Journal of Transport Geography xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Journal of Transport Geography

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

Estimating the residence zone of frequent public transport users to make travel pattern and time use analysis

Margarita Amaya[1], Ramón Cruzat[2], Marcela A. Munizaga*

Universidad de Chile, Casilla 228-3, Santiago, Chile

ARTICLE INFO

Keywords:
Smartcard
Public transport
Zone of residence
Travel time
Time use

ABSTRACT

Public transport systems with electronic fare collection devices continuously store data related to trips taken by users, which contain valuable information for planning and policy analysis. However, if the card is not personalized, there is no socioeconomic information available, which imposes a limitation on the types of analysis that can be performed. This work presents a simple method to estimate the residence zone of card users, which will allow socioeconomic variables to be estimated, thereby enriching the analytical possibilities. The method, which is based on the observation of morning transactions of frequent users, is applied to a sample of over 2 million cards. The method is evaluated using a sample from the Santiago ODS where users declared their card id and also declared their home address. A sample of 888,970 cards that are observed at least three days in a week and show spatial regularity for the morning transaction is used for zone of residence estimation and analysis of travel patterns and time use. The results show that users who live in the city center or in the wealthier East zone experience lower travel time, spend more time at home and less time at work.

1. Introduction

Cities have the purpose of satisfying the citizens' needs to interact and conduct different activities such as work, study or leisure. Transport systems are designed to allow the required movements within the city efficiently and effectively. However, these goals are not always achieved. Cities are complex systems, with a vast amount of elements, which are sometimes difficult to predict and to govern. Transport systems are also complex, with time-space relations that are evident and absolutely present for the user, but difficult to observe by an external modeler. To be able to understand the time use and travel patterns within a city, it is necessary to observe such behavior. The most common empirical studies of time use and mobility are conducted at a disaggregate (individual) level, with samples obtained from surveys. However, surveys are expensive and difficult to conduct, and there is a trade off between survey length and complexity, and sample size and accuracy (Jara-Díaz and Rosales, 2015). So the information available for conducting such studies is limited.

We now have information available from different technological devices such as mobile phones, which can be observed when they are connected to the antennas. The public transport systems are usually equipped with automatic vehicle location (AVL) devices installed in vehicles and automatic fare collection (AFC) systems. This information allows observing how people move through the city for large segments of the population, such as public transport users or the clients of a particular mobile phone company. However, usually there is no sociodemographic information associated to the movement traces observed, and sociodemographic characteristics are relevant to explain time use and travel behavior. The purpose of this study is to explore the possibility of using AVL and AFC data to make travel pattern and time use analysis. We make the hypothesis that zone of residence is closely related to sociodemographic characteristics. Therefore, a key aspect of the methodology is to estimate the zone of residence of users, which is used as a segmentation variable. The proposed method is applied to Santiago de Chile.

As in many large cities, the public transport system of Santiago, Chile (Transantiago) operates by using an AFC, where users validate when they board buses or enter a metro or bus station. The users are required to wave their cards near a validation device that reads the information and transacts the corresponding charge, allowing the passenger to enter the system. All of the payment transactions are recorded and contain information that is valuable for studying the public transport system. Several systems also have tap-off validation (Park et al., 2008; Ma et al., 2017), which can easily identify the origin and

* Corresponding author.
E-mail address: mamuniza@ing.uchile.cl (M.A. Munizaga).
[1] Currently at Citiplanning.
[2] Currently independent consultant.

http://dx.doi.org/10.1016/j.jtrangeo.2017.10.017
Received 5 May 2017; Received in revised form 12 September 2017; Accepted 24 October 2017
0966-6923/

destination stops of each trip stage. Others, such as Transantiago, do not require tap-off validation. In those cases, the alighting stop can be estimated by observing the daily sequence of transactions (Trepanier et al., 2007; Munizaga and Palma, 2012). Moreover, methods have been proposed to identify the trip destination, to distinguish transfer stops from activity destinations, and to assign an estimated purpose to each trip (Devillaine et al., 2012; Nassir et al., 2015). These results allow for the collection of valuable information regarding the travel behavior of public transportation users, such as travel patterns and time-use patterns.

From the time use perspective, previous studies have shown that income and other socioeconomic characteristics are relevant variables to explain time assignment. In Santiago it has been observed that users from different residential areas have different characteristics and this explains differences in the time use patterns of users from different residential areas (Jara-Díaz et al., 2013). If we are able to estimate residence zone for a sample of users, and observe the time use revealed by the trips made by them, we should expect to observe similar differences in the time use patterns to those observed by Jara-Díaz et al. (2013) using a sample obtained from the 2001 Santiago OD survey.

In this paper, a method to estimate the residence zones of frequent public transportation users is proposed. It is also applied, and the results are examined to verify if the assumption that justifies the effort does hold. Section 2 contains a brief description of the literature and a description of the context where this study was performed. The proposed method is presented in Section 3 and in Section 4 we present the results. Discussion and conclusions are presented in Section 5.

2. Background

The potential use of smartcard data as a complement of traditional travel survey methods was recognized by Bagchi and White (2005), who highlighted the quantity and temporal nature of this data source, and Utsonomiya et al. (2006), who performed their analysis using the Chicago transit system data. Pelletier et al. (2011) summarised the applications of these data performed until then, describing both the type of system and the uses given to the data. A number of studies have been dedicated to studying travel behavior, Morency et al. (2007) analyzed 10 consecutive months of data and observed weekly transactions to study the effect of holidays, among other things. Utsonomiya et al. (2006) studied the variability in the origin of the first trip of the day, focusing on a subsample of users whose addresses were known, to analyse differences by residence zone. Morency et al. (2007) analyzed 277 consecutive days of data and reached the conclusion that the most common behavior is to travel infrequently. The authors searched for spatial and time patterns using clustering methods. Lee and Hickman (2011) observe travel patterns of frequent public transport users in Minneapolis, US using smartcard data. Ma et al. (2013) also use cluster methods to search for time and space patterns in the massive Beijing, China dataset. Kusakabe and Asakura (2014) propose a data fusion methodology to estimate trip purpose, and use that information to make a frequency analysis over a long period that allows, for example, observing differences during summer time. Langlois et al. (2016) identify activity locations and perform cluster analysis to classify types of users according to the frequency of travel and locations visited. They perform a statistical analysis of the individuals in each cluster and found that they have different socioeconomic characteristics. Ma et al. (2017) use data mining methods to identify commuters, and characterize their travel patterns. As part of the analysis they identify home and work locations. They use a small survey to validate if their model labels the commuters in the sample correctly. The home and work locations are not validated.

Olguín et al. (2009) analyzed travel chains determined from the Santiago 2001 Origin-Destination survey (ODS) to study time-use patterns and determine differences by residence zone, age and gender. The study by Olguín et al. (2009) is interesting, though limited, because a

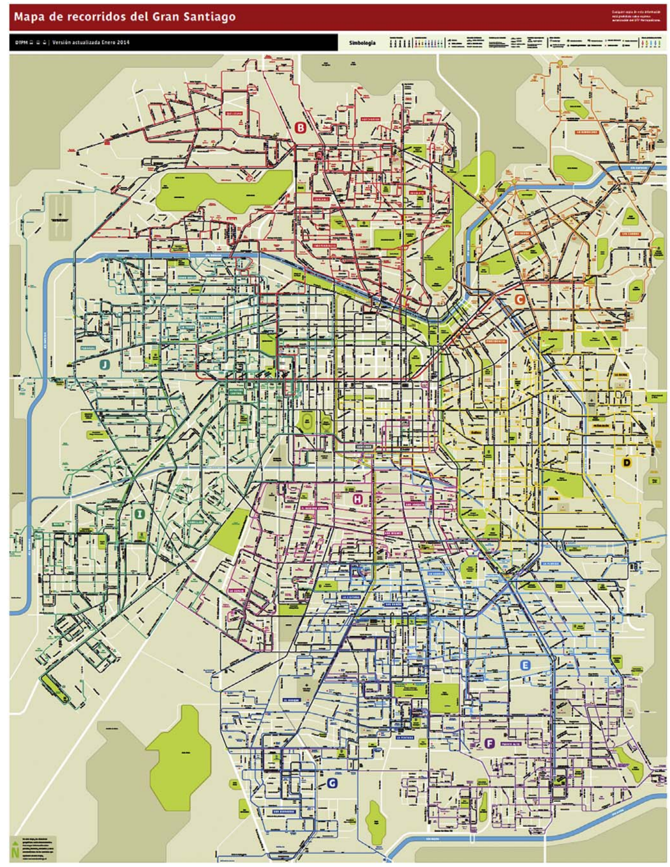

Fig. 1. Transantiago route map.
Source: http://www.transantiago.cl/imagenes/uploads/20160307134044-mapageneralfebrero2016.pdf.

large ODS was conducted only once every ten years and, due to its large cost, only a sample of users were surveyed, typically with one-day surveys. Devillaine et al. (2012) proposed to expand this analysis using massive smartcard data and, after applying a purpose imputation method, time-use profiles were generated. However, the results of this study can only be compared in aggregate because no socioeconomic information is available. Therefore, we cannot explore if the differences by residence zone found by Olguín et al. (2009) are also present in this larger, more reliable and replicable sample.

The public transport system available in Santiago, Chile, named Transantiago, has a wide coverage of the Santiago area and serves 34 of the 37 municipalities that are part of the greater metropolitan area (Fig. 1). The daily number of trips in a working day is 4621 thousands (Muñoz et al., 2015), which represents 27% of the total number of trips, and 41% of motorized trips. The public transport network has over 11,000 bus stops and 108 metro stations. 600 bus routes are operated with over 6500 buses; all equipped with GPS devices (see Gschwender et al., 2016).

Santiago is a spatially segregated city, and important differences in income by residence zone can be observed. Using the six large zones analysis proposed by Jara-Díaz et al. (2013) we built Fig. 2 using data from the 2012 Santiago ODS (Muñoz et al., 2015). Each zone is described in terms of average household characteristics, including household size, income and car ownership. Some mobility indicators are also provided. It can be seen that these large zones are indeed different. The East zone has the highest income, almost three times larger than the average income of the poorer zones, the highest car ownership rate and highest trip rate, but the lowest proportion of trips made by public transport. The Centre zone has the highest public transport market share, and the smallest household size. The rest of the