Using Smart Card Data to Identify Individual Passenger Behaviours during Disruption on Metro Rail Networks

George Goldberg , John W. Polak
Centre for Transport Studies, Imperial College London
Corresponding Author: george.goldberg08@imperial.ac.uk

Automatic Fare Collection (AFC) using Smart Cards is now widespread on public transport networks throughout the world. As well as providing convenience to passengers, these systems also provide a wealth of data which can be used by public transport operators to better understand how their networks are used.

A large amount of research into public transport passenger behaviour has been carried out using AFC data. However, the focus of much of this work has been on modelling flows and origin-destination matrices under normal conditions. More recently there has been an increase in the number of studies considering passenger behaviour during disrupted conditions. These studies typically fall into two broad categories. The first of these uses automatically collected data such as AFC records to analyse aggregate passenger behaviour and derive models that characterise the origin-destination demand. An example of this approach can be found in Silva et al. (2015), where the authors model how the entry and exit counts of stations on the London Underground are affected by unexpected disruptions to the operation of the network. The second approach adopts a disaggregate approach, typically using smaller samples of revealed or stated preference data to model the choices of individual passengers. An example of this approach can be seen in Pnevmatikou et al. (2015) where a disaggregate model of the route and modal choices made by passengers affected by a temporary closure of part of the Athens metro network is presented.

This study examines how passengers on the London Underground behave at a disaggregate level when faced with a disruption-causing incident on their journey. The analysis is based on AFC data collected over a period of 8 weeks from the Oyster smart card system used by nearly all passengers on the London Underground, together with a corresponding set of data detailing all incidents which took place on the network during the same time period. The approach taken lies somewhere between the two categories outlined above, working with a raw AFC data set giving nearly complete coverage of the travel histories of the system’s passengers, but attempting to identify individual behaviours rather than changes in aggregate entry and exit counts.
The method takes the form of a framework, within which the first step is to build a profile of individual smart cards from their travel histories and supplementary socio-economic information which can be attributed to them via inferences from their travel histories. Subsequently, the behaviours exhibited by each card on encountering disruption during the time period of the card histories are heuristically extracted. The list of behaviours considered are where the passengers:

1. Carry on with their journey as planned, and wait out any delays.
2. Abandon their journey altogether.
3. Change their origin and/or destination stations.
4. Postpone their journey until the incident has passed.
5. Change their route within the underground network to avoid the disruption.
6. Change to an alternative mode of transport to complete their journey.

This historical set of smart card profiles and their responses to incidents is then used to calibrate a model which in turn is validated against a distinct historical data set of smart card histories and corresponding incidents.

Learning from Guo (2013)’s experience building a similar framework for disrupted road traffic, the modelling step is defined flexibly to allow a range of different models, both discrete choice and machine learning, to be tested and their respective strengths and weaknesses evaluated. Of particular interest in this step is an evaluation of the trade-offs enforced on the different approaches when the requirement of real-time operation is applied to the model.

There are two key opportunities for applying the results of this work. The first is in real-time management of incident situations by network operators, allowing them to improve their ability to predict how the demand will adapt to the situation at hand, and thus where they need to make interventions to maintain the smooth operation of the system. The second is when considering the resilience of timetable or service pattern changes, where an improved understanding of how the system is used not only during normal operations but also during disruption will assist in avoiding changes that cause significant problems when the system is disrupted.

References


Silva, R., Kang, S.M. & Airoldi, E.M., 2015. Predicting traffic volumes and estimating the effects
of shocks in massive transportation systems. Proceedings of the National Academy of Sciences, p.201412908.