Mining Raw Trajectories for Network Optimization from Operating Vehicles

Lei Ning[1(✉)], Runzhou Zhang[1,2], Jing Pan[3], and Fajun Li[1]

[1] College of Big Data and Internet, Shenzhen Technology University, Shenzhen, China
ninglei@sztu.edu.cn
[2] College of Applied Technology, Shenzhen University, Shenzhen, China
[3] Chaincomp Technologies, Shenzhen, China

Abstract. Improving the user peak rate in hot-spots is one of the original intention of design for 5G networks. The cell radius shall be reduced to admit less users in a single cell with the given cell peak rate, namely Hyper-Dense Networks (HDN). Therefore, the feature extraction of the node trajectories will greatly facilitate the development of optimal algorithms for radio resource management in HDN. This paper presents a data mining of the raw GPS trajectories from the urban operating vehicles in the city of Shenzhen. As the widely recognized three features of human traces, the self-similarity, hot-spots and long-tails are evaluated. Mining results show that the vehicles to serve the daily trip of human in the city always take a short travel and activate in several hot-spots, but roaming randomly. However, the vehicles to serve the goods are showing the opposite characteristics.

Keywords: Trajectory mining · Vehicle mobility · Hyper-dense networks

1 Introduction

With the development of network and microelectronics technology in the past decades, the network world has gradually expanded its connection between people and things as well as between things from that between people. By 2025, the Internet of Things (IoT) will have more than 55 billion connections, the report said. The explosive growth of IoT expands the connectivity of the network and the way of data exchange, such as portable electronic equipment, household appliances, vehicles and manufacturing devices, and a series of integrated devices including electronics, software, sensors, drivers and networks. From consumer wearable devices to industrial production devices, these networked devices can sense environmental information, be controlled remotely, make decisions and take

Sponsored by the Young Innovative Project from Guangdong Province of China (No. 2018KQNCX403) and the Teaching Reform Project from Shenzhen Technology University (No. 2018105101002).

X. Wang et al. (Eds.): 6GN 2020, LNICST 337, pp. 207–217, 2020.
https://doi.org/10.1007/978-3-030-63941-9_15

actions by themselves; however, the number of users and their business requests are unevenly distributed in the network. Especially in the era of IoT, a large number of business requests are initiated in hot area collectively including indoor, business center, stadium, factory, farm and other node-intensive areas. Meanwhile, the business type of voice-intensive is changing to that of data-intensive and connection-intensive.

In order to meet the above requirements, in the fifth generation mobile communication technology (5G), its performance design index not only has high-speed rate, low delay and large connection, but also proposes the concept of hyper dense networking for hot areas, that is, improving the cell capacity per unit area under the condition of mixed multiple network systems. Due to the cellular characteristics of mobile networks, there are three technologies mainly to improve cell capacity as follows, increasing spectrum resources, improving wireless link performance and enhancing cell density respectively.

Hybrid Networking technology, as the multiple networks deployed in the same area, provides multi-mode terminals to select appropriate network access according to business needs and network status; Large Scale MIMO and New Multiple Access technology is to improve the wireless link performance in vertical space; while Small Cell technology improves the spectral efficiency in horizontal space by increasing the cell density in specific area [23]. In hot area, user's mobile behavior under the multiple and hyper dense deployment network greatly increases the difficulty for the network to guarantee the user service experience [22]. User's mobile behavior evolution under multiple networks cooperation is closely related to the development of hyper dense network technology, therefore, user's mobile behavior will significantly affect the data bearing distribution of different networks, as well as affect the overall performance and user experience of the network.

With the integrated development of IoT and wireless access technology, user's behavior in the network is complex and changeable, because its accessing users are no longer limited to human portable devices, but also include animals with sensors and communication devices, as well as machines with autonomous mobile functions. As the result, analyzing the characteristics of user's mobile trajectory under the hyper dense network, is helpful to optimize the network performance and promote the development of network technology for intelligent manufacturing and interconnection of all things [24].

Recently, Feng et al. [5] summarizes a survey on trajectory data mining, including main techniques and applications, which a wide spectrum of applications is driven by trajectory data mining, such as path discovery [3,4,10–12,18], movement behavior analysis [1,6,15,19], group behavior analysis [14], urban service [9,20,21] and so on.

Some of researches focus on human behavior analysis based on raw trajectory data, to understand people's real demand for transportation, driving preferences, extracting mobility behavioral patterns, which can be used to enhance utilization efficiency of public transportation [10], improve quality of user satisfaction [4], understand behavior of people moved in geographical context [15].

Moreover, another work [6] explores individual human mobility patterns by studying a large number of anonymous position data from mobile phone users and reveals a high degree of temporal and spatial regularity in human trajectories. While others focus on unhuman behavior analysis based on raw trajectory data, they believe that knowledge discovered with trajectory data mining techniques helps to improve quality of life in urban areas from several aspects [8,9,20,21]. Yuan et al. [21] address a problem of discovering regions of different functions in a city based on a large scale of trajectory data. Liu et al. [9] address a problem of map inference in a practical setting through inferring road maps from large-scale GPS traces. Especially, the vehicles behavior analysis based on trajectory data are of great importance in unhuman field, which abstract more attention. iPark in the literature of Yang et al.[20] aims to enable parking information, i.e., annotating an existing map with parking zones based on trajectory data of vehicles. Through analyzing a large scale of trajectory data collected from electronic vehicles, Li et al. [8] address that how to strategically deploy charging stations and charging points so that minimizing average time to the nearest charging station and average waiting time for an available charging point. Lee et al. [7] collected real users' data through Garmin-GPS-60CSx handheld terminal in the North American environment. Based on a large number of measured data collected above, a comprehensive moving model reflecting the characteristics of data samples is proposed. In such researches, real data plays more and more important role, and comparing with the researches on human behavior analysis. Thus, in this paper, the urban operating vehicles data will be used as the raw trajectories to mining and analyzing its characteristics and patterns and draw some conclusion, which will do some contribution for future modeling as well as network performance optimization. At the same time, user's mobile behavior tends to study the implicit constraints beyond the time and space information in user's mobile trajectory, in raw GPS trajectories from the urban operating vehicles data, there are many kinds of users in the network and their behaviors are complex and changeable. It is of great theoretical value and practical significance for the network performance optimization and intelligent deployment of IoT to study the mobile behavior characteristics of multiple types of users under the IoT and hyper dense network.

2 Trajectory Data Description

In this paper, the raw GPS data is published from Shenzhen Transportation Bureau. The trajectories contain 113,503 entries in a single day from the urban operating vehicles up to five types with the number of 29,218. This big data are gathered from 12:00 am, Oct. 08, 2018 to 11:59 pm Oct. 14, 2018 in the local time. The number of raw GPS coordinates in one typical trajectory is 295,966,347. For example, in Fig. 1, it shows the heap map of a random selected trajectory of the vehicle, which mainly take the activities around the central part ot the city.

In the following part, the basic definitions which are used throughout the paper are introduced for data descriptions.

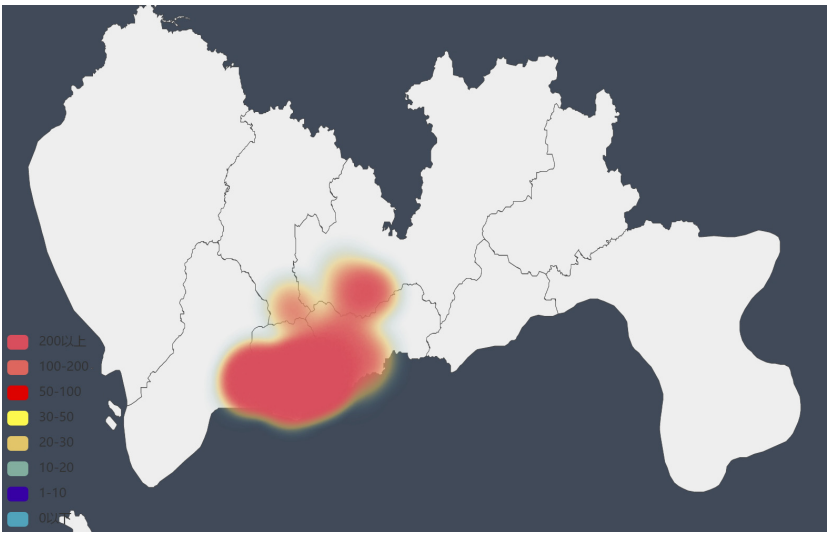

Fig. 1. The heap map of a random selected trajectory of the vehicle.

Definition 1 (Raw GPS Data): It is a 4-tuple of the form (lat, lon, t, p) where lat and lon is the vehicle's latitude and longitude respectively, t is the timestamp at which the record was tracked, p is the plate number of the vehicle that also indicates the type.

Definition 2 (Trajectory): It is a sequence of time-ordered raw GPS data for the specific vehicle in one day.

Definition 3 (Distance): It is a length of a line segment between two given coordinates. The earth radius shall be considered since the coordinates from the raw GPS data is presented by longitude and latitude respectively.

3 The Feature Extraction Method of the Trajectory

Previously some related work have been proposed that the human mobility has regularities, which are self-similar, hot-spots and heavy-tail [7,13]. Based on those discovered regularities and the raw GPS data of several vehicles, an extraction method of the trajectory is presented to evaluate the vehicle mobility features from the urban scale in this section.

3.1 Evaluation of Self-Similarity with Hurst Exponent

A system with Hurst statistical characteristics does not need the independent random event hypothesis of general probability statistics [2]. It reflects the result of a long series of interrelated events. What happens today will affect the future,

and what happened in the past will affect the present. Accordingly, the human always selects a familiar path from the constant location to the temporary destination, which is called self-similarity. Therefore, the Hurst exponent is adopted to analyze the self-similarity of trajectories.

The aggregated variance and the R/S methods are full-blown implementations of Hurst exponent algorithm. However, the candidate raw GPS data are huge. So the algorithm proposed by Bill Davidson named BD procedure in what follows is to quantify the self-similarity of way-points, which is far faster than the conventional algorithm.

Algorithm 1. BD procedure

get the set of raw GPS data $\boldsymbol{L}$
while $\boldsymbol{L}_{length} \geq \boldsymbol{L}_{threshold}$ **do**
  $\boldsymbol{L}_y = std(\boldsymbol{L})$
  $\boldsymbol{L}_x = \boldsymbol{L}_x \times 2$
  $\boldsymbol{L}_{length} = fix(\boldsymbol{L}_{length}/2)$
  for all $index \in \boldsymbol{L}_{length}$ **do**
    $\boldsymbol{L}_{tmp} = (\boldsymbol{L}(2 \times index) + \boldsymbol{L}((2 \times index) - 1)) \times 0.5$
    get new $\boldsymbol{L}$ from $\boldsymbol{L}_{tmp}$
  end for
end while
make the linear fit for $\boldsymbol{L}_y$ and $\boldsymbol{L}_x$
get Hurst exponent from the slope of the linear fit of log-log plot

3.2 Evaluation of Hot-Spots with Density-Based Clustering

From a macro perspective, human always activates in a constant area, which can be called hot-spots [16]. However, the urban operating vehicles have public and specific attributes. It is necessary to cluster trajectory of operating vehicles to explore whether there are hot-spots. Density-Based Spatial Clustering of Applications with Noise (DBSCAN) can divide the area with enough high density into groups, and find clusters of arbitrary shape in noisy spatial database, which can be applied to the big raw GPS data of operating vehicles [17].

3.3 Evaluation of Long-Tails with Cumulative Distribution Function

The head and tail are two statistical terms, where the head is a protruding part in the middle of normal curve and the tail is a relatively flat part on both sides. From the perspective of human mobility, most of the daily trip will focus on the head, which can be called popular, while the demand distributed in the long-tail is personalized, scattered and small.

In order to evaluate the long-tail effects of vehicle trajectories, the cumulative distribution function is introduced as follows.

Algorithm 2. DBSCAN procedure

get the set of raw GPS data $\boldsymbol{L}$
get the radius e and minimum points $MinPts$
for all $\boldsymbol{L}_i \in \boldsymbol{L}$ **do**
  if $\boldsymbol{L}_i$ is the core point **then**
    find all the objects that can reach the density from this point and form a cluster
  end if
end for
obtain the clusters with each center $Coordinate_{x,y}^{hot-spots}$

Theorem 1. *For all the discrete distance from the generated set* $\boldsymbol{D}$*, the cumulative distribution function is defined as the sum of occurrence probability of all values less than or equal to the specific distance* d*.*

$$F_{\boldsymbol{D}}(d) = P(\boldsymbol{D} \leq d) \tag{1}$$

Theorem 2. *The generated set* $\boldsymbol{D}$ *is defined as the time series of travel distance from vehicles with the most hot-spots as the center. Therefore, the distance* d_i *in the sampling time* t_i *as an element of* $\boldsymbol{D}$ *is calculated as follows.*

$$d_i = \|Coordinate_{x,y}^{t_i} - Coordinate_{x,y}^{hot-spots}\| \tag{2}$$

4 Performance Evaluation

Framework of Trajectory Data Mining. The basic data mining can be divided into two parts, which are data storage and calculation. For the big data that cannot dealing with the conventional tools, the MongoDB database and the SPARK calculation are selected in this paper respectively. Those two open source tools shall be a close combination with the Python language. As it is shown in Fig. 2, the MongoDB takes responsible for data compression, index, and storage, while the SPARK is in charge of the data preprocessing and mining of the trajectory features.

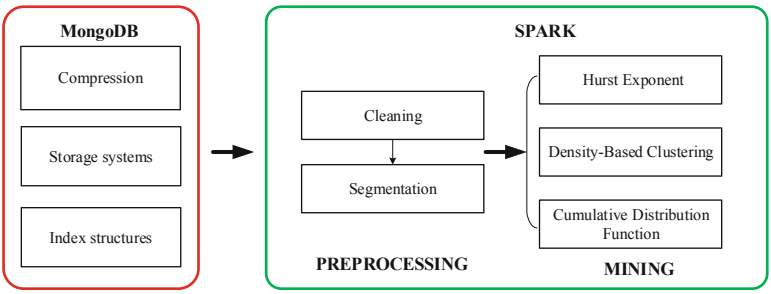

Fig. 2. The framework of trajectory data mining.

Self-Similarity. The BD procedure is adopted to calculate the Hurst value of the trajectories between the vehicles with blue and yellow license plate. In Fig. 3, it can be seen that the vehicle with the blue license plate has the lower value of Hurst, while the yellow vehicle has a litter higher one. From the definition of the Hurst parameter, the blue license plate that stands for the cabs is random roaming across the city, and the yellow license plate that stands for the trucks has a self-similarity path of moving.

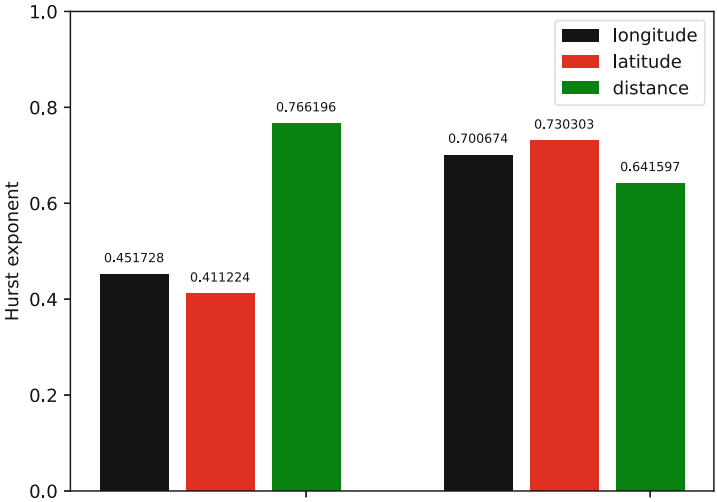

Fig. 3. The Hurst values between the vehicles with blue and yellow license plate respectively.

Hot-Spots. The DBSCAN procedure with the setup parameters listed in Table 1 is adopted to search for the clusters of the specific trajectory. Figure 4 that stands for the cabs and Fig. 5 that stands for the trucks has show the locations in one day for a specific type of the vehicle. After the DBSCAN procedure, the graph is colored based on the clustering results. Therefore, it is seen that the caps always roam in several certain hot-spots, while the trucks is always moving by a specific path.

Table 1. Data clustering for DBSCAN parameters.

Parameter	Value	Unit	Illustration
e	1000	m	Radius
$MinPts$	10	Null	Minimum points

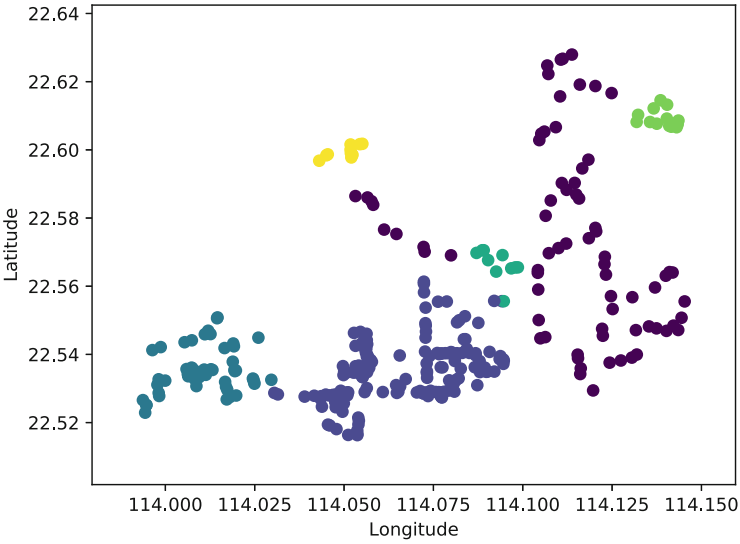

Fig. 4. Clustering results of locations by the caps activities in a specific day.

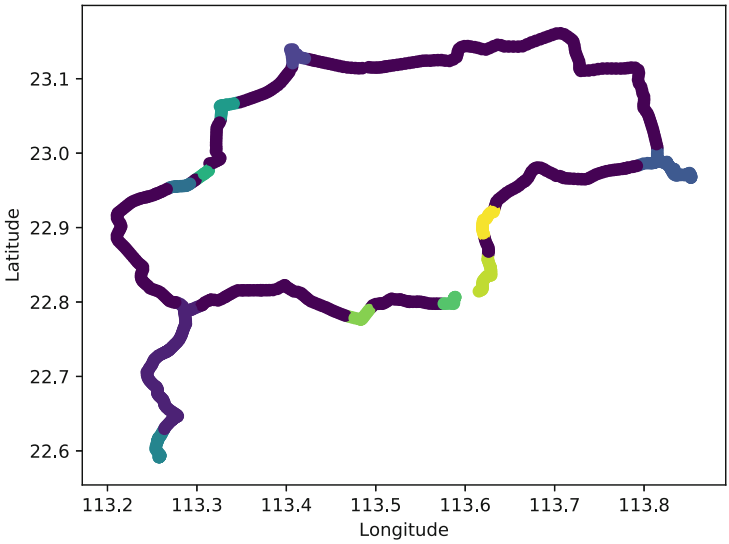

Fig. 5. Clustering results of locations by the trucks activities in a specific day.

Long-Tails. The cumulative distribution function is introduced in order to evaluate the long-tail effects of vehicle trajectories. Figure 6 shows that cabs in the city always take a short journey, while the distance that the trucks move is basically steady.

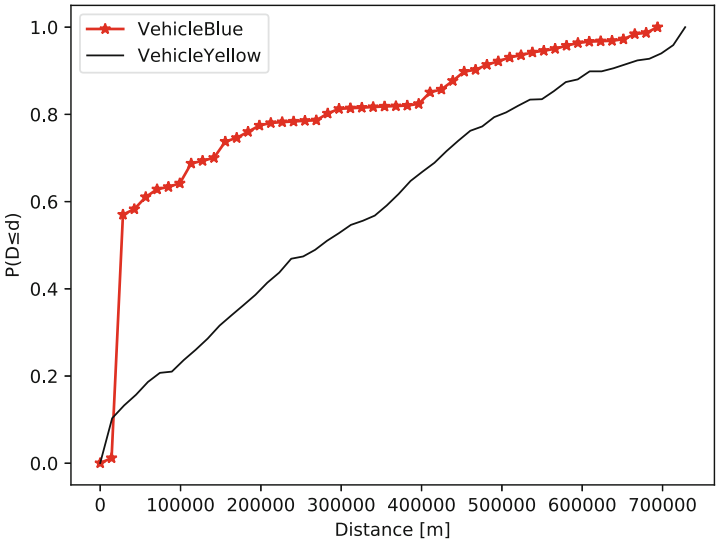

Fig. 6. The curve of the cumulative distribution function for the vehicles of cabs and trucks respectively.

5 Conclusion

In this paper, a data mining of the raw GPS trajectories is presented from the urban operating vehicles in the city of Shenzhen. The self-similarity, hot-spots and long-tails are evaluated as the widely recognized three features of human traces. Mining results show that the cabs which represent to serve the daily trip of human in the city always take a short travel and activate in several hot-spots, but roaming randomly. However, the trucks which represent to serve the goods are showing the opposite characteristics to cabs. Therefore, the vehicle types are the key feature of consideration to optimize algorithms for radio resource management in HDN.

References

1. Ando, S., Suzuki, E.: Role-behavior analysis from trajectory data by cross-domain learning (2012)
2. Chen, Y., Li, R., Zhao, Z., Zhang, H.: Fundamentals on base stations in urban cellular networks: from the perspective of algebraic topology. IEEE Wirel. Commun. Lett. **8**(2), 612–615 (2019). https://doi.org/10.1109/LWC.2018.2889041

3. Chen, Z., Shen, H.T., Zhou, X.: Discovering popular routes from trajectories. In: ICDE **6791**(9), 900–911 (2011)
4. Dai, J., Yang, B., Guo, C., Ding, Z.: Personalized route recommendation using big trajectory data (2015)
5. Feng, Z., Zhu, Y.: A survey on trajectory data mining: techniques and applications. IEEE Access **4**, 2056–2067 (2016). https://doi.org/10.1109/ACCESS.2016.2553681
6. Gonzalez, M.C., Hidalgo, C.A., Barabasi, A.L.: Understanding individual human mobility patterns. Nature **453**, 779–782 (2008)
7. Lee, K., Hong, S., Kim, S.J., Rhee, I., Chong, S.: Slaw: self-similar least-action human walk. IEEE/ACM Trans. Network. **20**(2), 515–529 (2012). https://doi.org/10.1109/TNET.2011.2172984
8. Li, Y., Luo, J., Chow, C.Y., Chan, K.L., Ding, Y., Zhang, F.: Growing the charging station network for electric vehicles with trajectory data analytics. In: IEEE 31st International Conference on Data Engineering, pp. 1376–1387. IEEE (2015)
9. Liu, X., Biagioni, J., Eriksson, J., Wang, Y., Forman, G., Zhu, Y.: Mining large-scale, sparse GPS traces for map inference: comparison of approaches. In: Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 669–677. ACM (2012)
10. Liu, Y., et al.: Exploiting heterogeneous human mobility patterns for intelligent bus routing (2015)
11. Lu, E.H.C., Lee, W.C., Tseng, V.S.: Mining fastest path from trajectories with multiple destinations in road networks. Knowl. Inf. Syst. **29**(1), 25–53 (2011)
12. Luo, W., Tan, H., Lei, C., Ni, L.M.: Finding time period-based most frequent path in big trajectory data. In: Proceedings of the 2013 ACM SIGMOD International Conference on Management of Data (2013)
13. Lv, Q., Qiao, Y., Ansari, N., Liu, J., Yang, J.: Big data driven hidden Markov model based individual mobility prediction at points of interest. IEEE Trans. Veh. Tech. **66**(6), 5204–5216 (2017). https://doi.org/10.1109/TVT.2016.2611654
14. McGuire, M.P., Janeja, V.P., Gangopadhyay, A.: Mining trajectories of moving dynamic spatio-temporal regions in sensor datasets. Data Min. Knowl. Discov. **28**(4), 961–1003 (2014)
15. Renso, C., Baglioni, M., de Macedo, J.A.F., Trasarti, R., Wachowicz, M.: How you move reveals who you are: understanding human behavior by analyzing trajectory data. Knowl. Inf. Syst. **37**(2), 331–362 (2013)
16. Rhee, I., Shin, M., Hong, S., Lee, K., Kim, S.J., Chong, S.: On the levy-walk nature of human mobility. IEEE/ACM Trans. Network. **19**(3), 630–643 (2011). https://doi.org/10.1109/TNET.2011.2120618
17. Wang, X., Huang, D.: A novel density-based clustering framework by using level set method. IEEE Trans. Knowl. Data Eng. **21**(11), 1515–1531 (2009). https://doi.org/10.1109/TKDE.2009.21
18. Wei, L.Y., Zheng, Y., Peng, W.C.: Constructing popular routes from uncertain trajectories (2012)
19. Xuan, S., Zhang, Q., Sekimoto, Y., Shibasaki, R.: Prediction of human emergency behavior and their mobility following large-scale disaster. ACM (2014)
20. Yang, B., Fantini, N., Jensen, C.S.: iPark: identifying parking spaces from trajectories (2013)
21. Yuan, N.J., Zheng, Y., Xie, X., Wang, Y., Zheng, K., Xiong, H.: Discovering urban functional zones using latent activity trajectories. IEEE Trans. Knowl. Data Eng. **27**(3), 712–725 (2015). https://doi.org/10.1109/TKDE.2014.2345405

22. Zhang, H., Huang, S., Jiang, C., Long, K., Leung, V.C.M., Poor, H.V.: Energy efficient user association and power allocation in millimeter-wave-based ultra dense networks with energy harvesting base stations. IEEE J. Sel. Areas Commun. **35**(9), 1936–1947 (2017). https://doi.org/10.1109/JSAC.2017.2720898
23. Zhang, H., Liu, H., Cheng, J., Leung, V.C.M.: Downlink energy efficiency of power allocation and wireless backhaul bandwidth allocation in heterogeneous small cell networks. IEEE Trans. Commun. **66**(4), 1705–1716 (2018). https://doi.org/10.1109/TCOMM.2017.2763623
24. Zhang, H., Liu, N., Chu, X., Long, K., Aghvami, A., Leung, V.C.M.: Network slicing based 5G and future mobile networks: mobility, resource management, and challenges. IEEE Commun. Mag. **55**(8), 138–145 (2017). https://doi.org/10.1109/MCOM.2017.1600940