Travel Behaviour and Society xxx (2016) xxx–xxx

Contents lists available at ScienceDirect

Travel Behaviour and Society

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

Inferring weekly primary activity patterns using public transport smart card data and a household travel survey

Sergio Arturo Ordóñez Medina

Future Cities Laboratory, 1 Create Way # 06-01, Singapore 138602, Singapore

ARTICLE INFO

Article history:
Received 2 June 2016
Received in revised form 21 September 2016
Accepted 23 November 2016
Available online xxxx

Keywords:
Primary activities
Smart card data
Discrete choice models
Clustering algorithms
DBSCAN

ABSTRACT

With the introduction of smart cards for electronic payment in public transport, massive spatio-temporal data have been recorded in many cities. This data represents how public transport users from a whole city or country travel during consecutive days or weeks. This work presents a method to identify temporal weekly patterns of primary activities performed by public transport users in Singapore. As continuous travel data from the same user during a week can be extracted, work-leisure cycles can be recognized. Using revealed trips from a 1% sample household travel survey, activities reached and left by public transport were extracted and used to develop two discrete choice models: one for workers and one for students. According to the start time and duration of an activity, these models estimate the likelihood of such activity to be *HOME*, *WORK*/*STUDY* or *OTHER* type. Then, *consistent activities* were extracted from public transport smart card transactions recorded during one week. Using the discrete choice models, these extracted activities were classified by type. For each user, a 14-dimension vector summarizes his/her primary activity pattern. Each vector is composed by start time and duration of his/her work or study activities during continuous 7 days. To recognize the most common primary activity patterns of the city, the *DBSCAN* clustering algorithm was applied to these vectors. As the input data of this method is commonly available, the method can be applied to many cities in the world.

1. Introduction

With the introduction of smart cards for electronic payment in public transport, massive spatio-temporal data have been recorded in many cities. This data represents how public transport users from a whole city or country travel during consecutive days or weeks. It is also a very valuable data source to study travel behavior. However, as highlighted by Pelletier et al. (2011), recognizing travel or activity patterns by smart card data is a challenge because those systems were not designed to directly support planning or performance measurement.

Several studies have been developed to extract patterns from public transport transactions data. Devillaine et al. (2012) detect the activity purpose of public transport trips using smart card data of Santiago in Chile and Gatineau in Canada to compare behavioral activity patterns of users in these cities. Chakirov and Erath (2012) and Ordonez Medina and Erath (2013) use smart card data from Singapore to detect work activities and estimate work capacities respectively. They rely on land use and transportation information to infer the type of the activity and possible locations where users go after alighting at the public transport stops. Ordonez Medina and Erath (2013) cluster daily work activities according to their start time and duration given a fixed number of clusters, to categorize workers. El Mahrsi et al. (2014) recognize weekly traveling patterns of public transport users clustering passengers based on temporal profiles in Rennes, France. As the authors are only interested on the travel patterns, no activity detection is proposed. However, they contextualize the clusters into socioeconomic data of the regions where public transport users board and alight. A similar approach is followed by Ma et al. (2013); they use the *DBSCAN* (Ester et al., 1996) clustering algorithm to identify typical trip daily chains of each individual by means of his/her temporal and geographical transactions during the course of a week. With this recognized trip daily chains, the regularity of each user is represented by a vector of 4 features. Then, they use *K-means++* algorithm (Arthur and Vassilvitskii, 2007) to find clusters of similar users calculating euclidean distances between these vectors.

This work presents a method to identify temporal weekly patterns of primary activities performed by public transport users in Singapore. This method uses smart card transactions recorded during one week (CEPAS) and a 1% sample household travel survey (HITS). As this information is commonly available, this method is relevant in many cities. These temporal patterns are very useful

E-mail address: ordonez@ivt.baug.ethz.ch

http://dx.doi.org/10.1016/j.tbs.2016.11.005
2214-367X/

for transport planners as primary activities represent the majority of people's trips (in Singapore 79% of trips are made to perform primary activities according to HITS). Furthermore, as continuous travel data from the same user during a week can be extracted, work-leisure cycles can also be studied. Fig. 1 displays the processes proposed in this chapter.

Activities reached and left by public transport were extracted from revealed trips of a 1% sample household travel survey, and used to develop two discrete choice models: one for workers and one for students. According to the start time and duration of an activity, these models estimate the likelihood of such activity to be *HOME*, *WORK*/*STUDY* or *OTHER* type. The first section presents details of the Singaporean travel survey, the discrete choice models' definition and their prediction accuracy.

In the following, consistent activities were extracted from public transport smart card transactions recorded during one week. Using the mentioned discrete choice models, these extracted activities were classified by type. Then, for each user a 14-dimension vector summarizes his/her primary activity pattern. Each vector is composed by start time and duration of his/her work or study activities during continuous 7 days. The *DBSCAN* clustering algorithm was applied to these vectors in order to recognize the most common travel behaviors. The second section explains in detail these procedures, and presents the identified working and studying patterns.

In the conclusions, proposals of how the method could be extended are presented pointing out future steps. As mentioned before, this method can easily be implemented for other cities with similar data; thus, comparing results from different cities is a natural step forward.

2. Type of activity detection

This section describes how discrete choice models were developed to detect activity types using revealed trips from the household travel survey in Singapore. The methodology presented below is based on the work presented by Chakirov and Erath (2012). They propose to estimate 2 independent models by splitting the observations in two: workers and students. These models are designed to be applied to activities extracted from public transport smart card data, which represent about 97% of public transport users in Singapore (Prakasam, 2008), during 7 consecutive days in 2012. In this section, a summary of the travel survey information is provided, followed by a description of the implementation and evaluation of the discrete choice models.

2.1. Household interview travel survey summary

The household interview travel survey (HITS), carried out in 2012, asked for the daily trips of about 1% of Singapore's citizens and permanent residents. Start time, duration, date and purpose of each trip were reported. Table 1 presents the number of people, trips and stages reported in HITS 2012 by day of the week.

It is assumed that during the time between two consecutive revealed trips, only one activity is performed by the respondent. Only *public transport consistent* activities were extracted. *Public transport consistent* activities must meet two conditions (see Chakirov and Erath (2012)):

- The activity location must be reached and left by public transport.
- The alighting station of the previous trip must be less than kilometer away from the boarding station of the next trip.

With these two conditions, only activities that can be observed in the public transport smart card data are included in the estimation. When the first reported activity is the same than the last one, only one activity is extracted with the sum of the durations. Extracted activities are divided in two groups, student and worker activities. As the public transport system in Singapore offers special fares for students it is possible to recognize some of them according to the type of smart card they use.

HITS respondents must be categorized as workers or students in a consistent manner with public transport users recorded in the Smart card data. Table 2 presents the types of cards used in Singapore and how they are recorded. As shown in this table, tertiary

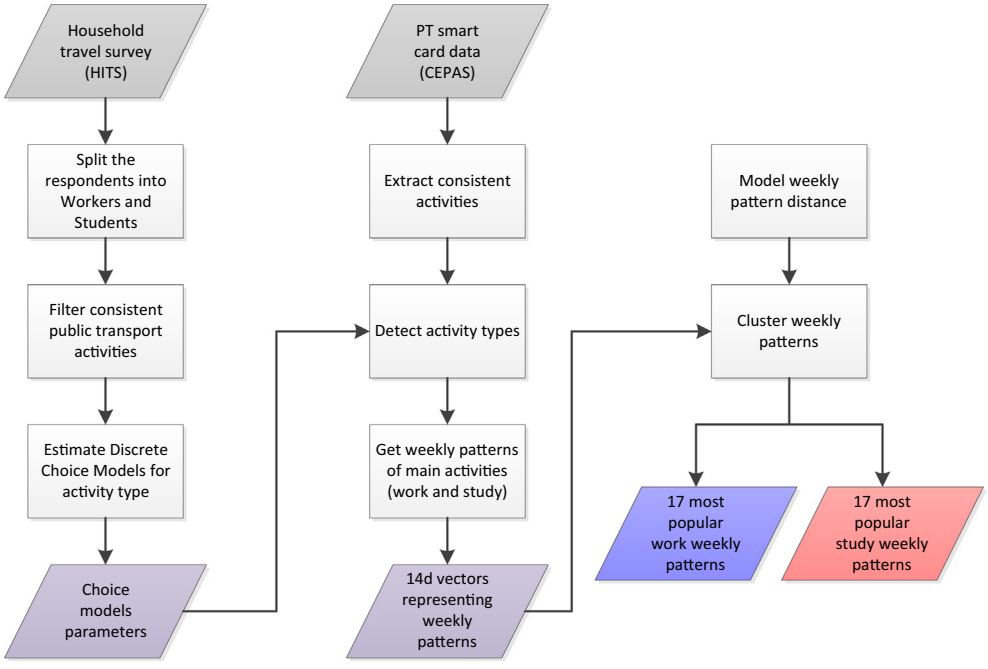

Fig. 1. Method to identify weekly activity patterns of public transport users in Singapore.