Editorial

Cynthia Chen¹ · Michael Batty² · Tom van Vuren³

Published online: 1 April 2015

© Springer Science+Business Media New York 2015

The rapid development of social media tools combined with location-aware sensors has made available the collection of passive datasets relevant to transportation modeling and planning. Passive data refers to those data not collected through active solicitation; it is data that is generated for purposes that are not originally intended but can potentially be used for research and practical transport planning applications. Examples of passive data include mobile phone sightings generated by phone operators for operations purposes and social media data generated voluntarily by users' online activities. Aside from the differences in the method of acquisition, passive data differs significantly from data collected through active solicitation in its coverage, length, and depth. Passive data often covers a substantial proportion of the population. In the case of human mobility patterns, the sampling rate of a typical household travel survey is about 1 % for a small metropolitan area and lower for larger areas, while the share of the population covered in typical passively generated mobile phone datasets is significantly higher and in some cases over 50 %. The length of the data for a passively generated data is often significantly longer (over multiple days to months to potentially years) compared to traditional household travel surveys that are often based on one-day diaries. Passive data, however, often lacks

✉ Cynthia Chen
qzchen@uw.edu

Michael Batty
m.batty@ucl.ac.uk

Tom van Vuren
Tom.vanVuren@mottmac.com

¹ Civil & Environmental Engineering, University of Washington, 201 More Hall,
Box 352700, Seattle, WA 98195-2700, USA

² Centre for Advanced Spatial Analysis, The Bartlett, UCL Faculty of the Built Environment, 90
Tottenham Court Road, London W1T 4TJ, UK

³ Mott MacDonald, 35 Newhall Street, Birmingham B3 3PU, UK

depth: unlike data collected through active solicitation that contains information designed for specific research purposes (e.g. socio-demographics and trip types), passive data often only contains limited data items (e.g. when and where a phone user is being sighted) while information critical for specific research purposes must be inferred or imputed by matching with other data sets. Furthermore, the limited information available in the passive data often contains a large amount of uncertainty.

The great promise of these emerging, passive data has in recent years generated a large number of studies that provide insights in diverse areas such as understanding human mobility patterns and their interactions with the environment and leveraging such data to improve the efficiency of the physical transportation system. Despite these exciting developments in many fields, important questions and critical challenges remain. This is particularly true for their professional use in transportation planning. The unique features associated with passive data are thus posing some critical research questions to be answered by the transportation planning community, including, for example:

- Can passive data be a potential substitute for data from active solicitation?
- What pre-processing and analysis methodologies are particularly suitable for such data?
- How do we validate the results inferred from passive data when there is no ground truth information?
- What are the applications in transportation planning for such passive data?
- What kind of competencies will transport researchers and practitioners need to be able to handle passive datasets effectively?

This special issue makes a very first attempt in answering these questions, with the hope of having many join the discussions and debates in the near future, as a wide variety of passive data is expected to sweep through the landscape of transportation planning. This special issue collects together nine papers, each providing some answers to one or more of the above-identified questions, but also posing further questions. Together, they provide a cross-sectional perspective and create synergies in our understanding of the challenges involved.

The paper by Advait Sarkar et al. (entitled “Comparing Cities’ Cycling Patterns Using Online Shared Bicycle Maps”) is an elegant example of how to carefully measure and mine publically accessible bike share data to identify intrinsic similarities from a set of 10 cities exhibiting heterogeneous behaviors. In one reviewer’s words, “it is an example of a journey being more important than the destination”. Via different perspectives and focuses, two papers demonstrate procedures for the validation of developed algorithms. This is a particularly important challenge resulting from the observation that passive data lacks depth and that ground-truth information is typically unavailable. The paper by Mahdiah Allahviranloo and Will Recker (entitled “Mining Activity Pattern Trajectories and Allocating Activities in the Network”) illustrates a procedure to generate pseudo GPS data from the California Household Travel Survey and the paper by Yijing Lu and Lei Zhang (entitled “Imputing Trip Purposes of Long-distance Travel”) develops a process of pulling existing data from multiple sources (American Travel Survey, land use data, and data from the Census Bureau) to create a simulated dataset for long-distance travel. These two studies also demonstrate the use of a variety of data mining techniques to impute trip purposes.

The special issue also contains papers using a variety of passive data to detect critical attributes for understanding human mobility patterns. Papers by Peter Widhalm et al. (entitled “Discovering Urban Activity Patterns in Cell Phone Data”), by Yang Xu et al.

(entitled “Understanding Aggregate Human Mobility Patterns using Passive Phone Location Data—A Home-based Approach”), and by Miguel Picornell et al. (entitled “Exploring the Potential of Phone Call Data to Characterize the Relationship between Social Network and Travel Behavior”) use mass mobile phone data to understand human mobility patterns and linkages between mobility patterns and social networks. Together, these papers demonstrate a variety of techniques to mine such data, involving a hierarchical clustering algorithm (by Xu et al.), a relational Markov Network (by Widhalm et al.) that cleverly leverages dependencies between activity types, scheduling and land use types, and a heuristic procedure (by Picornell et al.) that identifies frequently visited locations. The results of these papers offer great hope for potentially substituting household travel surveys with passively generated mobile phone data for understanding human mobility patterns. Recognizing that some passive data (e.g. mobile phone data) present challenges in terms of access, the paper by Zhenhua Chen and Laurie Schintler (entitled “Sensitivity of Location-Sharing Services Data: Evidence from American Travel Pattern”) examines the feasibility of mining travel patterns from publically-accessible location-sharing services (LSS) data. The findings point to the potential of using alternative types of passive data to understand travel patterns and show that the viability of using LSS data varies by types of environment. The paper by Neema Nassir et al. (entitled “Activity Detection and Transfer Identification for Public Transit Fare Card Data”) uses transit fare card data to detect travelers’ multi-leg journeys through the public transit system. Such studies not only offer great insights in understanding mobility patterns for transit users, but also reveal opportunities for a more efficient transit system that responds to demands.

The paper by Andrew Mondschein (entitled “Five-star Transportation: Using Online Activity Reviews to Examine Mode Choice to Non-work Destinations”) is an example of using passive data (in this case, Yelp reviews) for transportation planning applications. In this paper, the author shows how passive data can be mined to infer people’s mode choices and to understand accessibility to non-work destinations. The paper does not argue for a potential replacement by Yelp data of household travel surveys, yet it suggests that the Yelp data can potentially offer rich content and spatial information to augment a conventional household travel survey for transportation planning purposes.

The set of papers collected in this special issue illustrates the use of diverse data mining techniques, reveals insights in multiple modes of transportation and relates travel behaviors to other important aspects of contemporary social life (e.g. social interactions). By no means do they fully answer the questions raised earlier nor do they address many additional important questions that relate to passive data such as ways of linking such data where common keys are absent. One other issue relates to its representativeness, which is discussed in this special issue, but has not yet been tackled rigorously. As noted earlier, our hope is that this special issue will stimulate many more studies to answer questions that are intrinsically important to transportation planning.

Lastly, we want to express our sincere thanks to the many anonymous reviewers who have graciously provided their service to the production of this special issue, which was a substantial undertaking. For this special issue, we received 36 abstracts, from which 18 were invited for a full paper. Out of the 18 papers submitted, 9 papers were selected for this special issue.

Cynthia Chen is an Associate Professor of Civil and Environmental Engineering at the University of Washington, Seattle. Her research expertise and interests involve analyses of human mobility movements in time and space and the intersection of human and infrastructure systems in understanding sustainability and

resilience of cities and communities. She teaches classes in urban transportation planning, demand forecasting and resilience of infrastructure systems.

Michael Batty is the Bartlett Professor of Planning at University College London where he is Chair of the Centre for Advanced Spatial Analysis. His interests and those of his research group have become ever closer to the dynamics of cities in the very short term and the kinds of big data that are being collected passively involving mobility are more and more central to his work in understanding the structure of contemporary cities.

Tom van Vuren is a Divisional Director at Mott MacDonald, an international transport and engineering consultancy; he also holds a post as Visiting Professor at the University of Leeds. His interest in passive data stems from a desire to develop practical transport models better, cheaper and faster, based on an improved behavioural understanding, considerably lower data collection costs and an appropriate set of competencies in twenty first century transport modeling specialists.