Journal of Rail Transport Planning & Management xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Journal of Rail Transport Planning & Management

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

Use of mobile phone data for analysis of number of train travellers

Anette Østbø Sørensen[a,*], Johannes Bjelland[b], Heidi Bull-Berg[c], Andreas Dypvik Landmark[c], Muhammad Mohsin Akhtar[a], Nils O.E. Olsson[a]

[a] Norwegian University of Science and Technology, Dept. of Mechanical and Industrial Engineering, N-7491, Trondheim, Norway
[b] Telenor Research, Oslo, Norway
[c] SINTEF AS, Postboks 4760 Sluppen, N-7465, Trondheim, Norway

ARTICLE INFO

Keywords:
Railway
Ridership
Train
Mobile phone data

ABSTRACT

Several studies have pointed to the difficulties of obtaining good data on train ridership. There are at least two challenges regarding these data. First, train operators consider such data confidential business information, especially in high resolution. Second, the data that actually are available vary in quality and coverage. This paper studies mobile phone data as an alternative measure to obtain data about train ridership.

Handset counts were obtained from one telecom operator for selected mobile phone base stations and compared with timetable data and APC. The selected base stations are located so that it is likely that a large share of the mobile phone traffic is generated by train passengers. The number of units connected to a base station is found to correspond relatively well with the trains that pass close to the base stations. A ratio between the handset count and APC data appear as promising in utilizing handset count to calculate train ridership, with ratios around one in the rush hours. We discuss preliminary results as well as methodological and technical challenges.

To make sure that we do not violate privacy concerns, the data used in the study have been approved by personal privacy representatives.

1. Introduction

1.1. Research on train ridership is important

When analysing public transportation, including trains, ridership is an important factor. The number of travellers is a measure of demand for transportation services, which is important information for planning and evaluations. With updated ridership information, planners should be able to get a detailed, continuous and accurate vision of the travel behaviour of their customers. This is important in planning and improving the transportation service. Other uses of ridership numbers are calibration and validation of transport models. Boyle (1998) identifies four main reasons why ridership data are collected. Firstly, ridership is reported to external funding and oversight agencies. Secondly, it monitors trends over time. Thirdly, ridership is a key performance indicator at various levels of the transportation system. Finally, ridership data identifies locations with the greatest boarding and alighting activity, which is important not only for its own purpose, but because the safe management of the railway may depend upon it. Other issues that call for data on ridership on trains include fare equipment location optimization, fare policy change and train schedule (Li, 2000). In addition, revenue distribution in integrated public transportation systems can be based on ridership data.

According to Vuchic (2005), the purpose of obtaining data on passenger volume and load count is to monitor trends and travel behaviour over time. Such data show passenger volumes on different sections of a line, the maximum number of passengers on

* Corresponding author.
E-mail address: anette.o.sorensen@ntnu.no (A.Ø. Sørensen).

https://doi.org/10.1016/j.jrtpm.2018.06.002
Received 16 March 2018; Received in revised form 27 May 2018; Accepted 6 June 2018
2210-9706/

Please cite this article as: Sørensen, A.Ø., Journal of Rail Transport Planning & Management (2018), https://doi.org/10.1016/j.jrtpm.2018.06.002

different lines and when the maximum is reached, along with information on variations in passenger volume. Ridership data are also a key performance indicator in public transport (Vuchic, 2005).

On the other hand, the process of obtaining data on ridership creates least two major challenges. Firstly, train operators consider such data confidential business information, especially in high resolution (Vigren, 2017). Secondly, the data that actually are available vary in quality and coverage. Several studies highlight the unreliability of ridership data (including Chu and Chapleau, 2008 and Fowkes et al., 1985).

1.2. Mobile phone data and alternative technologies

Doi and Allen (1986) studied a rapid transit line for a period of about six and a half years (from 1978 to 1984) based on ridership data provided by a transit authority. Even though some studies combine several data sets, most studies on ridership rely on one data source. Wang et al. (2011) explored the application of archived data from Automated Data Collection Systems (ADCS) to transport planning with a focus on bus passengers' travel behaviour. They claimed that it was the first known attempt to validate the results by comparing automated ridership data with manual passenger survey data. Passenger distribution in the urban Copenhagen rail network is, according to Nielsen et al. (2014), tracked based on a combination of Electronic Weighing Equipment (EWE) and Automatic Passenger Count (APC). The two systems provide complementary information, since the weight-based estimation provides information about the total traffic volume and automatic passenger counting provides information on passenger flow. The two systems can also be used to perform quality assurance of each other's measurements. Zhao et al. (2007) combine data from the automated fare collection system and the automated vehicle location system to examine the rail-to-bus trip sequence to obtain a clearer picture of ridership patterns. De Regt et al. (2017) combine smart card and Global System for Mobile Communications (GSM) data to examine spatial and temporal patterns of public transport usage versus overall travel demand. The methodology was applied to a case study in Netherlands, and was shown to be valuable in supporting tactical transit planning and decision making.

Sørensen et al. (2017) identify several technologies for measuring ridership on trains. The technologies and approaches include (1) manual counts and surveys, (2) on-board sensors, such as door passing, weight, CCTV and Wi-Fi-use, (3) ticketing systems, ticket sales or ticket validation, and (4) tracking of travellers for larger part of the journey, such as tracking of mobile phones and payments. Pelletier et al. (2011) presents an overview of the first developments of smart card. Smart cards are used to store individual data such as identification, biometrics, photos, banking data, transportation fares, etc. In transit, the main purpose of smart cards is to collect revenue, but they also produce detailed data on onboard transactions which can be useful to transit planners on both a strategic, tactical and operational level. Smart cards in public transit are usually issued by the operators to be used on their own system, and the cards are typically tapped over the reader when the user enters the vehicle.

Pelletier et al. (2011) summarise the pros and cons of smart card use in public transit which they revealed in their literature review. Some of the pros and cons are also valid for mobile phone data. Disadvantages are for instance that the data cannot provide information on trip purpose or on user assessment of service, and that development cost is high. Furthermore, the ultimate destination is not provided. Advantages include that the user role in data collection previously achieved by the survey process is minimized, as well as improved data quality and increased amount of statistics available.

Data from on-board sensors and ticketing systems are typically managed by the transportation providers. However, surveys, payments statistics and mobile phone data may be available to stakeholders outside the public transportation system, which can be an advantage because access to ridership data can be an issue for business reasons. Furthermore, mobile phone data appears to be an interesting option because they can track complete journeys.

Mobile phone data can be used to derive good estimates of dynamic quantities, such as travel times, train occupancy levels and origin-destination flows, for transportation studies (Aguiléra et al., 2014). The advantages of mobile phones as sources of data include:

- the potential to generate information about travels that combine different modes of travel (such as walking, bus, train),
- to track journeys that include transfer between trains,
- to estimate commuting patterns,
- and to derive estimates of travel times, train occupancy levels and origin-destination flows.

Several studies on mobile phone data in transportation research utilise Call Detail Records (CDR) data. This paper studies a different type of mobile phone network source that will be presented in Section 3.2.1. Three main types of mobile phone data are collected using passive collection: CDR data, Probes data and Wi-Fi data (Larijani et al., 2015). CDRs are generated by phone communication activities and contain relevant information about the activity (e.g., caller/callee, time, duration) and the location of the cell phone tower that handles the communication (Zhao et al., 2016). Studies have shown that CDR data can be used to study habits and mobility patterns of mobile users (Bianchi et al., 2016; Zhao et al., 2016), to study user movements (Leo et al., 2016), and to calculate commuting matrices with a very high level of accuracy (Frias-Martinez et al., 2012). Studies have also looked at utilizing mobile data to estimate intra-city travel time (Kujala et al., 2016) and have shown that mobile data could be employed as a real-time traffic monitoring tool (Järv et al., 2012).

Studies point out that CDR data are coarse in space and sparse in time (Becker et al., 2013) because people's phone communication activities are unevenly distributed in space and time. The bias of CDR data in human mobility research depends on what research question one wants to answer and how frequently, as well as when and where, one uses the mobile phone to contact others (Zhao et al., 2016). It has therefore been suggested that researchers should use CDR data with caution.

CDR data contain information about the caller/callee, so they are not anonymous. Consequently, studies that utilise CDR data are