Characterizing Urban Dynamics Using Large Scale Taxicab Data

Xinwu Qian, Xianyuan Zhan and Satish V. Ukkusuri

Abstract Understanding urban dynamics is of fundamental importance for the efficient operation and sustainable development of large cities. In this paper, we present a comprehensive study on characterizing urban dynamics using the large scale taxi data in New York City. The pick-up and drop-off locations are firstly analyzed separately to reveal the general trip pattern across the city and the existence of unbalanced trips. The inherent similarities among taxi trips are further investigated using the two-step clustering algorithm. It builds up the relationship among detached areas in terms of land use types, travel distances and departure time. Moreover, human mobility pattern are inferred from the taxi trip displacements and is found to follow two stages: an exponential distribution with short trips and a truncated power law distribution for longer trips. The result indicates that the taxi trip may not fully represent human mobility and is heavily affected by trip expenses and the urban form and geography.

1 Introduction

The rapid urbanization process gives birth to megacities cities such as Tokyo, Shanghai and New York City (NYC). Not only are megacities big in terms of population density, they also bring up unprecedented opportunities and challenges. With large population density and tremendous human activities, one critical challenge is how to manage the giant urban system efficiently and sustainably. Urban dynamics represents the spatiotemporal principles followed by urban functioning evolvements [21].

X. Qian · X. Zhan · S.V. Ukkusuri (✉)
Purdue University, 550 Stadium Mall Dr, West Lafayette, IN, USA
e-mail: qian39@purdue.edu

X. Zhan
e-mail: zhanxianyuan@purdue.edu

S.V. Ukkusuri
e-mail: sukkusur@purdue.edu

N.D. Lagaros and M. Papadrakakis (eds.), *Engineering and Applied Sciences Optimization*, Computational Methods in Applied Sciences 38,
DOI 10.1007/978-3-319-18320-6_2

Understanding urban dynamics helps to capture the pulse of urban activities, which undoubtedly provides a huge step forward to address the problem.

In the past few decades, efforts have been made in modeling and simulating urban dynamics using data from transportation systems [1, 2, 8, 9]. However, the inherent complexity such as random behavior and the impact of geographical boundary can hardly be described properly using mathematical models. In the era of big data, we are widely exposed to various data sources and data-driven methods start to gain popularities. Compared with traditional data collected from surveys and questionnaires, the pervasive computing devices are able to collect abundant data in an efficient and accurate manner. Moreover, the digital footprints from mobile sensors such as GPS device and cellular mobile provides an opportunity to learn in-depth fundaments of human mobility. Several pioneering studies have implemented various data sources to reveal urban activity participation and individual mobility patterns [4, 7, 18, 19]. A case study in Milan discovered the urban spatiotemporal variations of activity intensity [18]. The intensity of activity locations is further used to locate hot spots and identify city structure by analyzing spatiotemporal signatures of Erlang data, which is a measurement of network bandwidth [19]. Gonzlez et al. revealed a highly regulated human mobility pattern [7] from 100,000 mobile phone users trajectories, and Calabrese et al. established a multivariate regression model to predict daily human mobility [4]. Hasan et al. [10] examined both aggregate and individual activity patterns from social media check-in data. Brockmann et al. [3] studied the distance of human travel from the distribution of bank notes. They found that distance distribution follows a power law and can be well approximated using continuous-time random walk.

In large cities, public transportation is the direct carrier of urban life and taxicab is an indispensable component of it. As of 2007, 10 % of total passenger volume are served by 18,000 taxicabs in Hong Kong [23]. By the end of 2012, 55,000 of taxis transport 1.5 million passengers daily [15]. Equipped with GPS devices, the taxi trip data enjoys the merit of sufficient temporal and spatial coverage due to the large passenger volume and 24-7 operation hours. Therefore, it is an advantageous data source for urban studies and has already received great attentions from researches. The pick-up and drop-off locations are processed with data mining and clustering algorithms to reveal urban activity patterns such as hotspots information for taxi drivers [5, 12, 24] and land use inference [14, 16]. However, urban dynamics are merely understood on the surface if pick-up/drop-off locations are only analyzed separately. Taxicab provides door-to-door service and is often used as a non-stop transportation tool. Therefore, the joint analysis of pick-up and drop-off location builds up a direct bridge between origin and destination and can aggregately reveal the underlying connections among detached urban places. Moreover, since each taxi trip is a form of human movement, the taxi trip data is also analyzed to disclose the uniformity of human mobility in large cities [11, 13, 17]. However, result discrepancies result are observed among very limited works. As a special case of human mobility, the displacement of taxi trips is restrained by the trip expenses and more importantly, the functionality structure of a city. Therefore the relationship between taxi trips and human mobility requires further investigation. In this paper, we make a comprehensive use of the large scale taxi trip data and present the study on urban dynamics pattern in NYC from three aspects. First, the spatiotemporal pattern of urban activities is examined from trip dynamics by aggregating pick-up and drop-off locations. Secondly, we explore the inherent similarities among taxi trips and reveal the underlying connections among detached places using two-step clustering algorithms. In the end, we investigate the relationship between the taxi trips and uniformity of human mobility.

The rest of the paper is organized as follows. Section 2 gives an overview of the data and Sect. 3 analyzes the demand pattern of overall study area and several hot spots. Then the similarity among different trips is captured and the mobility pattern of taxicabs is presented. Conclusions and limitations are discussed in the final part.

2 Data

The taxi trip data used in this research is collected by New York City Taxi & Limousine Commission (NYCTLC) from December, 2008 to January, 2010. About 300,000 to 500,000 daily trips are recorded during the time and an overview of the annual trip distribution in 2009 is given in Fig. 1. A repeated and stable pattern is observed for weekly trips over the year and drastic drops are detected on holidays such as Thanks Giving and Christmas. Approximately 300,000 to 500,000 daily trips are recorded during the study period.

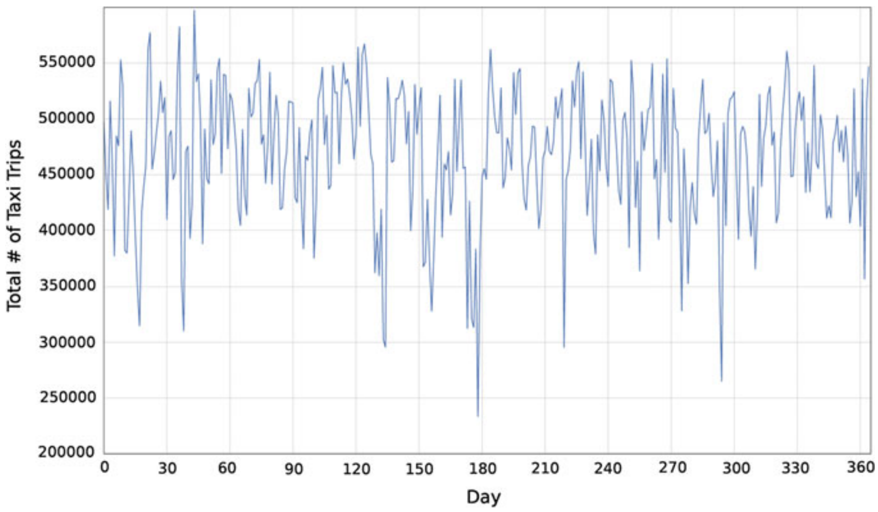

Fig. 1 Annual daily distribution of taxi trips

Table 1 Taxi data statistics

Date	Number of trips recorded	Number of trips after cleaning
10.5.2009	431,828	428,553
10.6.2009	467,649	464,273
10.7.2009	492,914	488,895
10.8.2009	517,079	512,781
10.9.2009	536,039	531,965
10.10.2009	532,179	528,032
10.11.2009	454,573	451,059

The dataset contains complete trip information, including the pick-up and drop-off timestamps and locations, the number of passengers onboard, the travel distance and the trip expense. Detailed trip trajectories are not available due to privacy concerns.

In addition to taxi data, census tract geography and land use information are also introduced in the analysis. The census tracts are extracted from the census tract area file provide in TRANSCAD. There are 2,211 census tracts within the study area, which cover Manhattan, Bronx, Queens, Brooklyn, Long Island, and a small portion of New Jersey. The land use map implemented in the study is obtained from New York City Department of City Planning (NYCDCP), which divides the city into three fundamental zoning districts: commercial (C), residential (R) and manufacturing (M). The three types are further categorized from low density to high density.

The taxi trip data from October 5th–11th are processed for further analysis, where no major social events were recorded during the period. The statistics of the one week data is presented in Table 1. Erroneous trip records are firstly removed, such as trips with zero travel distance or fare less than the initial price. Then all pick-up and drop-off locations are coupled with geography map to eliminate trips outside the study area. Finally, the remaining trips in the dataset are viewed as qualified and tagged with the overlaid census tract ID and land use type.

3 Trip Dynamics

3.1 Overall Pattern

In this section, patterns of urban activity participation are examined from the arrival and departure dynamics of taxi trips. NYC is one of the busiest cities in the world. Around 5.7 million passengers moving around the city during the study period, generating more than 3.4 million taxi trips. The pick-up and drop-off location of all trips are aggregated at the census tract level based on the geographical coordinates and the overall geographical distributions of taxi trips are visualized in Fig. 2.

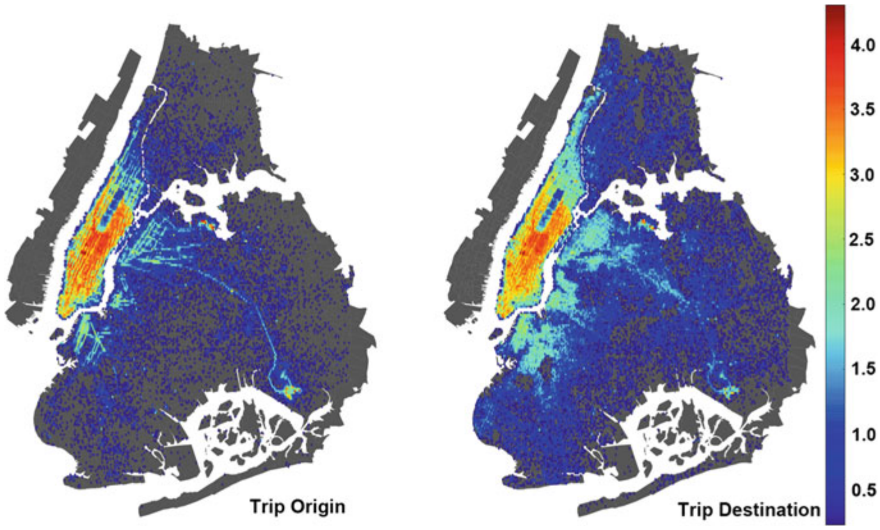

Fig. 2 Aggregated weekly density plot

The most appealing observation is that both trip origins and trip destinations exhibit highly centralized distribution towards Manhattan area. The result is not surprising since Manhattan serves as the business center of NYC. The number of trips decreases significantly with the increase of the distance to the city center, which reflects the typical sprawl of urban forms. While most places far from Manhattan have very low amount of trips, patterns at LaGuardia airport (LGA) and John F. Kennedy international airport (JFK) are entirely different. Approximately 90 % of total trips are associated with Manhattan area. While majority of the trips congregate at midtown Manhattan and lower Manhattan, the upper Manhattan area is apparently less preferred by both passengers and drivers.

3.2 Hot Spots

Hot spots refer to the most frequent visited places in a city and usually have great activity intensity. The analysis of hotspots dynamics helps to understand the urban functionality in depth. By ranking total trip frequencies, most popular places are identified and five specific tracts are selected which cover the LGA, JFK, Penn Station, Central Park and the Fifth Avenue (the segment between 49th street and 56th street). Each individual hotspot has indispensable functionality including transportation terminals (with different purposes), recreational place and commercial area. To analyze the dynamics at hotspots, the temporal patterns across the week are plotted in Fig. 3. Penn Station and Madison Square Garden locate in the census tract where

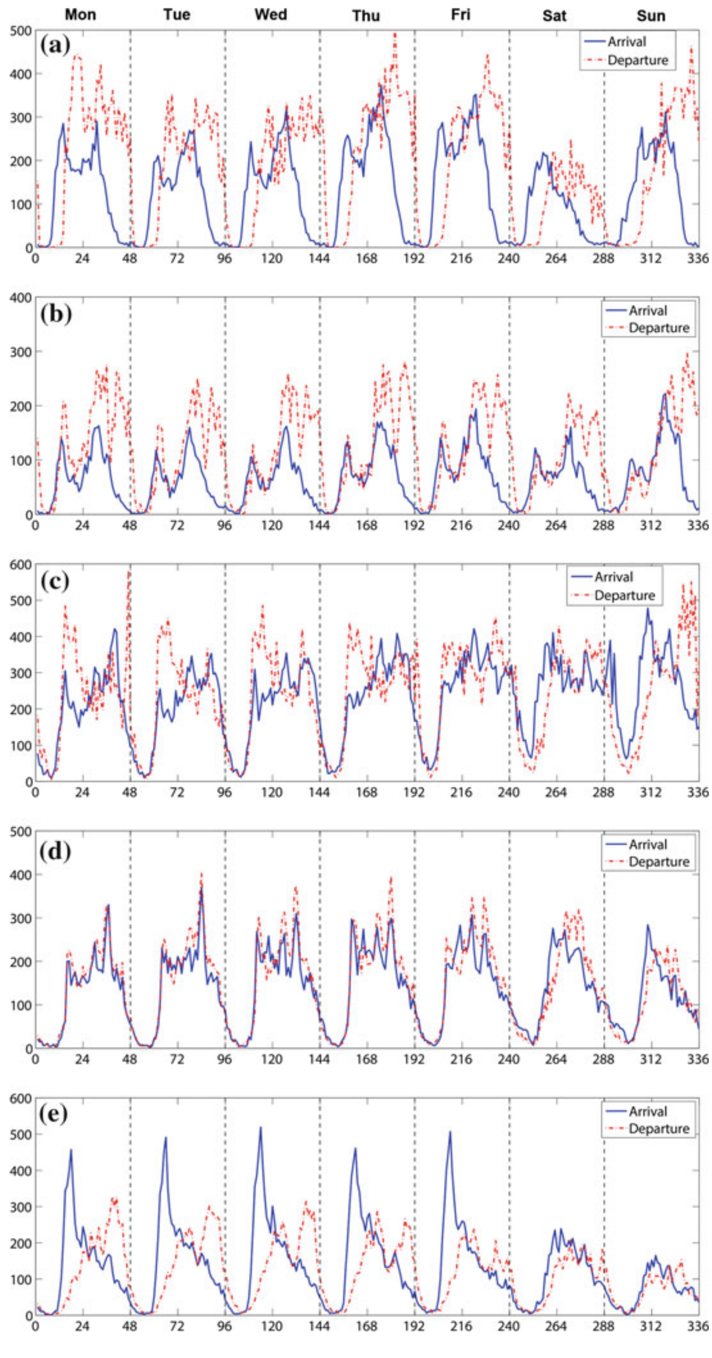

Fig. 3 Weekly trip pattern at hotspots x-axis is the time horizon and y-axis represents the number of trips **a** LaGuardia Airport, **b** JFK Airport, **c** Pen Station, **d** Central Park, **e** 5th Avenue

the greatest number of taxi trips are generated. Penn Station is not only the terminal for Amtrak trains, it also serves as the connection station for multiple subway lines. According to the morning arrival (trip origin) and evening departure (trip destination) peaks in weekdays, taxicab is very likely to function as the last and first mile transportation. Over the weekend, most arrivals and departures take place within daytime and at night. The pattern coincides with the functionality of Madison Square Garden, which is an entertainment place and is surrounded by many hotels.

Trip patterns at airports are distinct from that in central part. For both airports, while arrival curves are comparatively smooth, departure curves are observed to be noisy due to the periodical entry of flights. Besides, the intrinsic differences between the two airports are also disclosed from trip dynamics. Due to the effect of travel distance, the trip amount at JFK is significant lower than that at LGA. Secondly, since LGA are mainly used for domestic flights, the apparent morning peaks for flight arrivals during weekdays and the drop of trip amount on weekends. Moreover, as an airport mainly for international flights, JFK has more arrivals in the afternoon and the pattern is surprisingly consistent over the week. The result suggests that, during the week, the trip purpose is stable for international flights but varying significantly for domestic flights.

The Central Park is a recreational place. It occupies a larger area compared with other census tracts which contributes to its trip frequency. The comparison of trip dynamics between at the Central Park and at the Fifth Avenue perfectly interprets the functionality of corresponding land use attributes. The Fifth Avenue is a remarkable business street at midtown Manhattan and morning taxi arrival and evening taxi departure peaks are unsurprisingly retrieved. Reversely, due to the large portion of residential areas around the Central Park, most departures take place in the morning and majority of taxi arrivals are observed during evening rush hours.

3.3 Unbalanced Taxi Trip

Except for being able to capture activity dynamics at hotspots, the data also carries implicit yet significant insights such as the existence of unbalanced taxi trip. The number of taxi trips is closely associated with the level of economic development and the variation of urban functionality. Due to concerns such as trip margins and safety issues, taxi drivers usually have their preferred destinations, which eventually leads to the geographical discrimination. For example, taxi drivers may be unwilling to make trips to destinations where it is hardly possible to find potential passengers. The second type of unbalanced taxi trips is usually caused by sudden fluctuations in passenger demand. While the supply of taxis is fixed, the influx of commuters during peak hours makes it extremely hard to hail a vacant taxi.

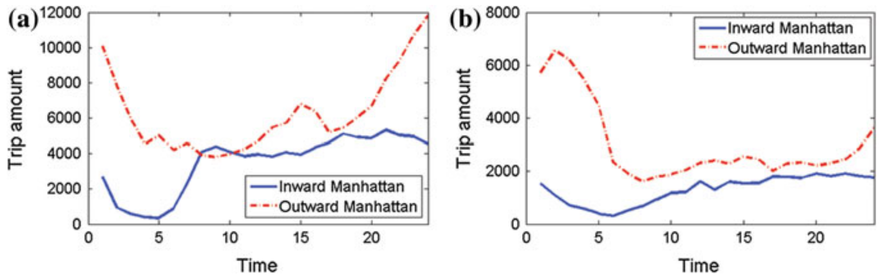

Fig. 4 Inward/outward-manhattan unbalanced trips **a** Weekday, **b** Weekend

From the overall spatial distribution, we observe a tremendous centrality of taxi trips at the developed Manhattan area. The great trip density suggests the easiness of finding passengers in Manhattan and stickiness of drivers to Manhattan area. As a result, we start looking into phenomenon and plot the temporal distributions for trips inwards and outwards Manhattan in Fig. 4. During daytime, the overall pattern for inward and outward trips turns out to be stable and balanced. However, when time goes to late night, we surprisingly witness an enormous gap: the highest amount of outward trips and the lowest amount of inward trips take place simultaneously. People may stay at Manhattan very late for entertainments and relaxations, while buses and metros having a reduced accessibility at the time. As taxi becomes very popular at a late time, drivers may refuse to leave Manhattan as they have to run the risk of returning empty. Hence, the unbalanced trip pattern implies the existence of geographical discrimination and a reduced level of service for taxi industry.

In order to reveal the unbalanced condition inbound Manhattan, we extract only weekday trips and spatial distributions of trip origins and destinations are presented in Fig. 5. Three typical time intervals are selected which cover off-peaks and morning and evening rush hours. Both morning peak and evening peak display eminent differences between trip origins and destinations and their patterns appear to be symmetric. Moreover, trips are found to be unbalanced with notable geographic characteristics. The northeastern part of midtown Manhattan is a large residential area and the midtown is mainly covered by commercial floors. As a result, most taxi trips inflow into midtown during morning peak and dissipate from the center area in the evening.

The existence of unbalanced taxi trips suggests an imminent need of designing policies to mitigate negative impacts. An additional fee can be charged or a subsidy can be assigned for trips outward Manhattan only after midnight as taxi drives are less likely to leave Manhattan at that time. Moreover, since morning and evening trips have distinct origins and destinations, the shuttle service following the direction of human migration should be to be effective. It can narrow the demand-supply gap of taxi service and reduce congestion at the same time.

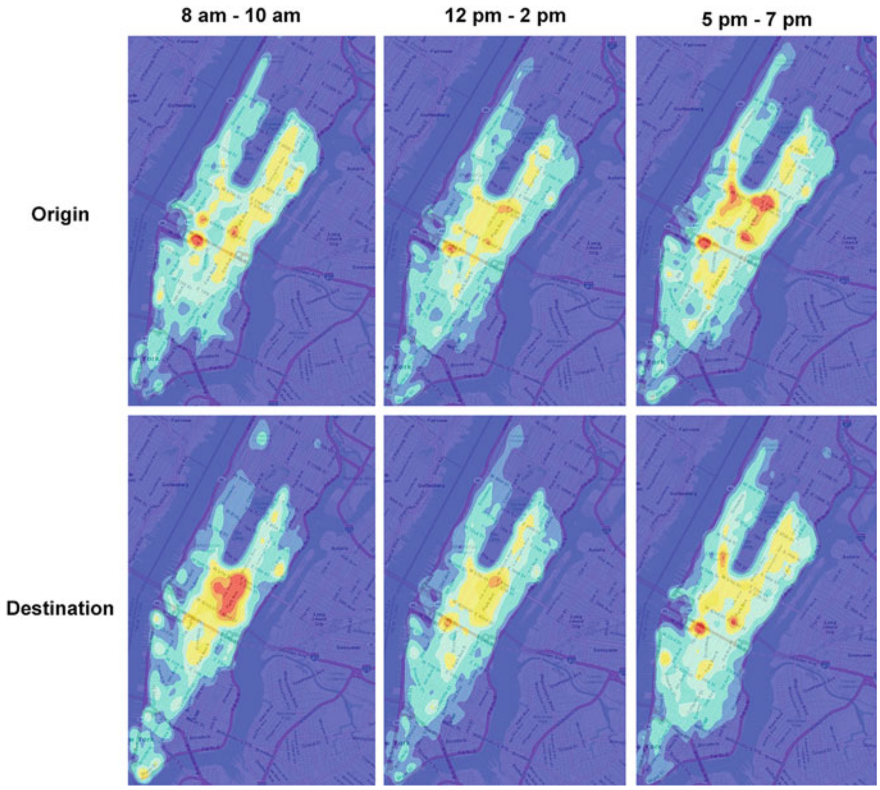

Fig. 5 Trip density plot inbound manhattan density increases from *blue* to *red*

4 Trip Classification

4.1 Clustering Algorithm

It is recognized that dynamics of trip origins and destinations are largely influenced by the geographical location, land use pattern and functionality of a particular place. Moreover, unlike other public transportation modes, the door-to-door service of taxicab builds up the straightforward connection between trip origin and destination. Therefore, how different urban areas are related can be understood by exploring the inherent similarities of taxi trips. Clustering algorithms are widely used to classify individual cases in large database into homogeneous groups. Considering spatial and temporal characteristics of taxi trips, each piece of taxi trip x_i can be represented as an eight dimensional tuple which takes the form:

$$x_i = (lat_i^o, long_i^o, lat_i^d, long_i^d, p_i^o, p_i^d, d_i, t_i) \quad (1)$$

Where o, d represent the trip origin and destination respectively, lat and $long$ are the latitude and longitude of trip locations, p refers to the land use attribute, d is the trip distance and t stands for the trip starting time. The clustering problem cannot be tackled by popular approaches such as k-means and DBSACN due to the presence of categorical variables (land use attribute).

Alternatively, the two-step clustering algorithm [6] is implemented to address the mixed variable clustering problem following two stages. The first stage is a pre-clustering approach which uses a sequential clustering method to generate initial sub-clusters. The second stage uses the agglomerative hierarchical approach which processes the sub-clusters from in the first stage recursively. The number of clusters is determined automatically by comparing BIC values. For interested readers, the detailed description for each step of the algorithm can be referred to SPSS manual [20].

4.2 Clustering Result

An overview of the clustering result is presented in Fig. 6. For both weekday and weekend taxi trips, the exactly same configuration with 7 distinct trip groups is obtained. Moreover, the percentage for the same cluster is pretty close. We name each cluster by its land use feature accordingly, including C-C, R-C, C-R, R-R, Mul (Mixed land use type)-Mul-S (short trip distance), Mul-Mul-L (long trip distance), and Mul-M trips. To better understand the characteristics of each cluster, the spatial distributions of trip origins and destinations in each cluster on weekdays are visualized in Fig. 7.

In general, C-C trips contribute to over one-third (36.0 % for weekday and 34.9 % for weekend) of the total taxi trips in NYC. Further, there are another 30 % of trips that

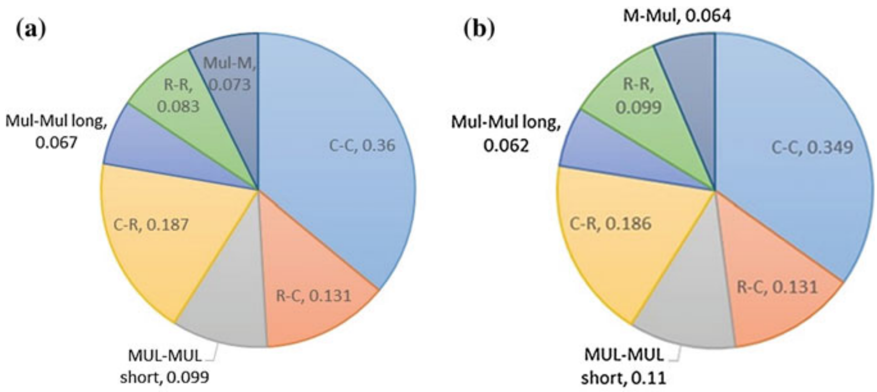

Fig. 6 Clustering result C-Commercial, R-Residential, M-Manufacturing, Mul-Mixture of the three Short/Long-Short/Long travel distance **a** Weekday, **b** Weekend

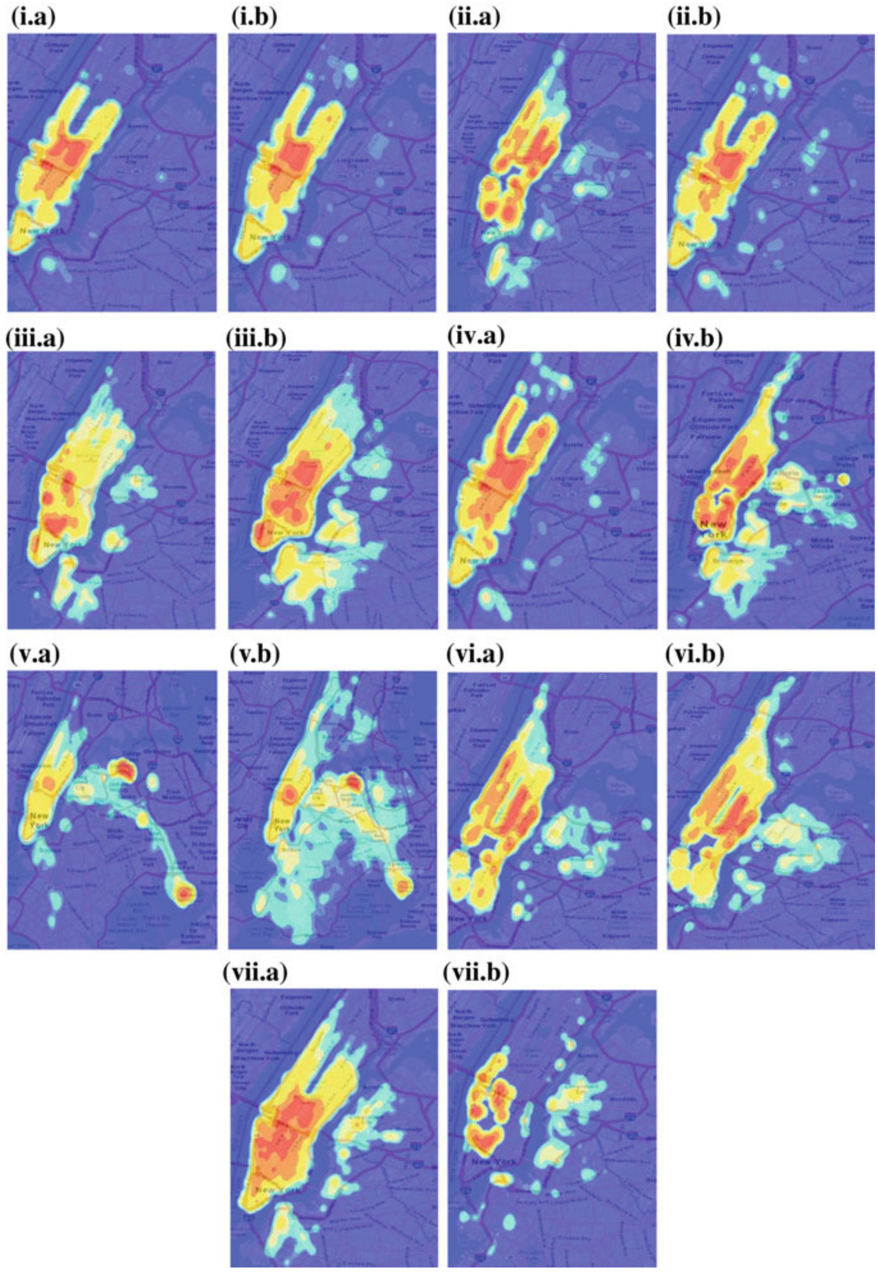

Fig. 7 Spatial density plot of cluster origins and destinations **a** for trip origin and **b** for destination; i:commericial to commercial; ii: residential to commercial; iii: mixed to mixed with short travel distance; iv:commercial to residential; v: mixed to mixed with long travel distance; vi: residential to residential; vii: mixed to manufacturing; density increases from *blue* to *red*

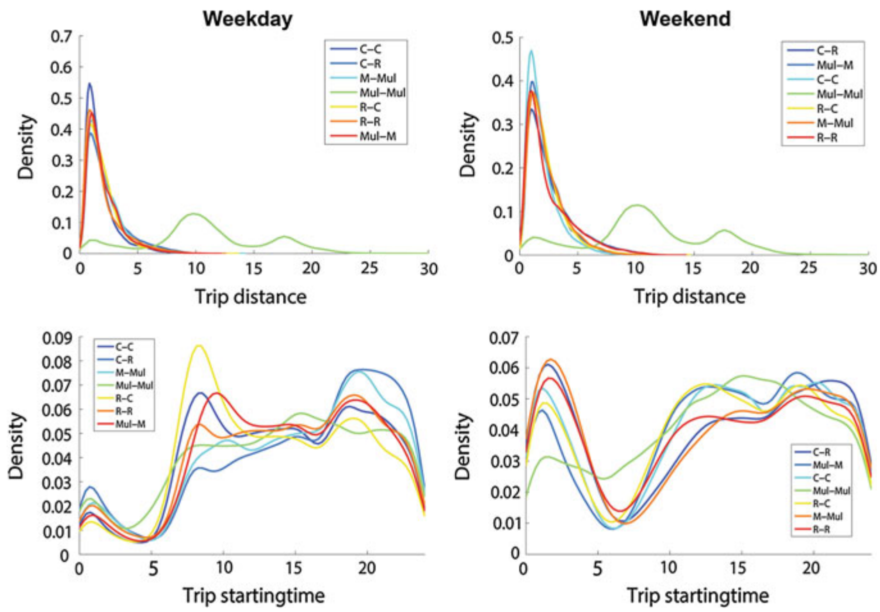

Fig. 8 Travel distance and trip starting time distribution for 7 clusters

are associated with commercial area (with either origin or destination in commercial area). This suggests the significant impact of land use pattern, especially commercial floors, on the amount of taxi trips. More specifically, commercial areas where trip originated from and arrived at cover the entire midtown and lower Manhattan. As a result, it is believed that most activities and functionalities of the city are concentrated in these places. Viewing the distribution of residential related trips, one can tell that there are considerable amount of people living on the peripheral area and they are connected to the city center by taxicab.

We also plot distributions of travel distance and trip starting time as important attributes for each cluster in Fig. 8. Apparently, the distance distribution suggests that taxi trips are heavily used for short-range travel, especially for trips less than 5 miles. Such pattern is mainly determined by the urban structure of NYC, as majority activities and functional places are agglomerated in a small area. While the distance distribution is stable over the week, there are prominent discrepancies observed for trip starting time between weekday and weekend. Firstly, all clusters except C-R and Mul-Mul-L trips have morning and evening peaks, reflecting that taxicabs are heavily used for work commuting in urban areas. Secondly, the temporal pattern of most urban activity is shifted from daytime to late night, as the trip intensity remains at a high level until 3 am.

Though taxi trips are mostly commercial and residential related, we observe that the Mul-Mul-L group is a very special type of taxi trips with unique characteristics. Based on the trip location distribution, these trips connect midtown Manhattan, LGA and JFK to the rest of NYC. While all other clusters have very short trip distance, the mean travel distance of the group is approximately 11 miles. Two peaks are revealed from the distance distribution, which locate at 10 miles and 17 miles. The two points are matched with the travel distance from Manhattan to LGA and JFK respectively. As a result, the exclusive pattern is largely caused by the urban forms, as airports are usually far from the city center but with very high passenger volumes. The group of trip should be treated separately during urban studies as it is heavily biased from the general mobility pattern of taxi trips.

5 Taxi Mobility

Individual mobility pattern have been realized barely random. Several studies using data from the movement of an online game [22], the dispersal of bank notes [3], as well as trajectories from cellular data [7] have found highly regulated pattern in human movement. And the human movement is observed to follow a heavy-tailed plot under logarithmic scale and can be well approximated by scaling law. With human beings as the main participants, the taxi trips are results of human movement in an urban context as well. Hence, we try to reveal the taxi mobility and examine the relationship with individual mobility.

To uncover the taxi mobility, we first plot the distribution of travel distance under logarithmic scale in Fig. 9a. From the observation, the distribution of travel distance can be divided into parts: an ascending ranges from 0 to 0.8 mile, and then gradually descending as trip distance increases. Two minor peaks around 10 miles and 20 miles in the distribution are mainly caused by trips to LGA and JFK airport. The interference of airport trips has been discussed in previous section. We remove the trips to and from the two airports as they have specific purposes and unique characteristics. A refined distribution is generated in Fig. 9b.

Trips with distance less than 0.8 mile take 16.89 % of total trips. As very short trips within walking radius, these trips differ from the general pattern of taxi mobility on a decision making process of whether to take taxis. The first part of the trips can be approximated with distribution:

$$P(d) \propto d^{\beta} \tag{2}$$

Where exponent $\beta = 1.2505$.

The distribution resembles a power-law like distribution (straight line under logarithmic scale), however, the exponent takes a positive value. As mentioned earlier this phenomenon captures model choice process in whether take a taxi. And it is intuitive that with the increase in distance, the probability of taking a taxi also increases until attaining its maximum around 0.8 miles.

The refined second part is used to capture urban mobility features of taxi trips. The trips greater than 0.8 mile contribute 83.11 % of total trips. It is found that the

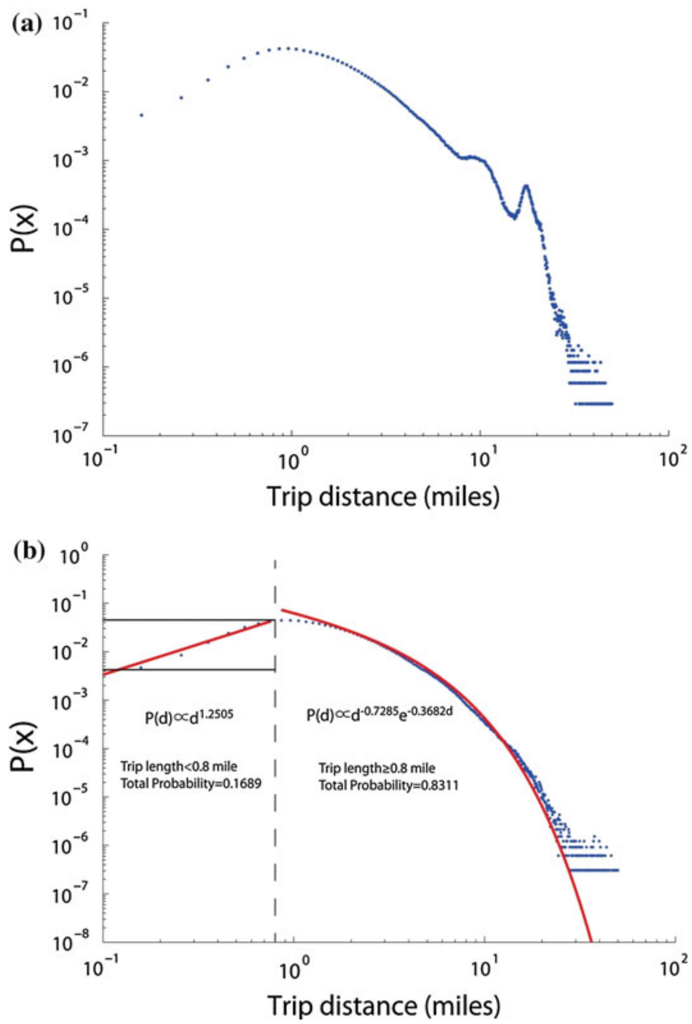

Fig. 9 Taxi trip distance distribution. **a** Distance distribution of all trips. **b** Distance of all but airport trips

distribution of taxi trip distance is well approximated by a power-law with exponential cut-off (also known as truncated power-law):

$$P(d) \propto d^{-\alpha} e^{-\lambda d} \tag{3}$$

With exponent $\alpha = 0.7285$ and $\lambda = 0.3682$. The distribution is found to be heavy-tailed. Unlike the power-law distribution of human movement reported (8, 18, 19), the taxi trip distance distribution has a faster probability decay in the tail part (the effect of the exponential cut-off term). This indicates that the unique effects of urban environment on the distribution of taxi trip distance. Since the underlying size of urban area limits the distance of taxi trip, very long trips (e.g.$>$30 miles) are less likely to happen, and the scale-free property of a typical power-law distribution fails. It is notable that as taxi trips are important component of urban human movement, the trip distance distribution reflects a unique perspective of human mobility. That is, the taxi mobility pattern reveals the hidden role of urban geographical boundaries in limiting urban human movement.

6 Conclusion and Future Work

In this paper, we exploit New York taxi trip data and comprehensively explore underlying patterns of urban taxi trips. We first look at the general level of demand and find out the spatial and temporal patterns for the most popular places. A potential unbalanced trip pattern is further discussed. Next, we use the two-step clustering algorithm to figure out the intrinsic taxi trip classes. Differences are discussed based on land use, travel distance and starting time distributions. In the end, taxi trip mobility is analyzed from the overall travel distance distribution.

Taxi data has been proved to be an efficient tool to understand urban dynamics and several interesting insights are raised in our paper. Unbalanced trips are common in taxi industry and should be carefully investigated to improve the level of service. Airport trips is a special part of taxi trips and differ from regular taxi trip patterns. Land use has significant impact on taxi trip types, and different types of taxi trips are able to uncover the structure of a city. Moreover, we find that the mobility of taxi trips are restricted by the urban geographical boundaries.

However, the paper also has several limitations. The current paper is primarily focused on exploring patterns. The following study will build a model from the patterns discovered to account for human movement within urban context. Moreover, more information such as social economics can be combined into the data analysis to provide more insights. Furthermore, it would be interesting to develop a methodology to infer urban land use type from taxi patterns. Also, attentions can be paid on extracting travel information from taxi dynamics and provide feedbacks to users.

References

1. Batty M, Xie Y (1994) From cells to cities. Environ Plan B 21:31
2. Batty M, Xie Y, Sun Z et al (1999) Modeling urban dynamics through GIS-based cellular automata. Comput Environ Urban Syst 23:205–233
3. Brockmann D, Hufnagel L, Geisel T et al (2006) The scaling laws of human travel. Nature 439:462–465
4. Calabrese F, Diao M, Di Lorenzo G et al (2013) Understanding individual mobility patterns from urban sensing data: a mobile phone trace example. Transp Res Part C Emerg Technol 26:301–313

5. Chang H, Tai Y, Chen H et al (2008) iTaxi: context-aware taxi demand hotspots prediction using ontology and data mining approaches. In: Proceedings of 13th conferance artificial intelligence and applications (TAAI 2008)
6. Chiu T, Fang D, Chen J et al (2001) A robust and scalable clustering algorithm for mixed type attributes in large database environment. In: Proceedings of seventh ACM SIGKDD international conferanceon of knowledge discovery and data mining—(KDD 01) 263–268
7. Gonzalez MC, Hidalgo CA, Barabasi A-L et al (2008) Understanding individual human mobility patterns. Nature 453:779–782
8. Guiliano G (2004) Land use impacts of transportation investments: highway and transit. In: Hanson S and Giuliano G (eds) Geography of urban transportation, 3rd edn. Guilford Press, New York
9. Harris B (1985) Urban simulation models in regional science. J Reg Sci 25:545–567
10. Hasan S, Zhan X, Ukkusuri SV et al (2013) Understanding urban human activity and mobility patterns using large-scale location-based data from online social media. In: Proceedings of 2nd ACM SIGKDD international work. Urban computing, p 6
11. Jiang B, Yin J, Zhao S et al (2009) Characterizing the human mobility pattern in a large street network. Phys Rev E 80:21136
12. Li B, Zhang D, Sun L et al (2011) Hunting or waiting? discovering passenger-finding strategies from a large-scale real-world taxi dataset. In: IEEE international conference on pervasive computing and communications workshops (PERCOM Workshops), pp 63–68
13. Liang X, Zheng X, Lv W et al (2012) The scaling of human mobility by taxis is exponential. Phys A Stat Mech Its Appl 391:2135–2144
14. Liu Y, Wang F, Xiao Y et al (2012) Urban land uses and traffic 'source-sink areas': evidence from GPS-enabled taxi data in Shanghai. Landsc Urban Plan 106:73–87
15. NYCTLC(2012) New York city taxi and limousine commission (2012) Annual Report
16. Pan G, Qi G, Wu Z et al (2013) Land-use cassification using taxi GPS traces. IEEE Trans Intell Transp Syst 14:113–123
17. Peng C, Jin X, Wong K-C et al (2012) Collective human mobility pattern from taxi trips in urban area. PLoS One 7:e34487
18. Ratti C, Pulselli RM, Williams S et al (2006) Mobile landscapes: using location data from cell phones for urban analysis. Environ Plan B Plan Des 33:727–748
19. Reades J, Calabrese F (2007) Cellular census: explorations in urban data collection. Pervasive Comput IEEE 6:30–38
20. SPSS INC (2001) The SPSS twoStep cluster component: a scalable component to segment your customers more effectively
21. Sun L, Chen C, Zhang D et al (2013) Understanding urban dynamics from Taxi GPS traces. Creat Pers Soc Urban Aware Through Pervasive Comput, p 299
22. Szell M, Sinatra R, Petri G et al (2012) Understanding mobility in a social petri dish. Sci Rep 2:457
23. Yang H, Fung CS, Wong KI et al (2010) Nonlinear pricing of taxi services. Transp Res Part A Policy Pract 44:337–348
24. Yuan J, Zheng Y, Zhang L et al (2011) Where to find my next passenger. In: Proceedings of 13th international conferance on ubiquitous computing, pp 109–118