Transportation Research Part C 26 (2013) 301–313

Contents lists available at SciVerse ScienceDirect

Transportation Research Part C

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

Understanding individual mobility patterns from urban sensing data: A mobile phone trace example

Francesco Calabrese [a,d], Mi Diao [b,e,*], Giusy Di Lorenzo [a,d], Joseph Ferreira Jr. [c], Carlo Ratti [a,c]

[a] SENSEable City Lab, Massachusetts Institute of Technology, USA
[b] Department of Real Estate, National University of Singapore, Singapore
[c] Department of Urban Studies and Planning, Massachusetts Institute of Technology, USA
[d] IBM Research, Dublin, Ireland
[e] Institute of Real Estate Studies, National University of Singapore, Singapore

ARTICLE INFO

Article history:
Received 18 October 2010
Received in revised form 22 September 2012
Accepted 24 September 2012

Keywords:
Mobility analysis
Mobile phone traces
Vehicle Kilometers Traveled (VKT)

ABSTRACT

Large-scale urban sensing data such as mobile phone traces are emerging as an important data source for urban modeling. This study represents a first step towards building a methodology whereby mobile phone data can be more usefully applied to transportation research. In this paper, we present techniques to extract useful mobility information from the mobile phone traces of millions of users to investigate individual mobility patterns within a metropolitan area. The mobile-phone-based mobility measures are compared to mobility measures computed using odometer readings from the annual safety inspections of all private vehicles in the region to check the validity of mobile phone data in characterizing individual mobility and to identify the differences between individual mobility and vehicular mobility. The empirical results can help us understand the intra-urban variation of mobility and the non-vehicular component of overall mobility. More importantly, this study suggests that mobile phone trace data represent a reasonable proxy for individual mobility and show enormous potential as an alternative and more frequently updatable data source and a compliment to the conventional travel surveys in mobility study.

1. Introduction

The transportation sector plays an important role in global sustainability. In 2004, it accounts for 22% of primary energy use and 27% of CO_2 emissions all over the world (de Ia Rue du Can and Price, 2008). Individual mobility consumes about two thirds of the total transportation energy use.[1] Understanding the intra-urban variation of individual mobility is important for policy makers to perform “what if” analysis of the environmental consequences of alternative development scenarios and land use controls, and develop regional growth strategies towards a more sustainable future.

The majority of empirical studies on mobility rely on travel surveys, because they provide detailed descriptions of demographics, place of residence, and travel attributes at an individual or household level to support modeling. However, travel surveys are not without limitations:

* Corresponding author at: Department of Real Estate, National University of Singapore, Singapore. Tel.: +65 65163418.
E-mail address: rstdm@nus.edu.sg (M. Diao).

[1] https://www.iea.org/impagr/cip/archived_bulletins/issue_no23.htm.

0968-090X/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.trc.2012.09.009

- The usage of small sample due to the high expense of travel surveys. For example, the National Household Travel Survey "add-on" program is among the largest available travel surveys, but it involves at most 10,000 observations for participating states and a few thousand observations for individual metropolitan areas. As a result, there are not many respondents included in any one neighborhood, which limits the efforts to adequately understand travel patterns for small areas (Handy, 1996).
- The limitation in spatial and temporal scales of the collected datasets. Privacy concerns often limit the geographic specificity with which trip origins and destinations can be revealed, thus spatial effects at fine-grained scales cannot be identified. The surveys normally only collect the travel diary of households within 1–2 days due to concerns in respondent burden. Therefore, complete household activity schedules cannot be observed and some important mobility patterns, such as intra-week and seasonal variations, are also neglected.
- The low update frequency. Survey data are normally updated every 5–10 years, which limit the responsiveness of related urban policies in addressing the rapid metropolitan growth and socio-economic, demographic, infrastructure and travel behavior changes that may have occurred or are projected to occur in the foreseeable future.

During the last two decades, we have seen an explosion in the deployment of pervasive systems like cellular networks, GPS devices, and WiFi hotspots that allow us to capture massive amounts of real-time data related to people and cities (Reades et al., 2007; Gonzalez et al., 2008; Wang et al., 2009). The usage of these datasets could enable researchers to better understand the laws governing people's movements and improve the efficiency and responsiveness of urban policies. This study aims to explore the potential value and challenges of these novel datasets in urban modeling, using the mobile phone trace data collected by mobile network operators as an example.

Compared to travel survey data, the mobile phone trace data provide researchers new opportunities to examine individual mobility from an alternative perspective with their lower collection cost, larger sample size, higher update frequency, and broader spatial and temporal coverage. The mobile phone locations are routinely collected by operators for network management purposes, therefore, the datasets are theoretically available at no cost to researchers. The datasets allow for studying individual mobility of millions of people across the metropolitan area over a longer time period compared to a few thousand households' movements within 1–2 days usually collected through travel surveys. They are updated on a real time basis, which could lead to more reliable and trackable urban performance indicators and support more prompt policy responses to emerging urban issues.

Meanwhile, the mobile phone trace data also have significant drawbacks for transportation research: (1) socio-economic and demographic attributes are not available due to privacy concerns, which are indispensable to calibrate models at disaggregate level to explore the underlying behavior mechanism of individual/household mobility choice; (2) mobile phone users might not represent a random sample of the population. The results need careful analyses to be properly interpreted; and (3) the datasets are not primarily designed for modeling purposes and are often not in an easy-to-use format, which restricts the usefulness of raw data without intensive processing.

This study represents a first step towards building a methodology to utilize mobile phone data for transportation research. In this study, we use mobile phone traces from about one million users in the Boston Metropolitan Areas, Massachusetts, USA over 3 months to characterize individual mobility and understand its spatial patterns within a metropolitan area.

To address the lack of socio-economic and demographic attributes in mobile phone data, we aggregate mobility measures generated from individual mobile phone traces to block groups, the most disaggregate level of census geography at which socio-economic and demographic information is available, and associate the aggregate mobility measures with census data. Given the aggregate nature of our analysis, we raise two cautions at the outset. First, the underlying behavior mechanism cannot be identified by this study. The second caution concerns the ecological fallacy, which is the fallacy related to inferring the nature of individuals based solely upon aggregate statistics collected for the group. Therefore, it should be noted that what we find in this study is the general spatial patterns of mobility within the metropolitan area and their relationships to neighborhood characteristics, but not how individuals' characteristics and built environment influence their own travel behavior.

Nonetheless, using aggregate data collected for a long time period could help screen the idiosyncratic factors at the individual level, identify the underlying trends, and explain the variation in intra-urban mobility patterns. As Yang (2008) and Yang and Ferreira (2008) demonstrate, even without individual preference, urban spatial structure alone could explain a significant portion of the variation in commuting distance.

Similar analytic approaches that involve data aggregation have been adopted by many previous studies in mobility research. Lindsey et al. (2011) explore the relationship between Vehicle Kilometers Traveled (VKT) and urban form characteristics at grid cell level. Yang (2008) examines the relationship between excess commuting distance and urban spatial structure at census tract level. Holtzclaw et al. (2002) find that the average annual distance driven per car at the Traffic Analysis Zone (TAZ) level is a strong function of density, income, household size and public transit. Wang (2001) explains intra-urban variations of commuting time and distance at TAZ level in Columbus, Ohio using Census Transportation Planning Package (CTPP) data. Miller and Ibrahim (1998) examine the relationship between urban form and work trip VKT at traffic zone level in Toronto. These studies provide useful insights into individual mobility, but their data are mostly aggregated from travel surveys. Therefore they have similar data-related shortcomings, such as high data collection expense and low update frequency.