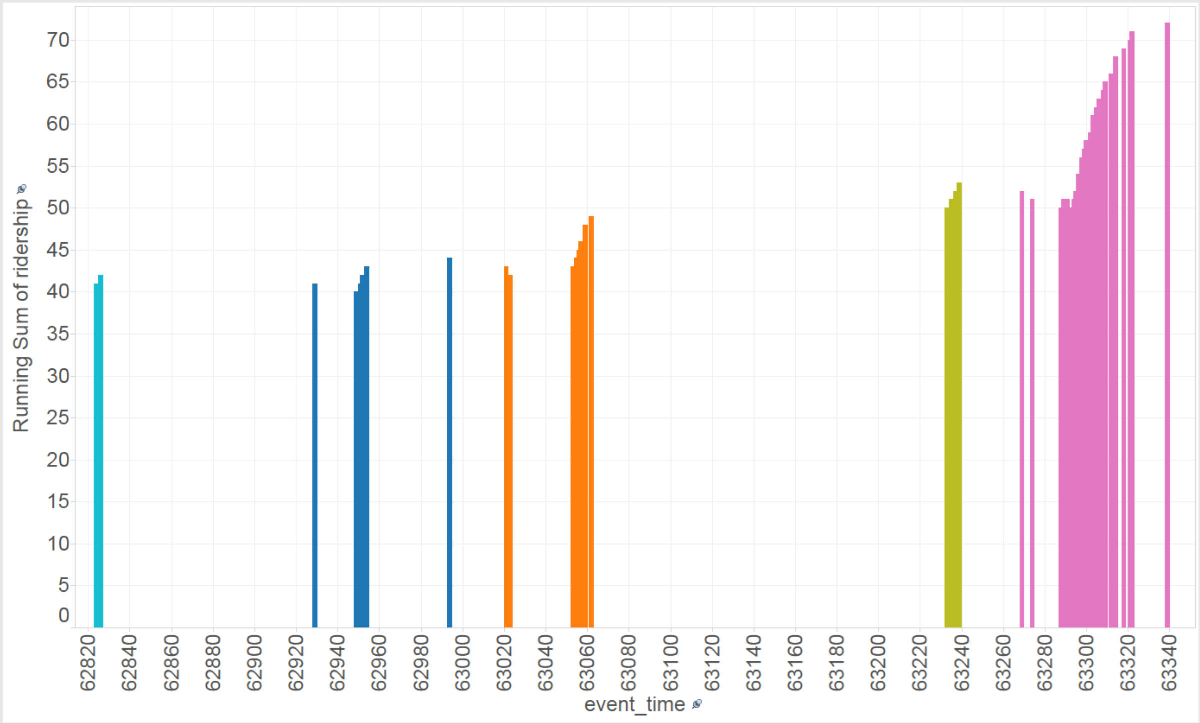

Figure 2. Bus ridership from transactions by time of day (seconds). Each boarding transaction increases ridership by one, each alighting decreases it by one. Transactions are colored by stop ID. For the dark blue transactions, note the early tap-out left and late tap-in right of the central cluster.

The timing of dwell events is adjusted to a minimum of 10 seconds. The sequence of dwell events is then grouped into separate route trips, by evaluating their order against the schedule. Missing dwell events (stops driven past with no transactions) are created in order to satisfy the requirements for the MATSim transit specification. These dwell events have a zero duration, and they are spaced at intervals proportional to the network distance between their stops (Figure 3).

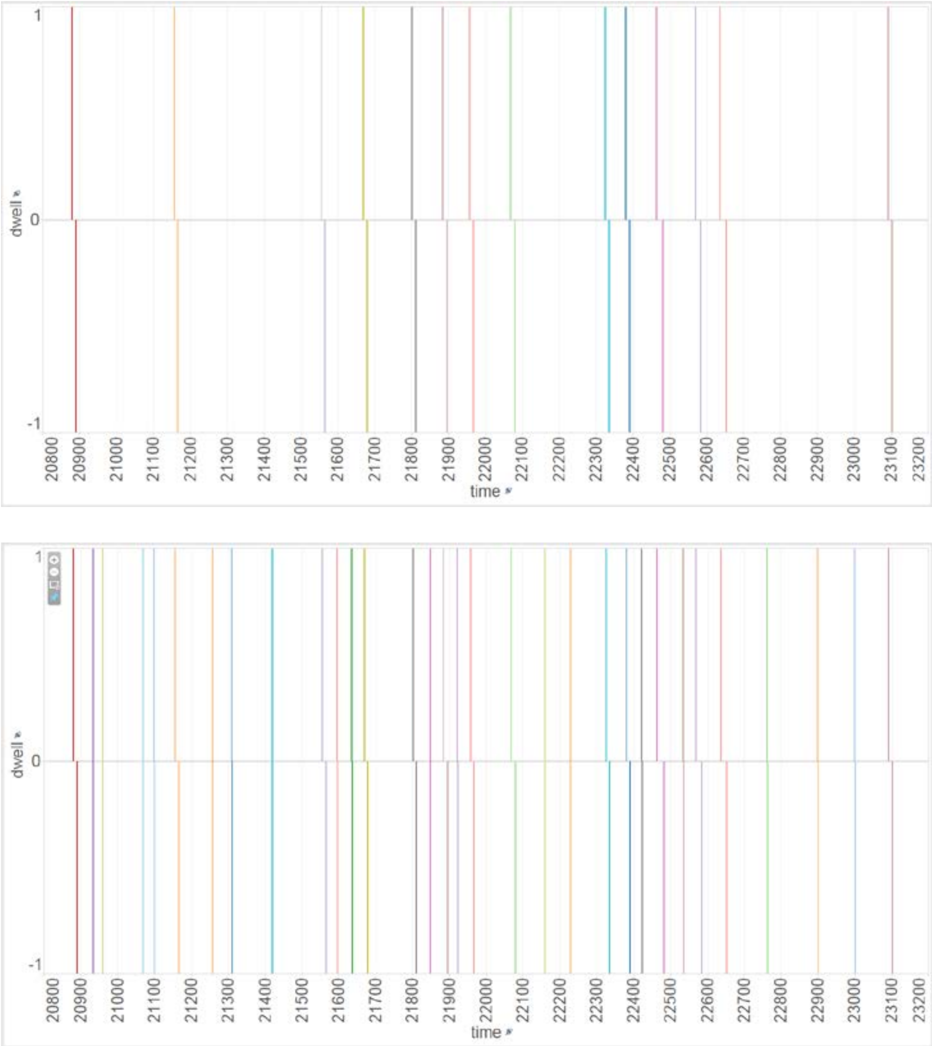

Figure 3. Bus dwell events vs. time before (top) and after missing dwell event insertion (bottom). Dwell events are coloured by stop ID, arrivals are denoted by a line above zero, departures by a line below.

Transaction times for early tap-outs and late tap-ins are adjusted to fall within their corresponding dwell event. Then, the vehicle trajectory is written out in a separate file for each vehicle in the MATSim XML format. The file records each dwell event and its associated transactions, and the sequence of links the vehicle traverses between consecutive dwell events, taken from the transit schedule. Given the time between consecutive dwell events, the time spent on each link is taken as (the total time between stops) multiplied by (the free-speed travel time of the link, not exceeding 80 km/h) divided by (the total free-speed travel time of all links between stops). Finally, the separate vehicle files are combined chronologically into a single output file using a merge-sort operation. The process currently discards approximately 3% of transactions due to the various exceptions discussed above. Improvement strategies to improve this rate will be discussed in the full version of this paper.

References

Erath, A., Fourie, P.J., Eggermond, M., Chakirov, A., Ordóñez Medina, S.A., Axhausen, K.W., 2012. Large-scale, agent-based transport demand model for Singapore. Presented at the 13th International Conference on Travel Behaviour Research (IATBR), Toronto.

Fourie, P., Illenberger, J., Nagel, K., 2013. Increased Convergence Rates in Multiagent Transport Simulations with Pseudosimulation. Transp. Res. Rec. J. Transp. Res. Board 2343, 68–76.

Itoh, M., Yokoyama, D., Toyoda, M., Tomita, Y., Kawamura, S., Kitsuregawa, M., 2013. Visualization of Passenger Flows on Metro. Technical report. University of Tokyo.

Ordóñez Medina, S.A., Erath, A., 2013. New dynamic events-based public transport router for agent-based simulations. Working paper. Future Cities Laboratory, Singapore.

Rieser, M., 2010. Adding transit to an agent-based transportation simulation. Ph. D. Thesis, Technical University Berlin, Berlin.

Sun, L., Lee, D.-H., Erath, A., Huang, X., 2012. Using Smart Card Data to Extract Passenger's Spatio-temporal Density and Train's Trajectory of MRT System, in: Proceedings of the ACM SIGKDD International Workshop on Urban Computing, UrbComp '12. ACM, New York, NY, USA, pp. 142–148.

Yang, Y., Du, B., Jiang, X., 2013. A Human Trajectory Estimate Based on Individual Mobility Pattern Library, Presented at the Green Computing and Communications (GreenCom), 2013 IEEE and Internet of Things (iThings/CPSCom)

Yuan, N.J., Wang, Y., Zhang, F., Xie, X., Sun, G., 2013. Reconstructing Individual Mobility from Smart Card Transactions: A Space Alignment Approach. Presented at the 2013 IEEE 13th International Conference on Data Mining (ICDM), pp. 877–886.