Physica A 438 (2015) 140–153

Contents lists available at ScienceDirect

Physica A

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

Uncovering urban human mobility from large scale taxi GPS data

Jinjun Tang[a,b,*], Fang Liu[c], Yinhai Wang[b], Hua Wang[a]

[a] School of Transportation Science and Engineering, Harbin Institute of Technology, Harbin 150001, China

[b] Department of Civil and Environmental Engineering, University of Washington, Seattle, WA 98195-2700, USA

[c] School of Energy and Transportation Engineering, Inner Mongolia Agricultural University, Hohhot 010018, China

HIGHLIGHTS

- We use taxi GPS data to analyze travel demand distributions.
- DBSCAN algorithm is used to cluster pick-up and drop-off locations.
- Spatial interaction models are calibrated and compared to study searching behavior.
- Travel distance, time and average speed are utilized to explore human mobility.
- We construct an entropy-maximizing model to estimate the traffic distribution.

ARTICLE INFO

Article history:
Received 19 October 2014
Received in revised form 2 May 2015
Available online 8 July 2015

Keywords:
Human mobility
Taxi GPS data
Travel distance and time
Trips distribution modeling

ABSTRACT

Taxi GPS trajectories data contain massive spatial and temporal information of urban human activity and mobility. Taking taxi as mobile sensors, the information derived from taxi trips benefits the city and transportation planning. The original data used in study are collected from more than 1100 taxi drivers in Harbin city. We firstly divide the city area into 400 different transportation districts and analyze the origin and destination distribution in urban area on weekday and weekend. The Density-Based Spatial Clustering of Applications with Noise (DBSCAN) algorithm is used to cluster pick-up and drop-off locations. Furthermore, four spatial interaction models are calibrated and compared based on trajectories in shopping center of Harbin city to study the pick-up location searching behavior. By extracting taxi trips from GPS data, travel distance, time and average speed in occupied and non-occupied status are then used to investigate human mobility. Finally, we use observed OD matrix of center area in Harbin city to model the traffic distribution patterns based on entropy-maximizing method, and the estimation performance verify its effectiveness in case study.

1. Introduction

Human travel behaviors are affected by a number of factors such as spatial structure of city, land use and road networks, understanding the regularity and characteristics of human mobility is of major importance to city and transportation planning. As questionnaire based approach is constrained by lack of data, it is difficult to use this traditional method to explore

* Corresponding author at: School of Transportation Science and Engineering, Harbin Institute of Technology, Harbin 150001, China.
E-mail address: jinjuntang@163.com (J. Tang).

http://dx.doi.org/10.1016/j.physa.2015.06.032

human mobility deeply and accurately. The fast development of information and communication technology makes it possible to understand travel behaviors of people by providing large-scale and granular data recording individual information chronologically. Various dataset including wireless network traces [1], GPS traces from probe vehicle data [2–5], mobile phone [6–15] and banking notes [16] are collected to study spatial–temporal feature of human movement. Jiang et al. [2] analyzed the human mobility pattern from over 72 000 people's moving trajectories collected from 50 taxicabs during six-month. Zheng et al. [3] proposed a graph-based post-processing algorithm to infer human movement modes from GPS trajectories data. In order to analyze the accessibility, Li et al. [5] introduced a new dynamic accessibility measure based on real-time travel speed extracted from probe vehicle data. Csáji et al. [6] used principal component analysis to reveal the relation between features of human behavior and their geographical location from mobile phone dataset. Sun et al. [7] also applied principal component analysis to discover the urban dynamics based on cell phones location information. Kang et al. [8] presented the distribution of human urban travel followed the exponential law, in which the exponents were affected by city size and shape, and they used Monte Carlo simulation to verify the relation between intra-urban human mobility and urban. By constructing mobile phone network, Hidalgo et al. [9] defined the persistence of ties to explore dynamics of human mobility. González et al. [10] used the trajectories from 100,000 mobile phone users during six months to show a highly regulated human mobility pattern. Calabrese et al. [11] presented a method to extract mobility information from mobile phone traces and established a multivariate regression model to predict human mobility. Song et al. [12,13] proposed novel models to explore human mobility patterns.

As a main part of public transportation system in cities, taxi undertakes massive citizens' travel for its accessibility and flexibility. Furthermore, taking GPS-equipped taxis as probe vehicles, these mobile sensors provide us new tools to discover spatial–temporal patterns of people movement and even origins and destinations distribution. Thus, compare to data source from cell phones, taxi locations data can reflect traveling characteristics more precisely as passengers who pick up taxis have certain origins and destinations. Recently, lots of interesting works focus on human activity recognizing, hotspot discovering, urban planning and transportation planning [17–28].

Liu et al. [17,18] introduced a new method to explore intra-urban human mobility and land use variations based on taxi trajectory data from Shanghai city. Castro et al. [19] proposed an overview of mechanisms for using taxi GPS data to analyze people's movements and activities, which includes three main categories: social dynamics, traffic dynamics and operational dynamics. Liang et al. [20] found the taxis' traveling displacements and elapsed time follow an exponential distribution instead of a power-law. In Refs. [21–23], Veloso and Phithakkitnukoon used taxi data collected in Lisbon city to study urban mobility, spatiotemporal variation of taxi services, relationships between pick-up and drop-off locations and drivers' behaviors. Zhu and Guo [24] proposed a hierarchical method to deal with the problem of how to extract clusters from similar flows in taxi trips. Liu et al. [25] used a two-level hierarchical polycentric city structure to study spatial interaction perspective in Shanghai city with large scale taxi data. Wu et al. [26] introduced a novel method to explore urban human mobility based on social media check-in data, in which they constructed transition probability to model travel demand distribution. Liu et al. [27] analyzed taxi drivers' spatial selection behavior, spatio-temporal operation behavior, route choice behavior, and operation tactics with taxi GPS traces. Pan et al. [28] discussed a new method by using taxi traces to classify the urban land-use features.

In this paper, we use taxi GPS data collected from more than 1100 drivers in Harbin city to characterize people travel movement. The distribution patterns of origins and destinations on weekday and weekend are firstly analyzed. Then, travel distance, time and speed are used to explore human mobility by extracting taxi trips from GPS trace data. Finally, we verify the effectiveness of entropy-maximizing method for modeling trip distribution.

The rest of this paper is organized as follows. Section 2 introduces the dataset used in paper, travel demand distributions and spatial interaction models. Section 3 describes the mobility pattern of taxi. Results of network-based method are discussed in Section 4. Conclusion is provided in final section.

2. Transportation demand analysis and attractiveness modeling

2.1. Data source

The taxi GPS data we used in this study are collected from about 1100 drivers in Harbin city, which locates in the northeast of China. The data start from July to December in 2012, the recording rate is 30 s, and total samples come to 2880 a day. Each data sample contains not only location information but also collecting time and status. Table 1 provides an overall description of taxi trajectory data. The "Time" indicates when the data be recorded, "Latitude" and "Longitude" provide location data of taxi vehicle, "Speed" is the instantaneous velocity of vehicle, the unit is kilometers per hour. "Orientation" represents driving direction, which is based on North. "Status" represents the taxi whether be occupied by passengers, "0" represents the taxi vehicle is vacant and "1" means it is occupied.

2.2. Distribution pattern of demand

We classify taxi trips into two parts based on their status: (1) pick up passengers from origins to destinations. (2) Roam on the road to find next passenger. The overall distributions of origins and destinations reflect the travel demand of citizens