Transportation Research Part C 68 (2016) 285–299

Contents lists available at ScienceDirect

Transportation Research Part C

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

Review

The promises of big data and small data for travel behavior (aka human mobility) analysis

Cynthia Chen [a,*], Jingtao Ma [b], Yusak Susilo [c], Yu Liu [d], Menglin Wang [e]

[a] Department of Civil & Environmental Engineering, University of Washington, Seattle, USA
[b] Traffic Technology Services, Portland, USA
[c] Royal Institute of Technology, KTH, Sweden
[d] Institute of Remote Sensing and Geographical Information Systems, Peking University, China
[e] Cambridge Systematics, Chicago, USA

ARTICLE INFO

Article history:
Received 8 July 2015
Received in revised form 31 March 2016
Accepted 5 April 2016

Keywords:
Big data
Small data
Human mobility
Travel behavior
Transportation planning

ABSTRACT

The last decade has witnessed very active development in two broad, but separate fields, both involving understanding and modeling of how individuals move in time and space (hereafter called "travel behavior analysis" or "human mobility analysis"). One field comprises transportation researchers who have been working in the field for decades and the other involves new comers from a wide range of disciplines, but primarily computer scientists and physicists. Researchers in these two fields work with different datasets, apply different methodologies, and answer different but overlapping questions. It is our view that there is much, hidden synergy between the two fields that needs to be brought out. It is thus the purpose of this paper to introduce datasets, concepts, knowledge and methods used in these two fields, and most importantly raise cross-discipline ideas for conversations and collaborations between the two. It is our hope that this paper will stimulate many future cross-cutting studies that involve researchers from both fields.

Contents

* Corresponding author.
E-mail address: qzchen@uw.edu (C. Chen).

http://dx.doi.org/10.1016/j.trc.2016.04.005
0968-090X/

1. Introduction

Research in human movement in time and space has been around for at least over five decades (Weiner, 1999). Motivated by the need to forecast future travel demand to better guide the investment of often mega-scale transportation projects, transportation researchers have long sought to develop models to predict how people travel in time and space and seek to understand the factors that affect travel-related choices. Recently, grand challenges such as global warming and air pollution can all be traced to the over-reliance on automobiles, further motivating transportation researchers and practitioners to develop effective strategies to move toward more sustainable modes of transportation (e.g., public transit and walking and biking). For decades, transportation researchers have largely used data of active solicitation, including, for example, travel surveys where subjects are asked to self-report their activities and travels via paper, web, or phone interviews; travel surveys coupled with GPS loggers during which subjects are asked to both complete questionnaires and carry GPS loggers; and pure GPS-based surveys during which subjects are only asked to carry GPS loggers (Wolf et al., 2001; Axhausen et al., 2003; Hato et al., 2006; Stopher et al., 2008a, 2008b; Bohte and Maat, 2009; Chen et al., 2010; Gong et al., 2011). In the last type, information about subjects' activities and travels still need to be inferred from the collected GPS traces. All these surveys share a common characteristics and that is: active solicitation—subjects and information on their travels are actively recruited. Probably because of this attribute, these surveys are limited by a relatively small sample size (TMIP, 2013). In this paper, we refer to data of active solicitation as *small data*.

Parallel to the continued use of small data in transportation research, the rapid rise and prevalence of mobile technologies have enabled the collection of a massive amount of passive data (*big data*), which have resulted in a surge of studies on human movement (e.g., Gonzalez et al., 2008; Kang et al., 2012a, 2012b; Calabrese et al., 2013). Passive data refers to those data not collected through active solicitation; rather it is generated for purposes that are not intended but can be potentially used for research. Examples include mobile phone sightings generated by phone operators for operation purposes (Calabrese et al., 2011), social media data generated voluntarily by users' online activities (Chen and Schintler, forthcoming), and smart-card data collected at many transit systems worldwide (Pelletier et al., 2011; Ma et al., 2013). Passively collected, such data is very different from data of active solicitation (*small data*) that are familiar to most transportation researchers and thus requires different methods and techniques for processing and modeling. The first purpose of this paper is to introduce passively collected big data to transportation researchers, provide a state of the art review of the methods used, and identify areas of gap that are particularly important for transportation planners.

More importantly, this paper seeks to identify cross-disciplinary concepts and opportunities for both transportation researchers who have traditionally used small data and big data researchers. Our discussion will be on three important subareas of travel behavior research: (1) *behavioral factors* where the interest is identifying factors that explain travel behaviors and uncover the underlying causal mechanisms; (2) *modeling travel behavior* where models are developed to predict human movement behaviors; and (3) *human mobility patterns* where pattern recognition is an important goal. It is our view that the recent advances made with the use of the big data have the potential to drive fundamental advances in research in human mobility and at the same time, knowledge accumulated in transportation research in the past many decades can guide big data studies to answer questions that matter to the society, in particular, those relating to transportation investment decisions and policies in urban environments.

The rest of the paper is organized as follows. In Section 2, an introduction of the big data as well as a review of the current methodologies is provided. Our focus is on passively generated mobile phone dataset. In analyzing human mobility patterns, passively generated mobile phone data has emerged as the most frequently used (and possibly the most reliable) data source (Gonzalez et al., 2008). Other data sources cannot capture the full spectrum of an individual's mobility pattern over multiple days, involving the use of multiple modes of transportation. Examples include the use of taxi data that is mostly suitable for studying drivers' patterns in searching for passengers (Jiang et al., 2009; Liu et al., 2012), use of transit smart card data that only captures the use of transit modes (Long and Thill, 2015), or the use of social media data whose spatial and temporal resolutions are much lower than those of mobile phone data and biased toward certain locations (Cheng et al., 2011; Noulas et al., 2012). The target audience of Section 2 is transportation researchers who are familiar with the small, survey data but not the big data that has been recently utilized. In Sections 3–5, we discuss cross-disciplinary concepts and ideas in the three subareas noted earlier. Concluding discussions are provided in Section 6.