Transport Policy 61 (2018) 84–95

Contents lists available at ScienceDirect

Transport Policy

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

Improving predictions of public transport usage during disturbances based on smart card data

M.D. Yap [a,*], S. Nijënstein [b], N. van Oort [a,c]

[a] Department of Transport & Planning, Delft University of Technology, The Netherlands
[b] HTM Personenvervoer N.V., The Netherlands
[c] Goudappel Coffeng B.V., The Netherlands

ARTICLE INFO

Keywords:
Disturbance
Passenger
Prediction
Public transport
Smart card

ABSTRACT

The availability of smart card data from public transport travelling the last decades allows analyzing current and predicting future public transport usage. Public transport models are commonly applied to predict ridership due to structural network changes, using a calibrated parameter set. Predicting the impact of planned disturbances, like temporary track closures, on public transport ridership is however an unexplored area. In the Netherlands, this area becomes increasingly important, given the many track closures operators are confronted with the last and upcoming years. We investigated the passenger impact of four planned disturbances on the public transport network of The Hague, the Netherlands, by comparing predicted and realized public transport ridership using smart card data. A three-step search procedure is applied to find a parameter set resulting in higher prediction accuracy. We found that in-vehicle time in rail-replacing bus services is perceived ≈1.1 times more negatively compared to in-vehicle time perception in the initial tram line. Waiting time for temporary rail-replacement bus services is found to be perceived ≈1.3 times higher, compared to waiting time perception for regular tram and bus services. Besides, passengers do not seem to perceive the theoretical benefit of the usually higher frequency of rail-replacement bus services compared to the frequency of the replaced tram line. For the different case studies, the new parameter set results in 3% up to 13% higher prediction accuracy compared to the default parameter set. It supports public transport operators to better predict the required supply of rail-replacement services and to predict the impact on their revenues.

1. Introduction

The last decade, in several cities worldwide automated fare collection (AFC) systems are introduced for the public transport system by public transport operators and authorities. For these AFC systems, passengers need to use a smart card for public transport travelling. Open systems in which passengers only need to tap-in, as well as closed systems in which both a tap-in and tap-out are required, are applied in practice. Although the main purpose of the introduction of AFC systems was to enable an easier way of revenue collection, additionally large amounts of data are generated which can be used to get more insight in passengers' travel behavior. Over the last years, data from AFC systems is used for many purposes by scientists and practitioners on a strategic, tactical and operational level (Pelletier et al., 2011). Data from AFC systems is for example used for destination inference in case of open systems with tap-in only (e.g. Trépanier et al., 2007; Nunes et al., 2016), transfer inference (e.g. Hofmann and O'Mahony, 2005; Jang, 2010) and journey inference to estimate origin-destination (OD) matrices (e.g. Seaborn et al., 2009; Wang et al., 2011; Munizaga and Palma, 2012; Zhao et al., 2007; Gordon et al., 2013). Other studies focus on fusion of smart card data of different operators (e.g. Nijënstein and Bussink, 2015) or clustering public transport stops in order to identify and classify public transport activity centers based on smart card data (Cats et al., 2015).

Next to the aforementioned studies which use smart card data to describe, analyze, cluster and visualize current travel patterns, there are also studies focusing on public transport ridership prediction based on smart card inferred travel patterns. Idris et al. (2015) developed several mode choice models based on revealed preference, contrary to traditional mode choice models having the tendency to overestimate public transport ridership. Wei and Chen (2012) developed a forecasting approach for short-term ridership predictions in metros using a combination of empirical mode decomposition and neural networks, whereas

* Corresponding author. P.O. Box 5048, 2600 GA Delft, The Netherlands.
E-mail address: M.D.Yap@TUDelft.nl (M.D. Yap).

https://doi.org/10.1016/j.tranpol.2017.10.010
Received 16 November 2016; Received in revised form 20 October 2017; Accepted 22 October 2017

Li et al. (2017) predict metro ridership under special events using a multiscale radial basis function (MSRBF) network. Ding et al. (2016) predict metro ridership using gradient boosting decision trees, thereby incorporating temporal features and bus transfer activities. In Van Oort et al. (2015a) a smart card based prediction model is developed which allows the prediction of effects of changes in public transport supply, like increasing the frequency or rerouting public transport services. This model considers the total urban public transport network and uses an elasticity approach, where parameter values are obtained based on revealed preference studies. Also effects of crowding can be incorporated in this short-term ridership prediction model (e.g. Van Oort et al., 2015b). This type of prediction model is of added value to improve prediction accuracy of the impact of structural network changes, which are usually implemented by operators on one or on a few fixed dates in the year. However, in practice many public transport operators are confronted with temporary closures of infrastructure many more times per year. These temporary infrastructure closures are for example caused by maintenance work, track renewal or redesign of public space. These closures usually result in longer travel time, more transfers, lower rider ship, lower passenger satisfaction, and less revenues. In the Netherlands, a tendency can be observed of more, larger and more long-lasting rail infrastructure closures. For example, HTM, the urban public transport operator in Den Haag, the Netherlands, was confronted with more than 20 temporary track closures in 2015. It therefore becomes more urgent for operators to predict the impact of these (planned) disturbances on their passengers, ridership and revenues. This impact of temporary track closures on demand and supply is different compared to the impact of structural network changes. Passengers might be willing to postpone a single trip, change their mode choice or route choice, or accept the use of rail-replacement bus services for temporary situations. Operators on the other hand have to accept the temporary reduction in level of service – because of rail-replacement bus services, additional travel time and transfers – and might accept the temporary additional operational costs for these bus services and communication. It can be concluded that the responses of passengers and operators differ in case of temporary network changes, compared to structural network changes. In order to predict passenger impacts of temporary network changes with sufficient accuracy, other/additional parameters and/or different parameter values in the public transport ridership prediction models are therefore required.

This study aims to improve the prediction accuracy of the impact of planned, temporary disturbances on public transport usage. To this end, in this study a new parameter set is calibrated and validated to predict public transport ridership in case of planned disturbances. This parameter set is based on smart card data derived from AFC systems during several planned disturbances which occurred in The Hague in 2015. The study results in a new set of parameter values allowing to better predict passenger impacts of planned disturbances in urban public transportation. With this result, more insight is gained in passenger behavior during disturbances. It also supports operators to predict the impact on their revenues, and to better align supply of rail-replacement services on alternative routes to the remaining demand, in order to efficiently use their scarce resources. This paper is structured as follows. Chapter 2 describes the methodology to calibrate and validate the parameter set of the ridership prediction model. Chapter 3 describes the case study network to which the methodology is applied. Chapter 4 discusses the results of this study. At last, in chapter 5 conclusions and recommendations for further research are formulated.

2. Methodology

2.1. Origin-destination matrix estimation

When travelling in trams or busses in the Netherlands by smart card, passengers are required to tap-in and tap-out at devices which are located within the vehicle. This means that in the Netherlands the passenger fare is based on the exact distance travelled in a specific public transport vehicle. Especially for busses, this is different from many other cities in the world where often an open, entry-only system with flat fare structure is applied, for example in London (Gordon et al., 2013) and Santiago, Chile (Munizaga and Palma, 2012). This means that for each individual transaction the boarding time and location, and the alighting time and location of each trip leg are known. Also, it is known in which public transport line and vehicle each passenger boarded and alighted with its unique smart card number. This closed within-vehicle system therefore eases the destination and journey inference, compared to open entry-only systems. When merging this closed within-vehicle AFC system with Automated Vehicle Location (AVL) data, also vehicle occupancies can be inferred directly from the transaction data for each line segment and vehicle.

For an urban public transportation network with tram and bus lines, journeys can be inferred by combining registered trip legs made with the same smart card ID. In this study we used a simple temporal criterion to determine whether a passenger alighting is considered as final destination or as transfer. When the boarding time to a vehicle follows within a certain time window after the alighting time of the previous trip leg made with that same card, two AFC transactions are considered as one journey. This approach is also used, for example, by Hofmann and O'Mahony (2005) and Seaborn et al. (2009). We are aware that in scientific literature more advanced transfer inference algorithms have been developed (e.g. Zhao et al., 2007; Munizaga and Palma, 2012; Gordon et al., 2013; Yap et al., 2017). In Dutch practice however, operators apply only a time window threshold between the previous alighting and next boarding as transfer inference criterion. In order to compare the prediction accuracy of the new proposed parameter set with the earlier operator predictions with the default parameter set, we decided not to adjust the transfer inference algorithm in this study. In this way, we can evaluate purely the effects of our new parameter set on the prediction accuracy, while not also changing the transfer inference algorithm simultaneously. In the Netherlands, a maximum threshold transfer time of 35 min is applied to classify trip legs made by the same smart card ID as one journey. By aggregating all journeys, a stop-to-stop smart card based OD matrix can be inferred. In the ridership prediction model, zones are located at the stop locations. Only stop codes which belong to the same stop from a passenger perspective, are aggregated to one zone. This means that stop codes of platforms of the same stop in opposite directions, or stops located at the same intersection, are represented by one zone. This is done to prevent passenger travel patterns to be relying too strong on the exact current stop codes of boarding and alighting in the undisturbed scenario. Under assumption that the distribution of destinations j from each origin i for non-card users is similar to the distribution of smart card users, which is in line with the assumption applied by Munizaga and Palma (2012) to correct for missing tap-outs, the zone-based OD matrix can be scaled based on the small percentage of non-card users in the Netherlands. Determination of the share of non-card users is based on passenger counts.

When travelling by train or metro in the Netherlands, there is also a closed system where transactions are required during boarding and alighting. For train and metro, devices are however located at the station gates. This means that train-train or metro-metro transfers, as well as exact chosen routes cannot be determined directly from the data, and that trip and transfer inference algorithms are necessary to obtain these insights.

2.2. Public transport ridership prediction model

For the prediction of public transport usage in case of planned disturbances, in this study the public transport ridership prediction model as described in Van Oort et al. (2015a) is used as basis. For an urban public transportation network, let the set of public transport stops and lines be denoted by S and L respectively. Each line $l \in L$ is defined by an ordered