Physica A 420 (2015) 134–147

Contents lists available at ScienceDirect

Physica A

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

A comparative analysis of intra-city human mobility by taxi

Wenjun Wang [a,b], Lin Pan [a,b,*], Ning Yuan [a,b,c], Sen Zhang [d], Dong Liu [a,b,e]

[a] School of Computer Science and Technology, Tianjin University, Tianjin 300072, China
[b] Tianjin Key Laboratory of Cognitive Computing and Application, Tianjin 300072, China
[c] Department of Basic Courses, Academy of Military Transportation, PLA, Tianjin 300161, China
[d] Department of Math, Computer Science and Statistics, State University of New York College at Oneonta, Oneonta, NY 13820, USA
[e] School of Computer and Information Technology, Henan Normal University, Xinxiang 453007, China

HIGHLIGHTS

- We observe common patterns of human mobility by taxi in several cities.
- The displacement distributions of taxi trips tend to follow exponential laws.
- The trip durations follow log-normal distributions.
- The interevent time distributions have log-normal bodies followed by power law tails.
- Airports attract large amounts of taxi traffic at a certain distance level.

ARTICLE INFO

Article history:
Received 8 April 2014
Received in revised form 26 October 2014
Available online 6 November 2014

Keywords:
Human mobility
Urban traffic
Taxi trajectory
Exponential distribution
Log-normal distribution

ABSTRACT

Quantitative understanding of human movement behaviors would provide helpful insights into the mechanisms of many socioeconomic phenomena. In this paper, we investigate human mobility patterns through analyzing taxi-trace datasets collected from five metropolitan cities in two countries. We focus on three statistics for each dataset: the displacement of each occupied trip, the duration of each occupied trip, and the time interval between successive occupied trips by the same taxi (interevent time). The results indicate that the displacement distributions of human travel by taxi tend to follow exponential laws in two displacement ranges rather than power laws; the trip duration distributions can be approximated by log-normal distributions; the interevent time distributions can be well characterized by log-normal bodies followed by power law tails. For each considered measure, the rescaled distributions of all cities collapsed into a master curve. These results provide empirical evidence supporting the common regularity of intra-city human mobility. Moreover, we show that airport locations could play a role in explaining the spikes of displacement distributions of taxi trips in certain cities.

1. Introduction

Understanding, modeling and predicting human spatio-temporal movements is an active research subject relevant to urban studies [1], traffic engineering [2], urban planning [3], smart cities [4], epidemics controlling [5,6], emergency management [7,8], location-based services [9], and so on. Traditionally, researchers relied on using and analyzing the human movements data collected from performing travel surveys or observations [10,11]. More recently, researchers have been able

* Corresponding author at: School of Computer Science and Technology, Tianjin University, Tianjin 300072, China.
E-mail address: linpan@tju.edu.cn (L. Pan).

http://dx.doi.org/10.1016/j.physa.2014.10.085
0378-4371/

to have access to massive individual movement data, including GPS traces of vehicles [12–15] and human beings [16,17], mobile phone records [18–20] and check-ins of online social network accounts [9,21]. These data have helped researchers to uncover the patterns of human mobility from more perspectives and gain deeper understanding of their underlying mechanisms than before.

A number of recent studies have found that statistical patterns of human mobility exhibit characteristics similar to Lévy flight [16–18,22], a type of random walk with the lengths of the steps following a certain probability distribution and the directions of the steps being isotropic and random. To be more specific, power law distribution was found in the dispersal of bank notes [22]. A power law with an exponential cutoff can be used to approximate the displacement distribution of human trajectories obtained from mobile phone datasets [18,20], GPS traces [16,17,23] and online location-based social networks [9]. Furthermore, Song et al. found a 93% potential predictability in individual's daily mobility [24]. In order to explain the observed scaling laws, González et al. suggested a convolution of the population-based heterogeneity and individual Lévy flights [18]. Song et al. developed an individual mobility model which incorporates exploration and preferential return principles [25]. Han et al. demonstrated that the power-law-like displacement distributions may originate from the hierarchical organization of transportation systems [26]. Yan et al. explained the aggregated scaling law under the principle of maximum entropy [27].

The studies mentioned above have provided valuable insights into the patterns and mechanisms of human mobility. However, these work studied human movements at various scales. Since most daily travel activities are bounded by geographic borders [28], routine movements of individuals within cities have also been of a particular interest to the research community. Through the analysis of GPS traces collected from 50 taxicabs, Jiang et al. found scaling properties in trip lengths [12]. However, power law tails were not suggested in many cities by analyzing the check-ins of Foursquare users [21]. Daily mobility lengths by private vehicles [13], taxis [14,15] and mobile phones [19] were reported to follow exponential distributions rather than power laws. Using a taxi dataset in Lisbon, Veloso et al. showed that trip distances can be fitted with a Gamma distribution, while the decreasing interval of trip distances can be fitted with an exponential distribution [29]. A study using a mobile phone dataset in Portugal showed that the commuting distances fit the log-normal distribution [30]. In short, according to various datasets compiled from different cities, many empirical studies have demonstrated that the displacements occurred in the cities deviate from power law distributions.

In this paper, we investigate statistical laws for human mobility by exploring large amounts of GPS traces collected from the taxi transportation system in five metropolitan cities. We focus on discovering common patterns of collective human mobility by taxi at the large city scale. More precisely, we analyze the statistical distributions of trip displacements (i.e., the displacement between pick-up and drop-off locations of an occupied trip), trip durations (i.e., the travel time of an occupied trip) and interevent time (i.e., the time spent between a pick-up and the previous drop-off). The main contributions of this paper can be summarized as follows:

- We comparatively study human mobility patterns in five metropolitan cities by analyzing massive trajectories of GPS equipped taxis.
- We observe that the statistical properties of taxi trips in all the five studied cities display similar behaviors. For each property, the distributions of all datasets collapse after being rescaled, suggesting common patterns of human mobility by taxi in cities.
- We find that the displacements of human movements by taxi approximately follow exponential distributions at the small distance scale (less than the diameter of the main urban area), and the displacements at the large distance scale tend to follow the exponential distributions rather than the power laws.
- We note that travel costs and airport locations should have important effects on human activities of traveling by taxi.
- We show that the trip durations could be well described by a log-normal distribution.
- We observe similar traffic impact in different cities by exploring the correlations between trip displacements and trip durations.
- We also observe log-normal bodies in the distributions of the interevent time between two consecutive taxi services. The tails of the interevent time distributions display similar scaling properties with exponents ranging from 1.52 to 1.60.

2. Datasets

This research uses six taxi datasets containing a total of more than 18 billion GPS records, which were generated by about a total number of 30,000 taxis from four major cities (Beijing, Tianjin, Shanghai and Nanjing) in China and the San Francisco Bay Area in the United States.

Dataset D_1 contained the GPS traces of approximately 15% of taxis in Beijing during March 2009. Both D_2 and D_3 were generated by over 4000 taxis in Tianjin, which are about 12% of the taxis in Tianjin, but respectively, from August 1, 2011 to December 31, 2011 (D_2) and from January 1, 2012 to December 31, 2012 (D_3). D_4 [31,32] was collected from over 4000 taxis in Shanghai, where there were nearly 50,000 taxis. D_5 consisted of the GPS traces recorded over a two-year period for almost all the taxis in Nanjing. D_6 [33,34] was collected through the cabspotting project [35], and it contained taxicab mobility traces that were captured in the San Francisco Bay Area from May 17 to June 10, 2008. Some basic statistics of the six datasets are summarized in Table 1.