Journal of Transport Geography 58 (2017) 135–145

Contents lists available at ScienceDirect

Journal of Transport Geography

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

Understanding commuting patterns using transit smart card data

Xiaolei Ma [a,b,*,1], Congcong Liu [b], Huimin Wen [c,d,**,1], Yunpeng Wang [b], Yao-Jan Wu [e]

[a] Beijing Advanced Innovation Center for Big Data and Brain Computing (BDBC), Beihang University, Beijing 100191, China
[b] School of Transportation Science and Engineering, Beijing Key Laboratory for Cooperative Vehicle Infrastructure, Systems, and Safety Control, Beihang University, Beijing 100191, China
[c] Beijing Transportation Research Center, Beijing 100073, China
[d] Beijing Key Laboratory of Energy Conservation and Emission Reduction Detection and Evaluation in Transport, Beijing 100073, China
[e] Dept. of Civil Engineering and Engineering Mechanics, The University of Arizona, 1209 E 2nd St. Room 324F, Tucson, AZ 85721, United States

ARTICLE INFO

Article history:
Received 30 April 2016
Accepted 1 December 2016
Available online 9 December 2016

Keywords:
Commuting
Human mobility
Public transportation
Travel behavior
Transit smart card data

ABSTRACT

Commuting reflects the long-term travel behavior of people and significantly impacts urban traffic congestion and emission. Recent advances in data availability provide new opportunities to understand commuting patterns efficiently and effectively. This study develops a series of data mining methods to identify the spatiotemporal commuting patterns of Beijing public transit riders. Using one-month transit smart card data, we measure spatio-temporal regularity of individual commuters, including residence, workplace, and departure time. This data could be used to identify transit commuters by leveraging spatial clustering and multi-criteria decision analysis approaches. A disaggregated-level survey is performed to demonstrate the effectiveness of the proposed methods with a commuter identification accuracy that reaches as high as 94.1%. By visualizing the spatial distribution of the homes and workplaces of transit commuters, we determine a clear disparity between commuters and noncommuters and confirm the existence of job–house imbalance in Beijing. The findings provide useful insights for policymakers to shape a more balanced job–housing relationship by adjusting the monocentric urban structure of Beijing. In addition, the extracted individual-level commuting patterns can be used as valuable information for public transit network design and optimization. These strategies are expected to reduce car dependency, shorten excess commute, and alleviate traffic congestion.

1. Introduction

As the primary component of personal daily travel, commuting considerably influences urban traffic conditions (Zhou et al., 2014). The job–housing spatial imbalance forces people to endure long commuting time, and thus, results in excessive travel time and fuel consumption (Van Acker and Witlox, 2011; Charron, 2007). To mitigate these adverse effects, transportation planners and operators strive to reduce travel demand and improve commuting efficiency through policies or technological countermeasures (Lovelace et al., 2014; Li et al., 2013a). Among these measures, prioritizing public transportation systems is considered as one of the most effective strategies because it can significantly reduce car dependency, mitigate traffic congestion, and alleviate air pollution (Ma and Wang, 2014; Li et al., 2013b; Zhao, 2013). Understanding transit commuting patterns offers valuable insights into the spatial and temporal relationship between transit commuters' residences and workplaces (Zhao, 2013; Louf and Barthelemy, 2014). These insights will highlight the critical need for properly designing public transport networks to establish job–housing balance (Zhou et al., 2014). Unlike traditional zonal or regional travel commuting behavioral studies, individual-level commuting patterns provide more detailed and useful information to refine existing travel demand forecasting models with high-resolution human mobility (Yang et al., 2014; Ren et al., 2014). In addition, targeting individual transit commuters using fare reduction for ridership attraction is necessary to improve public transportation systems usage rates (Ma et al., 2013). However, extracting the commuting behavior of an individual transit rider is not a straightforward task. Conventional public transport behavioral studies rely on household travel surveys or diaries to obtain personal profile, socioeconomic and demographic information, and travel patterns (Louail et al., 2014; Schneider et al., 2013; Jiang et al., 2012). This process is costly, time-consuming, and frequently results in a low sampling rate and a small population size. When ask to participate in multi-day travel surveys, people are usually reluctant to respond because of survey fatigue, which will further decrease data availability and accuracy (Mahrsi et al., 2014).

The recent developments of emerging data sources and statistical methods have created opportunities to analyze and determine transit commuting behavior at an individual level over long-term periods. Transit smart card data from automatic fare collection systems are

* Correspondence to: X. Ma, Beijing Advanced Innovation Center for Big Data and Brain Computing (BDBC), Beihang University, Beijing 100191, China.
** Correspondence to: H. Wen, Beijing Transportation Research Center, Beijing 100073, China.
E-mail addresses: xiaolei@buaa.edu.cn (X. Ma), vincent_liu@buaa.edu.cn (C. Liu), wenhm@bjtrc.org.cn (H. Wen), ypwang@buaa.edu.cn (Y. Wang), yaojan@email.arizona.edu (Y.-J. Wu).
[1] These authors contributed equally to this work.

http://dx.doi.org/10.1016/j.jtrangeo.2016.12.001
0966-6923/

widely adopted by transit authorities to manage revenue as well as to gather abundant passenger boarding and alighting information in a disaggregated manner (Kusakabe and Asakura, 2014). Compared with traditional data collection methods, smart card data can record the day-to-day variability in the travel patterns of an individual transit rider and have the potential to identify transit commuters and detect their spatial and temporal regularity through a continuous long-term observation period (Schneider et al., 2013). A large number of behavioral studies using smart card data have gained more and more popularities. Morency et al. (2007) applied data mining techniques to examine the spatial and temporal variability of transit use based on smart card data. Ma et al. (2013) developed a Density-Based Scanning Algorithm with Noise (DBSCAN) to categorize different groups of transit riders with varying travel patterns. Kieu et al. (2015a, b) improved the original DBSCAN algorithm proposed by Ma et al. (2013), and significantly reduced the algorithm complexity with the same clustering performance. Langlois et al. (2016) utilized four-week transit smart transaction data to identify transit rider heterogeneity, and generated 11 clusters with distinct activity sequences and demographic attributes. Ali et al. (2015) developed a large-scale activity based public transport simulation platform based on MATSim, and used the smart card as an input. They demonstrated the feasibility of applying smart card data into microsimulation travel demand models. As documented by Pelletier et al. (2011), new analytical methods and disaggregate approaches based on smart card data are renovating traditional travel behavior research.

In the current literature on transit behavioral studies on commuting travel patterns, the definition of a transit commuter is oversimplified. The majority of these studies only considered the repeatability of temporal activities (e.g., transit riders traveling for at least a number of days are determined to be transit commuters) (Long et al., 2012; Zhou et al., 2014). Several researchers simultaneously incorporated both the spatial and temporal regularities of recurring travels into transit passenger segmentation studies. Ortega-Tong (2013) analyzed the spatial and temporal travel patterns of London transit riders with Oyster Cards, and grouped them into 8 clusters based on different sociodemographic characteristics and activity patterns. The transit riders traveling 4 days a week or more were categorized as regular users. Kieu et al. (2015b) proposed a transit passenger segmentation method based on smart card data. The method identified the spatial travel pattern of each transit rider using a two-step DBSCAN algorithm, and a k-means algorithm was the applied to distinguish frequent and infrequent transit users based on the number of travel days and journeys made. Ma et al. (2013) developed a density-based clustering algorithm to mine each transit rider's spatial and temporal travel pattern in Beijing, and then proposed a K-means++ algorithm and Rough Set based approach to measure travel regularity. More than 40% transit riders were identified as frequent passengers. Kung et al. (2014) used mobile phone data to understand home-work commuting behaviors at three countries (Portugal, Ivory Coast, and Saudi Arabia) and one city (Boston). Individual's home and work locations can be identified by a spatial and temporal filtering approach. They found that the home-work commuting time distribution is independent of commute distance. However, the above studies are not specifically designed for transit commuter identification. Even if each transit rider's spatial and temporal travel pattern can be mined in some literatures, how the identified spatiotemporal travel pattern tie to one's commuting behavior is still not clear. Fortunately, the rapid development of data mining and statistical techniques has facilitated finding underlying and previously hidden information through large-scale data processing (Jiang et al., 2012), and thus can be applied to transit commuting pattern mining using smart card data. Both the spatial and temporal features of the commuting trips of transit riders should be considered in a quantitative and synthetic manner.

This study seeks to answer the following important questions: Which transit riders can be classified as transit commuters? Can we infer the residence and workplace of an individual transit commuter, as well as his/her commuting departure time, based on his/her long-term smart card transaction records? To answer these questions, we first extract the spatial and temporal features of an individual transit rider that can represent the regularity of commuting patterns. These features are calculated based on continuous trip-chaining behavior from one-month of smart card data in Beijing. These features fully incorporate the heterogeneity of individual route choices and the uncertainty of departure times using a modified DBSCAN clustering algorithm. A spatial clustering algorithm, called iterative self-organizing data analysis technique (ISODATA), is then utilized to categorize group transit riders into three clusters based on their spatiotemporal features. The three clusters are automatically determined and can reflect the intensity of transit commuting travel. Consequently, the residence and workplace of each transit commuter can be derived by summarizing his/her most frequently visited locations. For each cluster, we also develop a transit commuting score based on the technique for order of preference by similarity to ideal solution (TOPSIS) method. This method can score each individual transit rider based on his/her commuting patterns via a multi-criteria decision analytical framework. The score threshold for distinguishing commuters from noncommuters routinely remains at approximately 51.7 even when different groups of transit riders are randomly selected. Instead of clustering the total population of transit riders, we can identify transit commuters only by calculating the individual commuting score. This approach will significantly reduce computational power. Finally, the proposed methods are validated via a survey conducted in Beijing and visualized in map platform to demonstrate their effectiveness.

2. Methodology

2.1. Trip generation

The public transportation systems in Beijing consist of over 1000 bus routes and 18 subway lines in 2015 (Fig. 1), and these figures result in over 28,000 bus and subway stops. The percentage of subway and bus trips among all motorized trips reached 60.1% at the end of 2014 because of the expansion of subway lines and the governmental subsidies for public transit (Beijing Transportation Research Center, 2015). Over 90% of transit riders utilize smart cards to pay fares because a huge discount rate is received by smart card holders (i.e., 75% fare reduction for students and 50% fare reduction for regular passengers) (Ma et al., 2012). Since January 2015, all buses and subway lines have adopted distance-based fare strategies in which both passenger tap-in and tap-out data (e.g., route ID, transaction times, and boarding and alighting stops) for an individual transit rider are recorded. Beijing Automatic Fare Collection (AFC) system includes urban transit, rural transit and subway systems. Buses running in the rural area record passengers' boarding times/stops and alighting time/stops. Similarly, subway AFC system also contains passengers' full trip information. However, buses running in the urban area only store individual's boarding stop and alighting time/stop. The percentage of records with missing boarding times is approximately 40% of total smart card transactions. For the urban AFC system, the bus boarding time of each transit rider can be calculated by using the transaction time (i.e. alighting time) minus the average in-vehicle travel time. According to the 2015 Beijing Transport Annual Report, the average in-vehicle travel time during morning peak hours is 59.6 min, while this time increases to 65.3 min during evening peak hours (Beijing Transportation Research Center, 2015). We collected one-month smart card data (June 2015). A total of 18 million active transit riders were identified, thus generating a total of over 364 million smart card transaction records.

We generate individual trips on a daily basis. A trip is composed of a sequence of activities for a particular purpose (Primerano et al., 2008). In the context of public transport, a transit rider may transfer from one bus route to a subway line and continue to take another bus route to his/her destination. In the above example, the trip is associated with three smart card transactions (bus, subway and bus). Time