Applied Geography 70 (2016) 11–25

Contents lists available at ScienceDirect

Applied Geography

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

Explore spatiotemporal and demographic characteristics of human mobility via Twitter: A case study of Chicago

Feixiong Luo [a], Guofeng Cao [a,*], Kevin Mulligan [a], Xiang Li [b]

[a] Department of Geosciences, Texas Tech University, USA
[b] Key Lab of Geographical Information Science, East China Normal University, USA

ARTICLE INFO

Article history:
Received 5 August 2015
Received in revised form
27 February 2016
Accepted 2 March 2016

Keywords:
Human mobility
Geodemography
Social media
Twitter
Name analysis
Chicago

ABSTRACT

Characterizing human mobility patterns is essential for understanding human behaviors and the interactions with socioeconomic and natural environment, and plays a critical role in public health, urban planning, transportation engineering and related fields. With the widespread of location-aware mobile devices and continuing advancement of Web 2.0 technologies, location-based social media (LBSM) have been gaining widespread popularity in the past few years. With an access to locations of hundreds of million users, profiles and the contents of the social media posts, the LBSM data provided a novel modality of data source for human mobility study. By exploiting the explicit location footprints and mining the latent demographic information implied in the LBSM data, the purpose of this paper is to investigate the spatiotemporal characteristics of human mobility with a particular focus on the impact of demography. To serve this purpose, we first collect geo-tagged Twitter feeds posted in the conterminous United States area, and organize the collection of feeds using the concept of space-time trajectory corresponding to each Twitter user. Commonly human mobility measures, including detected home and activity centers, are derived for each user trajectory. We then select a subset of Twitter users that have detected home locations in the city of Chicago as a case study, and apply name analysis to the names provided in user profiles to learn the implicit demographic information of Twitter users, including race/ethnicity, gender and age. Finally we explore the spatiotemporal distribution and mobility characteristics of Chicago Twitter users, and investigate the demographic impact by comparing the differences across three demographic dimensions (race/ethnicity, gender and age). We found that, although the human mobility measures of different demographic groups generally follow the generic laws (e.g., power law distribution), the demographic information, particular the race/ethnicity group, significantly affects the urban human mobility patterns.

1. Introduction

Human mobility plays an important role in the study of traffic forecasting (Kitamura, Chen, Pendyala, & Narayanan, 2000), disease spreading (Colizza, Barrat, Barthelemy, Valleron, & Vespignani, 2007), urban planning (Horner & O'Kelly, 2001), and in the general science and engineering of smart cities (Batty, 2012). Conventionally, human mobility research heavily relies on traditional surveys of travel journals as major data sources. For example, based on 30 travel surveys in more than 10 countries, Schafer (2000) has shown that human travels exhibit strong regularities across space and time. The process of conventional survey is typically long, expensive, and often limited to a small number of samples, and hence difficult to realistically deal with the complex human behaviors. The advancement of Web and mobile technologies yields a set of reliable and cost-effective data sources that can provide fine-granularity of spatiotemporal information for large scale human behaviors and social dynamics. Typical examples of these data sources include on-line bank note tracking logs (Brockmann, Hufnagel, & Geisel, 2006), mobile phone calling records (Gonzalez, Hidalgo, & Barabasi, 2008), vehicle GPS trajectories (Yuan, Zheng, & Xie, 2012), and transactions of bank (or credit) cards (Lenormand et al., 2015; Sobolevsky et al., 2014) and transportation cards (Hasan, Schneider, Ukkusuri, & González, 2013). By taking advantage of these data sources, numerous findings, such as

* Corresponding author.
E-mail addresses: feixiong.luo@ttu.edu (F. Luo), guofeng.cao@ttu.edu (G. Cao), kevin.mulligan@ttu.edu (K. Mulligan), xli@geo.ecnu.edu.cn (X. Li).

http://dx.doi.org/10.1016/j.apgeog.2016.03.001
0143-6228/

spatiotemporal regularity and heavy-tail distribution of travel distances, have been recently reported (Barabasi, 2005; Yan, Han, Zhou, & Wang, 2011).

Most recently, social media (e.g., Twitter and Facebook), a set of on-line applications that allows users to create and exchange contents, have been experiencing a spectacular rise in popularity and attracting hundreds of millions of users for social networking, content generating and sharing. With the wide adoption of location-aware mobile smart devices and wireless communications, people tend to access social media from mobile smart devices, and thanks to the in-built positioning capability, location and whereabouts information can be attached to the social media messages (e.g., geo-tagged Twitter posts, Flickr photos and check-ins). The location-based social media (LBSM) data provide access to the locations as well as the contents of the social media activities, hence provide a promising modality of data source for studying the complex human behaviors and understanding socioeconomic dynamics (Liu et al., 2015). By capitalizing upon this new data source, novel and successful applications have started to emerge. In geography, location-based social media have been used to spatio-temporally and schematically characterize the geographic places (McKenzie & Janowicz, 2015; McKenzie, Janowicz, Gao, & Gong, 2015). Specifically in human mobility research, by analyzing 22 million check-in records in multiple location sharing services, Cheng, Caverlee, Lee, and Sui (2011) demonstrated that human mobility is a mixture of short, and random movements with occasional long jumps. Cho, Myers, and Leskovec (2011) found that social network relationships can explain approximately 10% and 30% of human movements and 50%–70% periodic behaviors based on check-in records and friendship networks from LBSM sites Gowalla and Brightkite. Hasan et al. (2013) explored an archive of Foursquare and Twitter posts, and reported that users tend to visit different urban places with diminishing regularity governed by a Zipf's law. Based on a collection of almost a billion tweets recorded in the year of 2012, Hawelka et al. (2014) demonstrated the characteristics of international travels and of human mobility across different countries. Wu, Zhi, Sui, and Liu (2014) combined activity-based analysis with a movement-based approach to model the intra-urban human mobility observed from about 15 million social media check-in records in China.

Most of the above mentioned research focuses on the spatial and temporal aspects of the human mobility and tends to ignore the effects of demographic and other (e.g., socioeconomic and health status) background information that have been known as significant factors in population distribution and human mobility. Exceptions include Cheng et al. (2011) and Li, Goodchild, and Xu (2013), which linked LBSM activities with associated geographic regions (e.g., via spatial join) to explore differences of LBSM activities as local geographic and economic landscapes vary at aggregated scales. Compared with the previously mentioned data sources (e.g., cell phone calling logs, bank notes) that have only access to the user locations, LBSM data also offer access to the profiles (e.g., profile names) of individual social media users and contents of social media messages. These information, if analyzed appropriately, can provide important background (e.g., demographic information and health status) of the social media users. These learned background information, together with location footprints, makes it possible to investigate the spatiotemporal and demographic characteristics of human mobility at a fine individual scale.

Personal names (both first and last names) have been demonstrated to carry rich amount of demography (Mateos, 2007), geography (King & Jobling, 2009), cultural (Zelinsky, 1970), linguistic and even genetic information (King & Jobling, 2009) about the name bearers, and name analysis has been used to study a range of problems in different disciplinary, e.g., ethnicity and population structure (Mateos & Longley, 2011), human migration (Piazza, Rendine, Zei, Moroni, & Cavalli-Sforza, 1986). Twitter requires new users to provide a full name (a pair of forename and surnames). Although real full names are not enforced, a significant fraction of active social media users are willing to provide real names (Peddinti, Ross, & Cappos, 2014), which have been successfully used to help demography breakdowns of social media users (e.g. Chen, Gallagher, & Girod, 2013; Gallagher & Chen, 2008; Oktay, Firat, & Ertem, 2014). In this paper, we apply a similar name analysis to the names provided in Twitter profiles to detect demographic groups, and then investigate the impact of the demographic background on spatiotemporal distribution and human mobility of Twitter users. Specifically, based on a public data stream of Twitter feeds posted in the conterminous United States area, we first represent the location footprints of each Twitter user in terms of a space-time trajectory, and for each of the trajectories, commonly used human mobility measures, including detected home locations, activity centers and radius of gyration, are computed and maintained. A subset of Twitter users whose detected home locations in the city boundary of Chicago are chosen for a case study. Name analysis is applied to the Chicago Twitter users to detect the implicit groups of three demographic factors, including race/ethnicity, age and gender. We then investigate the spatiotemporal distribution and mobility characteristics of Chicago Twitter users and compare the differences across the detected groups of the three demographic factors.

The remainder of this paper is structured as follows. Section 2 introduces the overall work flow of this paper including the methodology for detecting demographic information and activity centers of Twitter users. Section 3 implements the methodology, analyzes the spatiotemporal and demographic characteristics of human mobility of Chicago Twitter users. Section 4 concludes the paper and discusses the limitations and future work.

2. Methodology

Fig. 1 gives the overall flowchart of the proposed analysis of this paper. Five consecutive steps (components) are included: (1) Twitter feed collection, (2) space-time trajectories construction, (3) local residents identification, (4) name analysis for demographic groups, and finally (5) investigating spatiotemporal and demographic characteristics of human mobility of Chicago Twitter users.

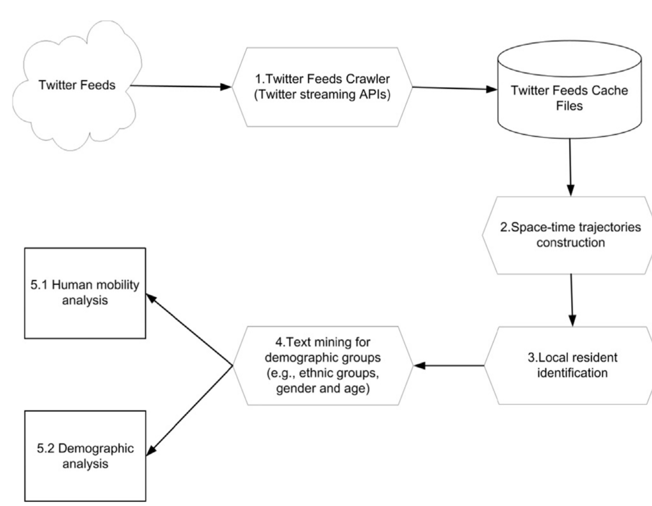

Fig. 1. The flowchart of the proposed method.