Journal of Transport Geography 61 (2017) 72–86

Contents lists available at ScienceDirect

Journal of Transport Geography

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

Revealing intra-urban travel patterns and service ranges from taxi trajectories

Shen Zhang[a], Jinjun Tang[b,*], Haixiao Wang[c], Yinhai Wang[d], Shi An[a]

[a] School of Transportation Science and Engineering, Harbin Institute of Technology, No. 73, Huang-He Street, 150090 Harbin, China
[b] School of Traffic and Transportation Engineering, Central South University, Changsha 410075, China
[c] School of Energy and Transportation Engineering, Inner Mongolia Agricultural University, Hohhot 010018, China
[d] Department of Civil & Environmental Engineering, University of Washington, Seattle, WA 98195-2700, United States

ARTICLE INFO

Keywords:
Taxi trajectory
Spatial point pattern
Utilization distribution
Movement-based kernel density estimation

ABSTRACT

As an important transport tool, taxi plays a significant role to meet travel demand in urban city. Understanding the travel patterns of taxis is important for addressing many urban sustainability challenges. Previous research has primarily focused on examining the statistical properties of taxi trips to characterize travel patterns, while it may be more appropriate to explore taxi service strategies on seasonal, weekly or daily time scale. Therefore, intra-urban taxi mobility is investigated by examining taxi trajectory data that were collected in Harbin, China, while 12-week corresponding to 12-month is chosen as the sampling period in our study. The multivariate spatial point pattern analysis is firstly adopted to characterize and model the spatial dependence, and infer significant positive spatial relationships between the picked up points (PUPs) and the dropped off points (DOPs). Secondly, the points of interest (POIs) are identified from DOPs using the emerging hot spot detection technique, then the taxi services and movement patterns surrounding POIs are further examined in details. Moreover, our study builds on and extends the existing work to examine the statistical regularities of trip distances, and we also validate and quantify the impacts posed by airport trips on the distance distributions. Finally, the movement-based kernel density estimation (MKDE) method is proposed to estimate taxis' service ranges within three isopleth levels (50, 75 and 95%) between summer/weekday and winter/weekend from taxi driver's perspective, and season as well as temperature factors are identified as the significant effect within certain service range levels. These results are expected to enhance current urban mobility research and suggest some interesting avenues for future research.

1. Introduction

As an important component of the public transportation sector, taxis provide flexible and convenient mobility solutions for urban residents. Locating the taxi-hailing demand of urban residents and optimizing the resource allocation of taxi service is critical to improving the quality of public transit services. However, the limitations of travel survey data reduce the quality of traditional travel demand studies and can result in misleading conclusions. Therefore, alternative data sources are needed to more accurately and comprehensively understand the spatial-temporal characteristics of travel demand.

The widespread adoptions of location-based services (LBS) provide unprecedented opportunities to study various mobility patterns from trillions of trails and footprints (Liang et al., 2012). A careful analysis of these digital footprints from taxi GPS localizers can provide an innovative strategy to facilitating urban public transit planning and operational decision-making, and thus have attracted considerable attentions from researchers.

Jiang et al. (2009) verified the scaling properties of taxi trip length and suggested that such property is attributed to the underlying street topology. Liu et al. (2012a, 2012b) 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. (2013) 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. (2012) found the taxis' traveling displacements and elapsed time follow an exponential distribution instead of a power-law. Veloso et al. (2011a, 2011b) used taxi data collected in Lisbon city to study urban mobility, spatiotemporal variation of taxi services, relationships be-

* Corresponding author.
E-mail addresses: shenzhang@hit.edu.cn (S. Zhang), jinjuntang@csu.edu.cn (J. Tang), teacherwhx@126.com (H. Wang), yinhai@u.washington.edu (Y. Wang), anshi@hit.edu.cn (S. An).

http://dx.doi.org/10.1016/j.jtrangeo.2017.04.009
Received 14 November 2016; Received in revised form 18 April 2017; Accepted 20 April 2017
Available online 05 May 2017
0966-6923/

tween pick-up and drop-off locations and drivers' behaviors. Wang et al. (2015) investigate statistical distributions of trip displacements, trip durations and interevent time by exploring large amounts of GPS traces collected in five metropolitan cities. Zhu and Guo (2014) proposed a hierarchical method to deal with the problem of how to extract clusters from similar flows in taxi trips. Liu et al. (2015) used a two-level hierarchical polycentric city structure to study spatial interaction perspective in Shanghai city with large scale taxi data. Wu et al. (2014) 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. (2010) analyzed taxi drivers' spatial selection behavior, spatio-temporal operation behavior, route choice behavior, and operation tactics with taxi GPS traces. Pan et al. (2013) worked on urban land-use classification using taxi GPS traces, and found pick-up/set-down dynamics exhibited clear patterns corresponding to the land-use classes of these regions. Tang et al. (2015, 2016a, 2016b) analyzed urban travel mobility, OD distribution, hotspots identification in Harbin city using large-scale taxi GPS trajectories. Since taxi travel pattern may also relate to vacant search behavior, Wong et al. (2014a, 2014b) further proposed a cell-based model to predict local customer-search movements of vacant taxi drivers.

These studies mentioned above have provided valuable insights into the patterns and mechanisms of taxi mobility, they tend to combine data on the location of origins and destinations, trip purpose, trip length, trip duration, departure and arrival times and travel modes in their analyses. Moreover, it is worth noting that the aforementioned studies generally focused on understanding travelers' behaviors on ordinary weekdays. Until recently, few studies have examined the variations of travel patterns between seasons, while the spatial-temporal properties of seasonal transport trips still tend to be regular and periodic. As a result, a purpose of this study is to model the travel patterns of taxi passengers on seasonal, weekly or daily time scale, and assess potential climate change effects on travel behavior. Additionally, most current studies merely focus on partial statistical properties of taxi trips (i.e., the trip displacements and durations). However, trip is not the only representative measure of taxi's travel pattern, while it may be more appropriate to estimate service coverage on an individual level from the perspective of public transport. Therefore, this research also aims to fill this gap by measuring taxis' service range sizes referred to home range studies (Braham et al., 2015; Dürr and Ward, 2014). The service range of a taxi is described in terms of a probabilistic model in our study, while the calculation of range size is required from the estimation of the utilization distribution (UD: the name given to the probability distribution of individual spatial location). Therefore, the service range model not only provides a useful global representation of space coverage patterns of taxis, but also represents the probability of taxi-hailing opportunities for passengers in a defined area, which is of great importance in public transport service studies. Such results can also help guide traffic modeling, public transportation planning, urban planning, and infrastructure development.

Our on-going work is focused on the analysis of taxi-GPS traces acquired in the city of Harbin, China, to better understand urban mobility from both driver's and passenger's perspectives. The contribution of this work lies on the following aspects: we investigate the multivariate spatial point pattern between pick-up and drop-off locations, analyze the local mobility patterns around POIs, examine the statistical distributions of trip distances, and explore the possibility of estimating the service range and identifying the seasonal effects from spatial area perspective.

2. Materials and methods

2.1. Study area

The case study is carried out in Harbin city, which locates in the northeast of China. The study area is restricted to Harbin urban area (126°57′–126°73 N lat and 45°68′–45°80E long) compared to its administrative area, which is rather large with latitude spanning 44° 04′–46° 40′ N, and longitude 125° 42′–130° 10′ E.

The area is characterized by a sub-humid continental monsoon climate with four distinct seasons, including a cold and dry winter (December–February) and a hot and rainy summer (June–August). The annual mean temperature is + 4.25 °C (39.6 °F), and extreme temperatures have ranged from − 42.6 °C (− 45 °F) to 39.2 °C (103 °F).

2.2. Data collection and study design

We conduct the study for the whole year of 2014. Since there are over 10 million GPS records in one day, 12-week corresponding to 12-month is chosen as the sampling period. Data is collected from 1000 distinct taxis, which account for nearly 12.5% of taxis in Harbin area. The sampling rate is 30 s, and total samples come to 2880 a day. Each data sample contains information of taxi id, timestamp, latitude, longitude, instantaneous velocity, and taximeter state (vacant or occupied). A data cleaning process is applied, removing trips with < 200 m and > 60 km (the realistic longest trips from one side of the city to the other could be around 55 km).

For illustration purposes, we have mapped the distribution of all sample taxi records in one day (Fig. 1) using open street map. Fig. 1 clearly shows that the majority of sample records are in the inner city area. Then, 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. Fig. 1a demonstrates a one-day trajectory of a taxi, where red dotted lines denote the occupied trip paths, and green dotted lines indicate the unoccupied trips. The spatial overlapping parts are caused by temporal factors. Based on the data, we can also identify the locations where passengers were picked up and dropped off (Fig. 1b and c), and thus the origin and destination of a completed occupied trip. PostgreSQL 9.4 and its spatial extension PostGIS are proposed for managing GPS data sets in our study, while ArcGIS 10.3 is adopted as the GIS interface for manipulating and visualizing the PostGIS data. Besides, the core of the analytical inspection and graphics are fully executed in the R 3.2.5 environment.

2.3. Trips classification

Taxi trip is a very important part of human beings movements in urban areas. In this section, three parameters including travel distance, time and average speed are used to explore taxi mobility. As we mentioned in above section, taxi drivers always exhibit different driving behaviors at different status: load up passengers and vacant. Thus, the trips can be classified into two parts. Dataset for occupied taxi k at time period τ can be expressed as: $R^o = (k, l^o, \tau^o)$, in which $l^o = (x^o, y^o)$ denotes utm32 geometry coordinates in the specific Universal Transverse Mercator (UTM) zone. Similarly, the dataset of non-occupied taxi can be defined as $R^n = (k, l^n, \tau^n)$ and $l^n = (x^n, y^n)$. So, the travel distance is defined as:

$$d = \sum_{i=1}^{N-1} \left| l_i - l_{i-1} \right| \tag{1}$$

where N is the total number of data samples in an unique trip with status of occupied or non-occupied, $|.|$ means the real distance in meters between two adjacent locations in geographic coordinates.

Finally, occupied and non-occupied trips are extracted for weekday/weekend and winter/summer respectively.

2.4. Multivariate spatial point pattern of origins and destinations

Since there are multiple types of spatial points in our study, which indicate the origin/destination points of occupied and unoccupied trips,