Computers, Environment and Urban Systems xxx (2015) xxx–xxx

Contents lists available at ScienceDirect

Computers, Environment and Urban Systems

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

Combining smart card data and household travel survey to analyze jobs–housing relationships in Beijing

Ying Long [a,*], Jean-Claude Thill [b]

[a] Beijing Institute of City Planning, Beijing 100045, China
[b] Department of Geography and Earth Sciences, The University of North Carolina at Charlotte, Charlotte, NC 28223, USA

ARTICLE INFO

Article history:
Available online xxxx

Keywords:
Bus smart card data
Jobs–housing spatial mismatch
Commuting trip
Rule-based
Beijing

ABSTRACT

Location Based Services (LBS) provide a new perspective for spatiotemporally analyzing dynamic urban systems. Research has investigated urban dynamics using LBS. However, less attention has been paid to the analysis of urban structure (especially commuting pattern) using smart card data (SCD), which are widely available in most large cities in China, and even in the world. This paper combines bus SCD for a one-week period with a oneday household travel survey, as well as a parcel-level land use map to identify job–housing locations and commuting trip routes in Beijing. Two data forms are proposed, one for jobs–housing identification and the other for commuting trip route identification. The results of the identification are aggregated in the bus stop and traffic analysis zone (TAZ) scales, respectively. Particularly, commuting trips from three typical residential communities to six main business zones are mapped and compared to analyze commuting patterns in Beijing. The identified commuting trips are validated by comparison with those from the survey in terms of commuting time and distance, and the positive validation results prove the applicability of our approach. Our experiment, as a first step toward enriching LBS data using conventional survey and urban GIS data, can obtain solid identification results based on rules extracted from existing surveys or censuses.

1. Introduction

This paper identifies job–housing location dyads and commuting patterns in Beijing using smart card data (SCD) that store the daily trip information of bus passengers. It proposes and implements a method for deriving commuting patterns from increasingly common SCD for informing city planners and transit system managers about patterns of transit usage across space and through time as well as about mobility patterns in a large and fast growing city region. Related research on jobs–housing relationships has conventionally used data acquired through surveys or censuses. The increasing pervasiveness of location-based services (LBS) associated with the prevalence of positioning technologies has led to the creation of large-scale and high-quality space-time datasets (Jiang & Yao, 2006). This development has also created opportunities to better describe and understand urban structures[1] in multiple dimensions. These datasets have been shown to be important for analyzing urban and environmental systems such as relationships between housing and jobs (Batty, 1990). Meanwhile, a geo-tagged smart card system is an effective alternative tool for individual data acquisition necessary to analyze urban spatial structures.

Various types of fine-granularity individual data generated by LBS technologies have been extensively leveraged to analyze urban structures (Ahas & Mark, 2005; Lu & Liu, 2012). With respect to handheld Global Positioning System (GPS) devices, Newhaus (2009) used location data to record and visualize urban diaries, while Gong, Chen, Bialostozky, and Lawson (2012a) elicited travel modes of travelers in New York City. Liu, Andris, and Ratti (2010) identified taxi drivers' behavior patterns from their daily digital trajectories, and Yue et al. (2012) used these trajectories to calibrate a spatial interaction model. With respect to mobile phone systems (see Steenbruggen, Borzacchiello, Nijkamp, & Scholten, 2013 for a review), Ratti, Pulselli, Williams, and Frenchman (2006) evaluated the density and spatiotemporal characteristics of urban activities using mobile phone data in Milan, Italy, whereas Wan and Lin (2013) studied fine-scale individual activities, Yuan, Raubal, and Liu (2012) correlated mobile phone usage and city-wide travel behavior in Harbin, China and Chi, Thill, Tong, Shi, and Liu (In press) exploited network properties of mobile phone

* Corresponding author. Tel.: +86 10 88073660; fax: +86 10 68031173.
E-mail address: longying1980@gmail.com (Y. Long).

[1] The concept of "urban structure" concerns the spatial arrangement of public and private spaces in cities and the degree of connectivity and accessibility. In this paper, the concept is focused particularly on the spatial concentration of resident population and employment (Anas, Arnott, & Small, 1998).

http://dx.doi.org/10.1016/j.compenvurbsys.2015.02.005
0198-9715/

data to reveal urban hierarchical structures at the regional scale. As for Wi-Fi, Rekimoto, Miyaki, and Ishizawa (2007) used Wi-Fi-based location detection technology to log the locations of device holders from received Wi-Fi beacon signals, a technology that works both indoors and outdoors. Torrens (2008) developed a system to detect Wi-Fi infrastructure and transmission and analyze their geographic properties, and tested this system in Salt Lake City, Utah.

Meanwhile, the discipline of time geography established by Hagerstrand (1970) also benefited from the development of LBS by retrieving more objective data. In sum, various LBS technologies have been successfully applied in urban studies. However, these technologies remain immature and most research on urban structure continues to employ data from the urban physical space or questionnaire surveys (with a few studies as exceptions, e.g. Kwan (2004)). Access to large-scale micro datasets remains a barrier to their widespread use for research, planning and management (Long & Shen, 2013).

A smart card that records full cardholder's bus trip information is an alternative form of location-acquisition technology. Smart card automated fare collection systems are increasingly deployed in public transit systems. Along with collecting revenue, such systems can capture a meaningful portion of travel patterns of cardholders, and the data are useful for monitoring and analyzing urban dynamics. Since the 1990s, the use of smart cards has become significant owing to the development of the Internet and the increased complexity of mobile communication technologies (Blythe, 2004). As of 2007, Intelligent Transportation Systems (ITS) that incorporate smart card automated fare systems either existed or were being established in over 100 Chinese cities, as well as in many other cities around the world (Zhou, Zhai, & Gao, 2007). The data generated by smart card systems track the detailed onboard transactions of each cardholder. We argue that smart card technology can deliver valuable information because it is a continuous data collection technique that provides a complete and real-time bus travel diary for all bus travelers. SCD can be used to validate traditional travel models applied to public transit. In contrast to SCD collection, conventional travel behavior surveys have the drawbacks of being expensive and infrequent. Notably, transit SCD collects data in fundamentally the same way as an AVI (automatic vehicle identification) system, which has been widely used in the United States to automatically identify vehicles. AVI is used in some states in the US for planning purposes. One such example is New York, where the E-ZPass tag is used as part of the TRANSMIT system.

Previous studies have advocated using SCD to make decisions on the planning and design of public transportation systems (see Pelletier, Trepanier, and Morency (2011) for a review). In South Korea, Joh and Hwang (2010) analyzed cardholder trip trajectories using bus SCD from ten million trips by four million individuals, and correlated these data with land use characteristics in the Seoul Metropolitan Area. Jang (2010) estimated travel time and transfer information using data on more than 100 million trips taken in Seoul on the same system. Roth, Kang, Batty, and Barthélemy (2011) used a real-time "Oyster" card database of individual traveler movements in the London subway to reveal the polycentric urban structure of London. Gong et al. (2012b) explored spatiotemporal characteristics of intra-city trips using metro SCD on 5 million trips in Shenzhen, China. Also, Sun, Axhausen, Lee, and Huang (2013) used bus SCD in Singapore to detect familiar "strangers".

There is considerable research on inferring home and job locations from individual trajectories like mobile phone call data records and location-based social networks (LBSN). For the identification of home locations, Lu, Wetter, Bharti, Tatem, and Bengtsson (2013) regarded the location of the last mobile signal of the day as the home location of a mobile user. The most frequently visited point-of-interest (POI) (Scellato, Noulas, Lambiotte, & Mascolo, 2011) or grid cell (Cheng, Caverlee, Lee, & Sui, 2011; Cho, Myers, & Leskovec, 2011) was regarded as a LBSN user's home location. It is not easy to infer home locations from LBSN with a high spatial resolution. Compared to approaches to home location identification, there are fewer studies on identifying job locations based on trajectories, with Cho et al. (2011) using LBSN and Isaacman et al. (2011) using cellular network data as notable exceptions. It should be mentioned that taxi trajectories are not well suited for identifying a passenger's home and job locations considering the passenger-sharing nature of taxis. However, less attention has been paid to using SCD to identify home and job locations as well as to analyze job–housing dyadic relationships and commuting patterns in a metropolitan region.

This paper regards job–housing dyadic relationships and commuting pattern analysis as a showcase for using SCD to urban spatial analysis. We argue that job and home locations, their dyadic relationships, and commuting trips can be identified from SCD and serve as valuable information on the modalities of use of the urban space in its residents. We propose a methodology to this effect and use Beijing as a case study to test its implementation. The identification results are validated using travel behavior survey data from Beijing. This paper is organized as follows. The retrieval of job–housing trips from conventional travel behavior surveys is discussed in Section 2, and the SCD and other related datasets used in our research are presented in Section 3. The approaches for identifying home and job locations, as well as commuting trips are elaborated in Section 4. In Section 5, the results of job–housing identification and commuting patterns are shown and analyzed in detail. Finally, we discuss our work and present concluding remarks in Sections 6 and 7, respectively.

2. Job–housing trips in conventional travel behavior surveys

Travel behavior surveys have been the primary means of data collection on urban resident travel behavior for planning and managing urban transit systems (Beijing Transportation Research Center, 2009). There is a well-established tradition in geography and urban planning to use surveys for tracking individual travel diaries (Gärling, Kwan, & Golledge, 1994; Schlich, Schönfelder, Hanson, & Axhausen, 2004). Travel behavior surveys track traveler socio-economic attributes, as well as trip origin and destination, time and duration, as well as trip purposes and travel modes. On the one hand, the traveler's home and job locations are directly recorded in the survey together with his/her socioeconomic attributes, and both locations are mostly aggregated in the traffic analysis zone (TAZ) scale. On the other hand, trips between work and home (i.e. commuting trips) can be screened using the purpose attribute. These trips are also recorded using the inter-TAZ scale rather than a finer spatial scale. Therefore, job–home location dyads and trips have already been recorded in conventional travel surveys, but mostly at the TAZ scale. Additionally, only a small portion of all households in a given city are surveyed due to time and cost constraints.

Compared with travel behavior surveys, mining the enormous volume of SCD can provide a more precise spatial resolution and a much larger sample, despite the SCD being unable to directly provide job–home location dyads and commuting trips. We will focus on using SCD to identify jobs–housing relationships. Patterns of commuting trips from typical residential communities or to typical business centers can be visualized by identifying the results at a finer scale than is available in travel surveys because residential communities or business zones are generally smaller than a TAZ. A more detailed commuting pattern is expected to reveal fresh information on jobs–housing relationships in a megacity such as