Journal of Transport Geography 70 (2018) 78–90

Contents lists available at ScienceDirect

Journal of Transport Geography

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

Do different datasets tell the same story about urban mobility — A comparative study of public transit and taxi usage

Xiaohu Zhang[a], Yang Xu[b,*], Wei Tu[c,d], Carlo Ratti[d]

[a] Singapore-MIT Alliance for Research and Technology, 1 Create Way, Singapore
[b] Department of Land-Surveying and Geo-Informatics, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong
[c] Shenzhen Key Laboratory of Spatial Information Smart Sensing and Services, School of Architecture and Urban Planning & Research Institute for Smart Cities, Shenzhen University, Shenzhen, China
[d] Senseable City Laboratory, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139, USA

ARTICLE INFO

Keywords:
Human mobility
Urban spatial structure
Smart card transactions
Taxi data
Spatiotemporal analysis

ABSTRACT

Understanding human movements and their interactions with the built environment has long been a research interest in transport geography. In recent years, two important types of urban mobility datasets — smart card transactions and taxi GPS trajectories — have been used extensively but often separately to quantify travel patterns as well as urban spatial structures. Despite the fruitful research outcomes, the relationships between different types of transport flows in the same geographic area remain poorly understood. In this research, we propose an analytical framework to compare urban mobility patterns extracted from these two data sources. Using Singapore as a case study, this research introduces a three-fold comparative analysis to understand: (1) the spatial distributions of public transit and taxi usages and their relative balance; (2) the distance decay of travel distance, and (3) the spatial interaction communities extracted from the two transport modes. The research findings reveal that the spatial distributions of travel demand extracted from the two transport modes exhibit high correlations. However, more in-depth analysis (based on rank-size distribution and log odds ratio) reveals a higher degree of spatial heterogeneity in public transit usage. The travel distance of trips from public transit decays faster than that of taxi trips, highlighting the importance of taxis in facilitating long-distance travels. Both types of trips decay much faster when travel distance is beyond 20 km, which corresponds to the average distance from the urban periphery to the center. The spatial interaction communities derived from public transit are different on weekdays and weekends, while those of taxis show similar patterns. Both transport modes yield communities that reveal the city's polycentric structure, but their differences indicate that each of the transport modes plays a specific role in connecting certain places in the city. The study demonstrates the importance of comparative data analytics to urban and transportation research.

1. Introduction

The past two decades have witnessed an exponential growth of scientific research that characterizes human mobility and their interactions with the built environment. The rapid developments of information and pervasive sensing technologies have produced – especially in urban settings – a wide spectrum of human mobility datasets, empowering researchers to tackle critical questions in transport planning (Santi et al., 2014; Alexander et al., 2015; Tu et al., 2016), disease control (Bengtsson et al., 2011; Wesolowski et al., 2012), and social dynamics (Cho et al., 2011; Xu et al., 2017; Sun et al., 2013). The big data evolution has spurred “a new science” or many new sciences of cities, from which urban environments can be better understood as systems of networks and flows (Batty, 2013).

The networks and flows embedded in cities are defined by researchers through different types of datasets, resulting in a multi-faceted view of urban mobility patterns. For example, many studies have been conducted in recent years to quantify intra-urban mobility patterns based on taxicab usages (Wang et al., 2015; Liu et al., 2015; Kang and Qin, 2016), public transit data (Zhong et al., 2015; Liu et al., 2009), and mobile phone records (Gao et al., 2013; Ahas et al., 2010; Xu et al., 2016). Despite the fruitful research outcomes, most of the existing studies focus on a single type of human mobility dataset, which yields into insights that are somewhat isolated. The relationships between different types of networks and flows in the same geographic area – such as a city – remain poorly understood (Tu et al., 2018). It is, therefore, important to combine different data sources to obtain a

* Corresponding author.
E-mail address: yang.ls.xu@polyu.edu.hk (Y. Xu).

https://doi.org/10.1016/j.jtrangeo.2018.05.002
Received 31 October 2017; Received in revised form 25 March 2018; Accepted 3 May 2018
0966-6923/

comprehensive view of the spatial structures and organizations of cities. This will shed light on the bias when each data type is used alone to represent the dynamics of urban systems. More importantly, it would generate a deeper understanding of the interplay among different socio-economic processes.

In this research, we propose an analytic framework to compare urban travel patterns and the associated urban spatial structures extracted from smart card transactions and taxi GPS trajectories. The two types of datasets are widely used but often separately in revealing urban mobility patterns. Using Singapore as a case study, this work aims to fill the research gap by answering the following research question – do public transit and taxi usages in a city produce similar patterns of travel demand, travel distance, and urban spatial structures?

Based on origin-destination (OD) trips extracted from smart card transactions and taxi GPS trajectories — both cover a one-week period in Singapore — this study performs a three-fold comparative analysis. First, we analyze trip origins and destinations separately by focusing on the spatial distributions of outgoing and incoming trips. Two measures, namely rank-size distribution and log odds ratio, are introduced to quantify and compare spatial heterogeneity of travel demand extracted from the two datasets. We then examine spatial variations and statistical properties of travel distance (e.g., distance decay effect) to better understand the service radius of public transport and taxis in different parts of the city. Finally, we apply a community detection algorithm to the OD matrices to uncover the hidden spatial structures embedded in these two transport modes.

The remainder of this article is organized as follows. Section 3 provides an overview of related work of this research. Section 4 introduces the study area and the two mobility datasets. In Section 5, we introduce the approaches and measures for conducting the three-fold comparative analysis. We then present analysis results in Section 6. Finally, in Section 7, we conclude our findings and discuss future research directions.

2. Interplay between public transit and taxi services

Urban travel patterns are the outcome of the complex interactions between land use configuration and individual characteristics. The land use system governs spatial distribution of opportunities (in commercial, industrial areas) and the demand for these opportunities (in residential areas) (Geurs and van Wee, 2004). It determines intra-urban movement of people and goods from a macro level, which is modeled by trip generation and trip distribution in the "four-step" urban transportation planning process (Pas, 1995). The classical urban transport model assumes the trip amount between different zones is proportional to the number of households at the origin and the number of opportunities at the destination. However, when zooming into local neighborhood, the share disparity of different transport modes is more sensitive to individual characteristics such as income, education and vehicle ownership. This share disparity is captured in mode choice model which is used to forecast individuals' travel behaviors based on microeconomic theory (Ben-akiva and Bierlaire, 2003).

Public transit and taxi — two typical transport modes — are different in its service deployment and customer behaviors. The public transportation system is only considered by city government when population density reaches a certain level. Accessibility to transit service is higher around bus stops or metro stations than other places. It means intra-urban connectivity by public transit is enhanced at areas around bus stops or metro stations but not uniformly along bus/metro route. Unlike public transit, deployment of taxi service does not require high population density. It can be taken from anywhere, not constrained by a limited number of pick-up locations, although the service may be more accessible in some areas. In the areas where no public transit is provided, taxi complements its service; while it also competes with public transit as an alternative transport means with high flexibility in time and at a higher cost.

Human mobility datasets generated by public transportation and taxi can be used to reveal urban travel patterns. Mining the variances of the travel patterns will provide hints related to the local land use, socio-demographics, and the city structure. For example, a temporal profile of outgoing and incoming trips can be used to infer land use (Liu et al., 2012a). The mode share disparity may suggest differences of socio-demographics in different areas. Also, the flow of people and goods connects urban spaces which may have implications for spatial interactions in different regions. These spatial interactions reflect economic activities and reveal the underlying urban structures (Sun et al., 2014).

3. Related work

Understanding human mobility patterns has been a long standing research interest in areas such as urban planning, transportation, and geography. Before information and communication technologies (ICT) pervaded, travel surveys were used as the primary data source to support studies of human travel and daily activities. These studies cover important subjects such as trip chaining analysis (Hanson, 1980; Kitamura, 1984), characterization of human activity space (Newsome et al., 1998; Dijst, 1999; Schönfelder and Axhausen, 2003; Tu et al., 2017), and relationships between travel behavior and socio-economic characteristics (Hanson and Hanson, 1981; Kwan, 1999). The sampling schemes in these survey-based studies are often carefully designed, and the datasets usually contain detailed information of respondents. On the downside, however, the sample sizes are usually limited by the human and financial resources available.

With rapid developments of information and pervasive sensing technologies, researchers nowadays are able to access bigger and more diversified datasets, leading to a new paradigm of data-intensive science (Hey et al., 2009). The ways human and urban mobility can be measured are greatly enriched by datasets such as smart card transactions (Liu et al., 2009), taxi GPS trajectories (Li et al., 2011), and mobile phone records (Blondel et al., 2015). These datasets have both pros and cons when used in human mobility research, and they often reflect different or sometimes overlapping dimensions of human activities. For instance, taxi GPS trajectories record movements of taxicabs as well as their occupancy status over space and time. Such data have been used in previous studies to gain insights into cabdrivers' operation strategies (Li et al., 2011; Kang and Qin, 2016; Liu et al., 2010), urban traffic conditions (Castro et al., 2012), hot spots of taxi pick-up and drop-off points (Wang et al., 2009), and benefits of ridesharing in cities (Santi et al., 2014). Smart card transactions, on the other hand, often collect information about people's usage of public transit (e.g., card id as well as location and time for boarding/alighting). Due to the abilities to capture public transit usage for large populations, such data have been widely used to derive ridership statistics and performance indicators, and to guide transit planning and service improvements (Pelletier et al., 2011). In recent years, mobile phone data have gained increasing attention on human mobility research. Call detail records (CDRs) – a typical type of mobile phone data – have been widely used to characterize intrinsic properties of human moevements (Gonzalez et al., 2008; Song et al., 2010a, 2010b), people's use of urban space (Becker et al., 2013; Xu et al., 2015), and the interplay between human travel and social relations (Cho et al., 2011). Unlike taxi tracking and smart card transactions that are tied to specific means of transportation, mobile phone data capture snapshots of activities for large phone user pools, enabling a broader but mixed view of travelers' mobility patterns.

Since all the three types of data mentioned above include useful information about how people move from one place to another, they have all been used in previous studies to understand dynamics of population flows and intra-urban spatial structures. However, most of the