

Contents lists available at [ScienceDirect](http://www.sciencedirect.com)

Transportation Research Part C

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

Potentials of using social media to infer the longitudinal travel behavior: A sequential model-based clustering method

Zhenhua Zhang^a, Qing He^{b,*}, Shanjiang Zhu^c^a Here Global B.V., 425 W Randolph St, Chicago, IL 60606, United States^b Department of Civil, Structural and Environmental Engineering, Department of Industrial and Systems Engineering, The State University of New York, Buffalo, NY 14260, United States^c Sid and Reva Dewberry Department of Civil, Environmental, and Infrastructure Engineering, George Mason University, Fairfax, VA 22030, United States

ARTICLE INFO

Keywords:

Longitudinal travel behavior
Social media
Travel survey
Clustering

ABSTRACT

This study explores the possibility of employing social media data to infer the longitudinal travel behavior. The geo-tagged social media data show some unique features including location-aggregated features, distance-separated features, and Gaussian distributed features. Compared to conventional household travel survey, social media data is less expensive, easier to obtain and the most importantly can monitor the individual's longitudinal travel behavior features over a much longer observation period. This paper proposes a sequential model-based clustering method to group the high-resolution Twitter locations and extract the Twitter displacements. Further, this study details the unique features of displacements extracted from Twitter including the demographics of Twitter user, as well as the advantages and limitations. The results are even compared with those from traditional household travel survey, showing promises in using displacement distribution, length, duration and start time to infer individual's travel behavior. On this basis, one can also see the potential of employing social media to infer longitudinal travel behavior, as well as a large quantity of short-distance Twitter displacements. The results will supplement the traditional travel survey and support travel behavior modeling in a metropolitan area.

1. Introduction

Traditional household travel survey can directly provide trip features and unveil regional traveler demographics. The travel survey is of crucial importance for transportation planners and policy makers because the detailed individual data in the travel survey cannot only explain current travel behavior but also to forecast future travel demand. Current survey methodologies vary in the form of telephone interview, email or mail, home visit, GPS device tracking, etc. and the process is usually well designed to guarantee accuracy. The respondents are also drawn scientifically from a cross-section of age, gender, household income, educational background, etc.

However, even a well-designed travel survey suffers from several well-known issues. First, the sample size per person is usually small, and the response rate is low. For instance, due to high costs for these methods, 2009 National Household Travel Survey (Santos et al., 2011) only covers about 150,000 people across the U.S. with both the national samples and the States and MPO add-on samples combined. Second, NHTS only provides one-day snapshot on the travel behavior of the sampled household and current household travel survey studies usually have some well-documented quality problems. Not all respondents are able to provide accurate trip

* Corresponding author.

E-mail addresses: zhenhua.zhang@here.com (Z. Zhang), qinghe@buffalo.edu (Q. He), szhu3@gmu.edu (S. Zhu).

<http://dx.doi.org/10.1016/j.trc.2017.10.005>

Received 12 February 2017; Received in revised form 3 October 2017; Accepted 3 October 2017

0968-090X/ © 2017 Elsevier Ltd. All rights reserved.

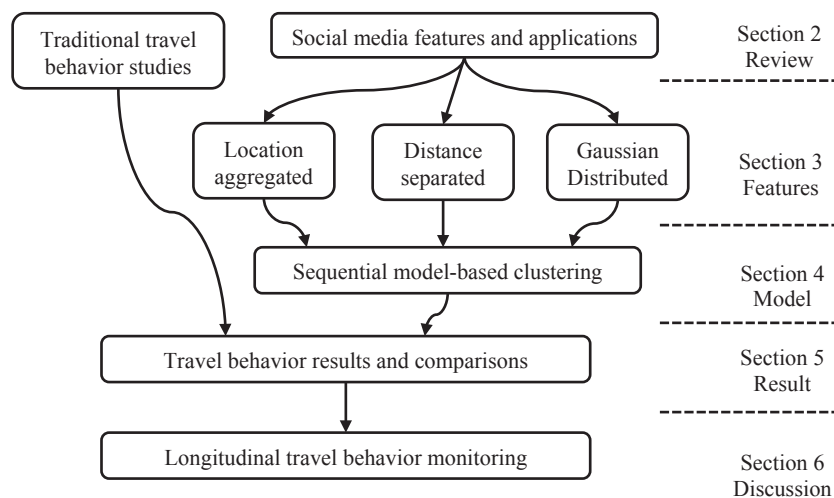


Fig. 1. Paper structure.

histories. For example, the respondents may not accurately recall their trip details, especially for non-recurrent and occasional short-distance trips. What's more, the travel survey is always costly, time-consuming, and labor-consuming and almost no organizations or institutes can afford an annual travel survey. For instance, the National Household Travel Survey after millennia is conducted only in 1969, 1977, 1983, 1990, 1995, 2001, 2009 (Santos et al., 2011) and 2016. Therefore, the conventional survey methods cannot provide up-to-date, continuous and longitudinal observations.

To mitigate above shortcomings and broaden the horizons of travel behavior studies, data crowdsourcing is a good option in complementing the traditional methods. Similar attempts can be seen in (Chen et al., 2016; Lin et al., 2015; Wang et al., 2016; Zheng et al., 2016). In this study, we introduce social media as a viable tool to infer the longitudinal travel behavior and activity patterns. The study area is the metropolitan area in Northern Virginia (NOVA) which has long been known for its heavy congestion. Its fluctuated traffic conditions and complex road networks make it a good testbed for studies such as traffic flow patterns (Zhang et al., 2016b), and travel time reliability (Zhang et al., 2016a). Our main contributions lie in the following three aspects. First, we unveil the characteristics of individual travel behavior especially the location aggregation, separation distances between locations and the clustering features of certain groups of locations. Second, based on these features, we propose a sequential model-based clustering method to capture the clustering features of traveler's hourly locations and extract directed travel between primary locations. Third, the results of travel behavior obtained from Twitter are compared with those from the household travel survey and advantages and disadvantages of Twitter travel behavior are discussed.

Our findings and modeling results are fully detailed in Fig. 1: Section 2 reviews the visions in the traditional travel survey and current efforts of social media in travel behavior studies. Based on the review, Section 3 examines the Twitter location data and finds three features; In Section 4, an algorithm of the sequential model-based clustering method is proposed which fully considered the data features of the Twitter locations and this algorithm can cluster the locations and eliminate the effects of undirected travel. Section 5 unveils the travel behavior features of the Twitter displacements, compared and validated by the household travel survey. Based on these Twitter displacement features, Section 6 identifies the potential of Twitter to conduct the longitudinal travel behavior studies. The paper ends with an elaborate conclusion and discussions in Section 7.

2. Literature review

2.1. Review of traditional travel behavior studies

Traditional travel behavior studies based on the travel survey or other methods are good references for our study. Most of the works are well founded after years of dedications, and their conclusions are insightful. They can not only unravel the interesting properties of the underlying mobility patterns (Schneider et al., 2013) but also have some practical meanings such as features of epidemics spreading (Anderson et al., 1992), population diffusion (Petrovskii and Morozov, 2009), and social networking (Centola, 2010).

One of the key elements in travel behavior studies is the movement. Past studies tend to clarify the movements by predefining the scales of zones (Balcan et al., 2009; Lenormand et al., 2014). The zone scale in each trip may vary from hundreds of meters to a few miles. Usually, movement between the zones is taken as directed travel while those within the zones is undirected travel (Schneider et al., 2013). The directed travel is the primal trip activity which holds the common view that human mobility is a motif-driven activity. These motifs are general human mobility characteristics that can account for the activities mostly. The primal trip activity is commonly seen and driven by purposes, such as traveling to a special event (Ni et al., 2014), commuting to the workplace, going to school, etc. In contrast, undirected travel represents the secondary trip activity (as this activity usually cannot be taken as a trip, we

Download English Version:

<https://daneshyari.com/en/article/4968393>

Download Persian Version:

<https://daneshyari.com/article/4968393>

[Daneshyari.com](https://daneshyari.com)