



## **USING SMART CARD DATA FOR BETTER DISRUPTION MANAGEMENT IN PUBLIC TRANSPORT**

### **Predicting travel behavior of passengers**

**Evelien van der Hurk MSc, Prof. dr. Leo Kroon, Dr. Ting Li, Dr. Gábor Maróti,  
Prof. mr. dr. Peter Vervest**

Rotterdam School of Management, Department of Decision and Information Sciences,  
Erasmus University Rotterdam, the Netherlands

### **ABSTRACT**

Public transport operators are traditionally focused on resources in case of a disruption. Since the change to privatization of public transport, service is more emphasized, causing a shift in focus from resources to passengers in disruption management. For passenger oriented disruption management, information is needed on passenger movements and behavior in case of a disruption. The introduction of ticketing through smart cards has led to a new source of data on time, origin and destination of journeys. We study how this information can provide insight in passenger behavior. We present ongoing research on the value of smart cards for analysis and prediction of passenger travel patterns, and evaluation of system performance. We use Dutch smart cards data for a case study.

### **KEYWORDS**

Disruption management, operations research, smart card data, econometric models

### **INTRODUCTION**

Public transport operators face disruptions regularly. For instance, NS faces several disruptions daily. What to do as an airline or railway transport operator in case of a disruption is and has been an important field of research within Operations Research; see e.g. Jespersen-Groth et al. (2009). Most of the research in this area has focused primarily on operations, timetable issues (dispatching), resource management, capacity management and crew scheduling. One area of disruption management focuses on the rolling stock and its characteristics, such as in Goverde (2010). In this research our perspective is that of the passenger and the operated lines.

The public transport operator is currently viewed as a service provider to its passengers. It is this emphasis on the service that has caused a shift in focus of transport operators: they want to go from the traditional focus on trains, to a focus on passengers in disruption management. Information on passenger travel behavior is therefore essential. With the introduction of electronic ticketing through smart cards a lot of data has become available on passenger travel behavior.

Smart card data contains information on individual passenger behavior and therefore provides an opportunity to analyze behavior of a single passenger over time. This information is far more detailed than former data resources such as survey data and ticket sales. As a result, a more detailed estimation of passenger flows through the system will be possible. It is still an estimation, as though departure time and location are stored in the data, the actual route through time and place is not. Also, data is not in real time available. Hence in case of a disruption we need to make a prediction of the flows through the system. Knowing when and where passengers are at a certain moment *as well as* where they are going is essential to make disruption management passenger oriented – and is a valuable input for e.g. capacity planning in disruption management or information management. Furthermore, smart card data can be used as a system evaluation tool, analyzing past delay of passengers due to a disruption. This analysis in hindsight allows an operator also to estimate how a different information or capacity management may affect the delays experienced by passengers.

Smart card data is a fairly new source of data and currently mostly considered as a good substitute for traditional survey data. It is however much more than a possible substitute. It provides objective information of actual passenger movements. In that manner, it is a very valuable source of information on passenger flows, which are of special interest in passenger oriented disruption management. It is this value of smart card data that we want to demonstrate in this paper, from the perspective of disruption management.

We present ongoing research. Our aim is to be able to predict the flows through the public transport network at any time. This estimation should consist of the flows over the arcs of the system, as well as the destination of the passengers in the flows. In case of a disruption, we could estimate the number of passengers effected. Knowing the destination of the passengers, we may also predict the effects for the train routes and stations. Adding knowledge on passenger travel behavior in case of a disruption, e.g. wait, reroute, or travel differently, we can further improve this prediction. So, for passenger oriented disruption management, we want to predict the flows for all combinations of origin and destination over time.

Smart card data provides data on origin and destination of any journey made by a smart card, and becomes available to the transport operator generally one day after the date of the journey. Hence smart card data is a source of historic data. It however does not include information over the route choice of the passenger. This is a first challenge. The second challenge is to deduce travel behavior in case of a disruption. Both together are needed for the prediction of passenger flows for disruption management. We will address both challenges in the case study.

Furthermore, smart card data can be used for measurement of system performance. It provides valuable business information for a public transport operator concerning its service level: the experienced delay of passengers. For planners it can become an even more valuable tool, as it provides information on the effect of specific chosen disruption management strategy in hind sight. Knowledge about the passenger flows during that disruption can be used to predict the

effect of other disruption management strategies. Also, in hindsight with complete information about the disruptions and the flows, we could estimate an upper bound on the reduction in delay through disruption management. Hence, the data provides also a basis for evaluation of disruption management.

In this paper we will address all three topics at the hand of the situation of the Dutch Railways. We will use Dutch smart card data for an illustrative case study. We present preliminary results and ongoing research.

## **LITERATURE REVIEW**

Smart cards have been around for a while, with the first introduction of the Octopus card in Hong Kong in 1997 ([www.octopus.com.hk](http://www.octopus.com.hk)). Previous studies on smart card usage can be categorized into two groups. One group is focused on how smart card data can be used to replace the traditionally used survey data. The second group focuses on behavioral inference models, usually using econometric models.

### **Smart card data as a replacement for survey data**

An interesting illustration of this topic in research is the article by Park et al. (2008). They analyzed smart card data from the Seoul subway system to see whether these can be used instead of survey data. The results indicate that smart card data is an improvement compared to survey data, since it can provide information on the lines separately. Smart card data can be used to analyze statistics such as the number of transfers, boarding time, hourly trip distribution of the number of trips for different transit modes, and travel time distribution for all transit modes and user types.

Bagchia and White (2005) use smart card data of a UK bus to study card statistics, such as the number of trips and the number of transfers as well as the usage of the card. With that they estimate also the churn: turnover of the customer population of the bus company.

### **Smart card data as a source for behavioral inference**

Morency et al. (2007) use smart card data to deduce passenger behavior from smart card data for a Canadian bus line in Quebec. The regularity of passenger travel is studied through clustering and data analysis. The results indicate that for most people, the travel pattern is static and one can divide people based on travel pattern in several distinguished groups.

The recent research of Gonzalez et al. (2008) is worth mentioning here as well. Although it does not use smart card data but cell phone data, it is similar in that it studies mobility patterns of people. The research demonstrates that movement of people is not random, as is the traditional assumption. This is shown through data analysis of cell phone activity data. They showed that usually a person moves between a limited set of locations, and that there is a strong correlation between the time and the location. Therefore they conclude that the location of people is predictable.

## **CASE STUDY**

In this section, we present how smart card data can be used to predict passenger flows and analyze the delays of passengers due to a disruption at the hand of an case study using Dutch “OV-chipkaart” (smart card) data. We present some preliminary results on the predictability

of passenger travel behavior. In addition, we outline our ongoing research on deducing passenger delays.

## **Assessing passenger flows**

The smart card data contains a start and end location, including a time and date, though the check in (*ci*) and check out (*co*) of the card. A technical limitation is that no information on vehicle or route is stored in the case of Dutch railways. Now by inferring passenger behavior from this data, we mean linking a single pair of a start and end location to one specific path through place and time, e.g. a specific combination of trains used to travel from end to start. If knowing this for all passengers, we can deduce capacity usage from aggregating all time-space paths. Secondly, we would like to be able to infer from past capacity usage the future capacity usage, as estimations of current and expected flows in your network are valuable information in case of a major disruption. It enables anticipating the effects of a disruption for the passengers in the network, answering questions like the need for capacity for trains on disrupted lines, which rerouting options will be popular, and whether current capacity on these routes is sufficient. Therefore we start out by inferring this passenger behavior from past data.

The start location with date and time is the check-in of the card (*ci*), the end location with place and time is the check out (*co*). Currently around 700 000 trips a week are made at the Dutch Railways using a smart card. We can link a *ci* to a *co*. Based on time and place, we find all possible routes through the network for the trip. In this manner, only 40% of data can be non-trivially matched to a single route and vehicle. However, using an algorithm based on a set of logic rules we expect to be able to match about 80% of all trips. Aggregating results for all trips we can deduce the loading of the network. Furthermore, once linked we can see the arrival pattern and departure pattern of passengers to and from the station. Furthermore, by comparing several days or using prediction models we can study the predictive power of historic data for future flows.

## **Analyzing system performance**

We want to deduce the delay of passengers due to a disruption. Therefore, we will compare a disrupted situation to a non disrupted situation in historic data. We again, want to link a *ci,co* pair to a specific route through time and place. Succeeding in that, we could deduce whether people reroute or wait for the disruption to end. Also, shortage in capacity might be deductible. We want to see whether we can validate a set of general assumptions internally used at NS on passenger behavior in a disruption, or formulate new ones. These rules on passenger behavior in case of a disruption, e.g. rerouting, waiting or choosing different transportation, are essential for estimating needed capacity after the disruption. Also, we would like to show the effects on the network of a single disruption from a passengers perspective.

A similar study has been conducted in the thesis of Frumin (2008), where it is shown that delay can be observed from the data. We want to extend these analyses and show the effects of the disruption on capacity usage levels.

Finally, we want to see whether with better information we could reroute passengers faster through the network. As noted in Landex and Nielsen (2008) the level of assumed available information is important. Although as a disruption occurs, not all information is directly available, we will assume we have perfect information on occurrence and length of a

disruption once the disruption has occurred. In this manner we find an upper bound on the possible time savings of passengers through information.

## CONCLUSIONS

Smart card data is a valuable source of data for inferring passenger behaviour and capacity usage. Knowing passenger travel behaviour enables predicting the demand for capacity over time and location in the transportation network. This input can be used to learn the demand for capacity in a disrupted situation, and enables disruption management to shift focus from optimizing operations to providing service to its customers: the passengers. Smart card data can be used to evaluate capacity usage in disrupted and non-disrupted cases. Also, the smart card data can be used to evaluate the delays of passengers in case of a disruption. As we can deduce the delay and origin-destination pair of a passenger, we can also investigate whether the delay can be reduced by providing passengers with better rerouting options. Future research will examine how public transport operators can increase the informedness of passengers to reduce their delays and achieve a more even capacity usage.

## REFERENCES

- Bagchia M., P.R. White (2005) The potential of public transport smart card data, in: *Transport Policy*, Vol. 12, pp. 464-474.
- Frumin, F, N. Wilson (2008) *Oysterbased performance metrics for the London overground*, Master Thesis, School of Civil and Environmental Engineering, MIT.
- Gonzales, M. C., C. A. Hidalgo, A.L. Barabasi (2008) Understanding individual human mobility patterns, in: *Nature*, Vol. 453, pp. 779-782.
- Goverde, R.M.P. (2010) A delay propagation algorithm for large-scale railway traffic networks, in: *Transportation Research Part C*, Vol 18, Issue 3, pp. 269-287.
- Jespersen-Groth J., D. Potthoff, J. Clausen, D. Huisman, L. Kroon, G. Maroti, M. N. Nielsen (2009) Disruption Management in Passenger Railway Transportation, in: *Lecture Notes in Computer Science*, Springer, Berlin/Heidelberg, pp. 399 - 421.
- Landex, A, O. A. Nielsen (2008) *Methods to estimate railway capacity and passenger delays*, PhD thesis, Department of Transport, Technical University of Denmark.
- Morency C., M. Trepanier, B. Agard (2007) Measuring transit use variability with smart card data, in: *Transport Policy*, Vol. 14, pp. 193-203.
- Octopus Travel card Hong Kong official website, *Hong Kong Octopus Card* [www.octopus.com.hk]
- Park J.Y., D. Kim, Y. Lim (2008) Use of smart card data to define public transit in Seoul, South Korea, in: *Transportation Research Record*, Vol. 2063, pp. 3-9.
- Schöbel A. (2004) *Integer programming for solving the delay management problem*, Technical Report 92, Technische Universität Kaiserslautern, Institute für Techno- und Wirtschaftsmathematik.