Unlicensed Taxis Detection Service based on Large-Scale Vehicles Mobility Data

Yujie Wang[1,2], Xiaoliang Fan[1,2*], Xiao Liu[3], Chuanpan Zheng[1], Longbiao Chen[1], Cheng Wang[1], Jonathan Li[1]
[1] Fujian Key Laboratory of Sensing and Computing for Smart City, Xiamen University, Xiamen, China
[2] School of Information Science and Engineering, Lanzhou University, Lanzhou, China
[3] School of Information Technology, Deakin University, Melbourne, Australia
Email: {wangyujie15, fanxiaoliang}@lzu.edu.cn, xiao.liu@deakin.edu.au, zhengchuanpan@stu.xmu.edu.cn, {longbiaochen, cwang, junli}@xmu.edu.cn

***Abstract*—Unlicensed taxis are widely considered as major obstacles to city traffic regulation and public safety. Thus, many governments have issued restrictions for car-hailing services and alleged that the use of unlicensed vehicles was illegal. However, it is very challenging that traffic administrative enforcements face limited manpower to prohibit unlicensed taxis, due to costly and time-consuming procedure of on-site evidence collection. In this paper, we propose an effective service to incorporate human mobility mechanism into unlicensed taxis detection from massive city-wide vehicles. We first extract 276 spatio-temporal features, which are grouped into two categories, including daily behaviors and sustainable behaviors to capture the mobility characteristics of unlicensed taxis. Second, we investigate the detection accuracy of three machine learning techniques, viz. support vector machines, decision tree, and logical regression. We illustrate our approach using real-world vehicle license plate recognition dataset in Xiamen, China, which contains 336 million passing records for 6.2 million vehicles filmed by 439 devices in August 2016. Experimental results reveal that LR outperforms SVM and DT in prediction accuracy and F-score measurement, while SVM is capable of identifying the largest number of unlicensed taxis.**

***Keywords*—unlicensed taxi, anomaly detection, human mobility, big data analytics.**

I. INTRODUCTION

Unlicensed taxis (UT) are private vehicles that illegally ferry passengers for the commercial purpose without a taxi license [1]. Nowadays, many metropolitan cities have considered unlicensed taxis as major obstacles to city traffic regulation and public safety[1]. Especially, UT activities are harmful to many passengers due to issues, such as overcharging price by taking unnecessary detours, after-accident risk with inadequate liability insurance, and potential risk victims of crimes. Therefore, many governments have issued detailed restrictions for car-hailing services and alleged that the use of both unlicensed and crowd-sourced vehicles was unsafe and illegal[2].

Example: Reported from Shanghai Municipal Transportation Commission, 7,578 unlicensed taxis were caught and seized in 2015 in Shanghai, China, with year-on-year growth of 30.27%[3].

In fact, the phenomenon of UT operations is not solely existing in China. Big cities such as New York City, Paris, and Abu Dhabi, are reported to crack down illegal drivers who could face a large number of fines, up to 90-day in jail, or both[4]. However, due to the costly and time-consuming procedure of collecting evidences from massive regular private vehicles on-site, it is quite challenging for the traffic administrative enforcement department with limited manpower to prohibit UT.

With the advancement of sensing and location-aware technologies, human mobility studies try to employ a data-driven approach, instead of manual efforts, to solve such kind of outlier detection problem [5]. In fact, it is possible to identify UT using large-scale vehicle mobility records because human behavioral patterns can be discovered in their mobility footprints. Examples of such suspicious behaviors include, driving daily for extraordinarily many trips, making frequent stops at transportation hubs and hot scenic spots, and being extremely active in both weekends and late nights, and so on.

Massive human mobility data, such as call details records [6], GPS trajectories [7], and social network check-ins [8], provide the possibility to uncover mobility patterns to separate suspicious unlicensed taxis from regular private vehicles. More recently, a novel dataset, vehicle license plate recognition (VLPR) data, are considered as an emerging analytics tool for traffic law enforcement [9]. However, there are inherent complexities involved in unlicensed taxis activities that it is a non-trivial task for detecting such activities from large-scale VLPR records generated from millions of regular private vehicles. Specially, we need to address the following research challenges. The first challenge is city-scale detection. Identifying UT from massive private vehicles is computationally intensive. Second, it is difficult to classify UT behaviors. In particular, private vehicles may be occasionally used as unlicensed taxis in peak hours or weekends to make profits. For example, Fig. 1 shows that the difference between a known unlicensed taxi (i.e., *b*) and several regular commuting vehicles (i.e., *a, c, d,* and *e*) in morning peak is quite small.

Therefore, an efficient feature extraction model is needed to uncover the UT mobility pattern. In this paper, we propose an effective service which incorporates human mobility

* Correspondence author. This work was done when the first author was an intern at Xiamen University.

[1] Residents warned against unlicensed taxi services, The free press, May 18, 2016, http://www.thefreepress.ca/news/379964501.html

[2] NYC Cracks Down on Unlicensed Airport Taxis, The Voice of the Orthodox, November 24, 2009 http://www.vosizneias.com/43699/2009/11/24/new-york-city-nyc-cracks-down-on-unlicensed-airport-taxis/

[3] Shanghai Traffic Law Enforcement made comprehensive regulation on unlicensed taxis, Ministry of Transport of the People's Republic of China, http://www.moc.gov.cn/difangxinwen/xxlb_fabu/fbpd_shanghai/201601/t20160120_1979141.html

[4] 50 unlicensed Abu Dhabi taxi drivers arrested, The national, August 23, 2016, http://www.thenational.ae/uae/transport/50-unlicensed-abu-dhabi-taxi-drivers-arrested

978-1-5386-0752-7/17 $31.00 © 2017 IEEE
DOI 10.1109/ICWS.2017.106

mechanism into unlicensed taxis detection from massive city-wide vehicles. We first extract 276 spatio-temporal features which are grouped into two categories, including daily behaviors and sustainable behaviors to capture the mobility characteristics of UT. Second, we investigate the detection accuracy of three machine learning techniques, viz. support vector machines (SVM), decision tree (DT), and logical regression (LR). We illustrate our approach using real-world VLPR dataset [5], which contains nearly 336 million passing records generated by 6.2 million all kinds of vehicles and filmed by 439 VLPR devices in Xiamen, China, in August 2016. Experimental results reveal that LR outperforms SVM and DT in prediction accuracy and F-score measurement, while SVM is capable of identifying the largest number of unlicensed taxis.

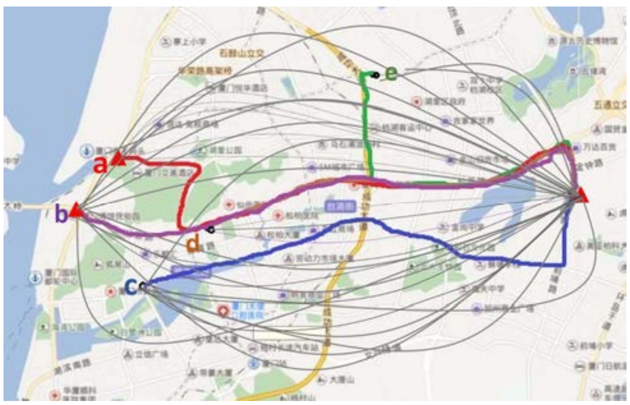

Figure 1. VLPR trajectories of unlicensed taxis and regular private vehicles

The remain of this paper is organized as follows. Section II summarizes related works. Section III discusses the preliminary and observations of datasets. Section IV proposes the algorithm. Section V describes the experiments and evaluations. Finally, Section VI concludes our work.

II. Related Works

Comprehensive works have been investigated to detect the abnormal behaviors based on urban sensing data in two categories: 1) location-based and 2) trajectory-based. First, to cope with location-based anomaly detection problem, Yuan et al. [2] presented a framework that could learn the contexts of different functional regions in a city with a feature extraction approach. Identification of such anomalies could enable the perception of abnormal activities. Thus, it could provide decision support for traffic management, such as black holes detection in traffic flows [3]. Second, the main objective of trajectory-based anomaly detection is to identify a small group of people whose movement traces are different from the ordinary people using data mining techniques. One typical example is to identify fraudulent taxi driving behaviors using the Dempster-Shafer theory [4].

Furthermore, it has been suggested that human mobility patterns follow a high degree of spatial and temporal regularity, and are thus highly predictable. Indeed, large-scale human mobility data, such as call details records [6], GPS trajectories [7], social network check-ins [8] and VLPR data [9], could provide the possibility of uncovering mobility patterns to separate suspicious unlicensed taxis from regular private vehicles. Specifically, VLPR records could be considered as a trajectory data if each passing record of a vehicle is linked sequentially. The advantage of VLPR dataset is the large-scale volume of records which could be used for city-wide UT detection. For instance, there are 500,000 private vehicles filmed by VLPR devices daily in Xiamen, compared with GPS trajectories which are generated only by 5,600 taxicabs.

For unlicensed taxis detection, Yuan et al. [10] presented a two-phase model to identify unlicensed taxis based on large-scale records generated from electronic traffic bayonet devices. However, the limitations of the proposed method in [10] include: (1) for feature extraction, they paid inadequate attentions on the sustainable patterns over time (i.e., one month) of UT; and (2) they employed only features related to passing records without taking into account mobility features for each vehicle such as travel distance and radius of gyration. To overcome these problems, in this paper, our approach incorporates a human mobility mechanism into both the daily behavior and sustainable behavior of UT, to characterize their abnormal mobility patterns from massive regular private vehicles in the city level.

III. Preliminary and Mobility Characteristics

A. Preliminary

Definition 1 *(Vehicle License Plate Recognition (VLPR) Device): A VLPR device* is a tuple D_i =<d_i,lo_i, la_i>, where d_i is a ID of the device, and lo_i and la_i refer to the longitude and latitude ofd_i, respectively.

Definition 2 *(Vehicle License Plate Recognition (VLPR) Record):* An VLPR Record is a record of plate numbers captured by VLPR Device. Each record can be denoted as $VR_{i,k}$ = *(*d_k*,* v_i*,* $time_{i,k}$*)*, where d_k is the ID of the VLPR device, v_i denotes the recognized vehicle plate number of the vehicle passing by d_k, and $time_{i,k}$ is the time stamp of the record.

Definition 3 *(Vehicle Trajectory):* A vehicle trajectory VT_{v_i} is represented by the sequence of VLPR records for a vehicle v_i, i.e., $VT_{v_i} = (VR_{i,1}, VR_{i,2}, \dots, VR_{i,l} \dots, VR_{i,m})$, where $VR_{i,l}$ is the *l*-th passing record of vehicle v_i.

Definition 4 *(Visited Location):* If the interval time between two consecutive VLPR records for a vehicle v_i is bigger than a threshold, a visited location is presented by $VL_{i,m}$, where $VL_{i,m}$ is the *m*-th visited location of vehicle v_i.

B. Problem Definition

Given a set of vehicles $= \{v_1, v_2, \dots, v_i, \dots v_n, 1 \le i \le n\}$, the unlicensed taxis detected problem is to design a binary classifier F to identify whether a vehicle is an unlicensed taxi (UT) or a regular private vehicle (RPV), formally defined as:

$$F(v_i) = \begin{cases} 1\ (UT) \\ -1\ (RPV) \end{cases} \qquad (1)$$

[5] Our study has been approved by the research ethnics board of Institute of Computer Architecture, Lanzhou University. The data is legally used without leaking any sensitive information.

C. VLPR Datasets

The VLPR sensors could recognize a vehicle's license plate number with advanced techniques such as image processing and pattern recognition [9]. Recently, VLPR sensors have been widely used in urban traffic applications, such as road traffic monitoring, traffic law enforcement, access control for parking lots, and so on. For example, there are over 439 VLPR sensors deployed in Xiamen, China, which is a coastal city with 3.8 million populations and 1.4 million local vehicles. Thus, hundreds of VLPR sensors deployed on city-wide road networks are capable of generating a large-scale mobility dataset, which depicts the traffic congestion, accident, as well as human moving patterns by vehicles.

There are two VLPR datasets, including VLPR records for passing vehicles, and VLPR information. Table I shows a sample of VLPR records. Specifically, a VLPR record contains attributes as follows: 1) Device ID, which is an identifier for each VLPR sensor; 2) driving direction of the vehicle; 3) the number of vehicle's license plate; 4) the color of vehicle's license plate; 5) lane number of the vehicle; and 6) a time stamp. Specifically, license plate numbers have been redacted for privacy- protection issues. Furthermore, the VLPR information is the location and type of the VLPR sensors, described by Device ID, Device Type, Device Location, Longitude, Latitude. An example of VLPR information is "620115, 0304, *North Hubin* Road and *Yuxiu* Road (East→West), 118.117, 24.488".

Table I. Sample of VLPR records for passing vehicles

Device ID	Direction	License plate No.	Color	Lane No.	Time Stamp
62068	2A	000001	2	2	2016-5-11-07:04:57
62143	2B	000002	2	1	2016-5-11-07:04:59
60201	2A	000003	2	2	2016-5-11-07:05:03

D. DB Mobility Characteristics

The daily behavior contains five features as follows.

1) ***Number of VLPR records***, is the total number of VLPR records for a vehicle v_i during the interval t, described as:

$$N_{VLPR,t} = |VT_{v_i,t}| \qquad (2)$$

2) ***Number of VLPR devices***, is the total number of vehicle v_i passing devices in time period t, which is described as:

$$N_{VLPRD,t} = (|VD_{v_i,t}|, VD_{v_i,t} \in VT_{v_i}) \qquad (3)$$

where $VD_{v_i,t}$ is the set of VLPR devices passed by vehicle v_i in a time period t.

As shown in Fig. 2(a), the observations are: 1) unlicensed taxis have both more VLPR records and passing devices than regular private vehicles; and 2) there is an interval of curve between UT's records and devices. This indicates that UT could visit a certain number of locations in a frequent pattern, because they pass several VLPR devices repeatedly.

3) ***Travel distance***, represents the distance a vehicle could drive within the time period t, described as:

$$TD_{i,t} = \sum_{l=1}^{l=m-1} distance(VR_{i,l+1,t}, VR_{i,l,t}) \;\; VR_{i,l,t} \in VT_{v_i,t} \qquad (4)$$

where $distance(VR_{i,l+1,t}, VR_{i,l,t})$ is the distance between the two consecutive VLPR records $VR_{i,l}$ and $VR_{i,l+1}$.

Fig. 2(b) shows the CDF (i.e., cumulative distribution function) curve of the average distance of unlicensed taxis and regular private vehicles during 6am~6pm of week days in a month. The average travel distance of UTs is longer than RPVs. Indeed, unlicensed taxis could drive a much longer distance because they are used by plenty of passengers.

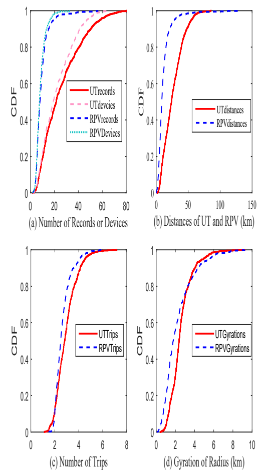

Figure 2. the CDF curve of (a) the average number of VLPR records and VLPR devices during 6am~6pm of week days in a month; (b) the average distance of unlicensed taxis and regular private vehicles during 6am~6pm of week days in a month; (c) trip number of unlicensed taxis and regular private vehicle; (d) radius of gyration for unlicensed taxis and regular private vehicle during 6am~6pm of week days in a month.

4) Trip number. In the Definition 4 of Section III.A, we have given the definition of visited location. We consider that the travel from $VL_{i,n}$ to $VL_{i,n+1}$ is a trip, presented by tri. Also, a vehicle trajectory VT_{v_i} in the time period t is represented by the sequence of tri for a vehicle v_i, i.e. $VT_{v_i,t} = TRI_{v_i,t} = \{tri_{i,1,t}, tri_{i,2,t}, \dots, tri_{i,l,t} \dots, tri_{i,m,t}\}$, where $tri_{i,l,t}$ is the *l*-th visited location of vehicle v_i in a time period t. The trip number is defined as:

$$N_{v_i,tri,t} = |TRI_{v_i,t}| \qquad (5)$$

Fig. 2(c) shows that the average trip number for unlicensed taxis is more than that of regular private vehicles.

5) Radius of gyration [6] is widely used by physicians to characterize the distance which a person would travel during a given time period t, which can be defined as:

$$r_g^{v_i}(t) = \sqrt{\frac{1}{n}\sum_{l=1}^{n}(r_l^{v_i} - r_{cm}^{v_i}(t))^2} \qquad (6)$$

where $r_l^{v_i}$ denotes the l-th location of the vehicle v_i, and $r_{cm}^{v_i}(t)$ is the center of mass that represents the centroid location of all passing records for v_i. Fig. 2(d) demonstrates that the average radius of gyration for unlicensed taxis is larger than that of regular private vehicles.

E. SB Mobility Characteristics

The sustainable behaviors are considered as long-term or periodical patterns of a vehicle in a time duration. We use mean, the standard deviation and the frequency of the daily features observed in a month (e.g., 31 days) for each vehicle to present their sustainable behaviors. Specifically, the standard deviation indicates the degree of variation of the daily behaviors of vehicles. For instance, regular private vehicles following routine trajectories are normally with less variations. In addition, the frequency indicates the activeness of each vehicle.

[6] Radius of gyration, Wikipedia, https://en.wikipedia.org/wiki/Radius_of_gyration

IV. The Proposed Solution

A. Framework

The general architecture of our system for detecting unlicensed taxis is presented in Fig. 3. In this section, we describe the input and output of the system, and the functional components, including active vehicle filtering, feature extraction and classification, as well as data flows connecting components.

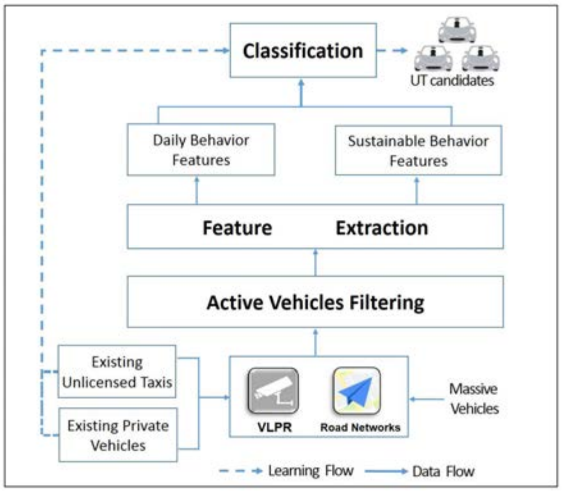

Figure 3. Overview of system architecture

B. Algorithm

In order to distinguish unlicensed taxi from a large number of private vehicles with high accuracy and low false-positive, we develop a three-step framework including active vehicles filtering, feature extraction and suspect detection.

Step 1. Active Vehicles Filtering.

First, we select private vehicles from massive VLPR records of all kinds of vehicles (i.e., 6.2 million vehicles in August 2016). Then, we selected VLPR records of private vehicles whose license plate color is blue since most unlicensed taxis are private vehicles. As a result, nearly 5 million private vehicles (including vehicles with both local and non-local license plate) are obtained.

Second, we extract active private vehicles based on the high frequency of vehicles in August 2016. We filter out those with lower than 10 active days, (here an active day means that a vehicle should have at least 9 VLPR records). Then, we obtained 295,674 active vehicles from all private vehicles.

Step 2. Feature Extraction.

According to two categories of features mentioned in Section III.B (viz. daily behavior and sustainable behavior), we extract six features before training of data. First, for daily behaviors, we take into account five features (viz. number of VLPR records, number of passing devices, travel distance, trip number and radius of gyration). More specifically, we extract different features in different time periods as described in Table 1. As a result, daily behavior of each vehicle has $23 * 4 * 2 + 8 * 4 + 31 = 247$ daily features.

Second, for sustainable behavior, we consider other three features (viz. the mean, standard deviation, and frequency of vehicle appearance) for each vehicle, representing the sustainability of the vehicle behavior in a month. As a result, each vehicle has $4 * 2 * 3 + 2 + 3 = 29$ sustainable features. In summary, as shown in Table II, the total number of spatio-temporal features for each vehicle is $247 + 29 = 276$.

TABLE II. Time period of features

Features		**Time periods**	**No. of Features**
DB Features	Number of records; number of devices; travel distances; radius of gyration	6am~6pm in weekdays (23); 6 pm~6am in weekdays (23); and holidays (8)	4*23*2+4*8 = 216
	Trip number	Every day (31)	31
SB Features	Mean and standard deviation	6am~6pm in weekdays (23); 6 pm~6am in weekdays (23); and holidays (8)	4*2*3+2 = 26
	Frequency of vehicle appearance	Weekdays morning, weekdays evening and holidays	3

Step 3. Suspect Detection

We adopt the binary classification methodology that are capable of incorporating heterogeneous features to effectively detect the unlicensed taxis. Specifically, we first train a supervised classifier model with known unlicensed taxis, and then detect unlicensed taxi from candidates using this model. The whole procedure is presented as below.

Algorithm: Unlicensed taxi detection

Input: **V**: the vehicles dataset; **Dataset:** VLPR records dataset; **SD**: sample dataset
Output: $F(v_i)$: the type of vehicle v_i in **V**
for each vehicle v_i in **V do**
  if (plate color = = blue && number of active days of vehicle v_i>10;) // Active Vehicles Filtering
    $v_i \in A_{cv}$ // A_{cv} is the set of Active Vehicles
    AF(v_i)=features (v_i) //extracted features of vehicle v_i; **AF(v_i)** is all features of vehicles v_i;
  end
end
train classifier F using **SD**;
  for each vehicle v_i in active vehicles set A_{cv}
    calculate $F(v_i)$ using **AF(v_i)**;
    if $F(v_i)$=1
      vehicles v_i is UT;
    else
      vehicles v_i is RPV;
  end
end

V. Experiments and Evaluations

In this section, we present experiments and evaluations with our proposed framework. First, we describe the experimental environment and the detail of datasets. Then we demonstrate the effectiveness of our framework.

A. Experimental Settings

We conduct our experiments on real-world datasets containing nearly 336 million passing records generated by 6.2 million vehicles in August 2016. The records are generated by 439 VLPR devices installed in the city. In addition, we use 800 known unlicensed taxis[7] as positive examples and 3200 private

[7] The ground truth data of 800 unlicensed taxis was retrieved from the Xiamen traffic administrative enforcement website, http://www.xmjtzf.com/

vehicles as negative examples to train the model. To evaluate the performance of the proposed model, we split all examples into training set and test set. The simulation and experiment are developed with MATLAB 2014b and JDK1.7, and conducted on an ASUS K55V PC with Windows 7 operating system, 32 GB RAM and Intel Core I7 3.6 GHZ CPU.

B. Evaluation Metrics

We investigate the detection accuracy of three machine learning techniques, viz. support vector machines (SVM), decision tree (DT), and logical regression (LR). For all the methods with parameters, we optimize the parameters with 5-fold cross-validation by dividing the training set into 80% for model fitting and 20% for parameter validation. In addition, we repeat the experiment five times and report the average results.

We use accuracy, precision, recall, and F-score as metrics to evaluate the performance of our method. Specially, accuracy is defined as the number of correctly identified instances divided by the number of identified test instances; precision is defined as the number of correctly identified positives divided by the number of identified positives instances; recall is defined as the number of correctly identified positives divided by the number of all positive instances in the test set; and the F-score is defined as:

$$\text{F-score} = 2 \times \frac{precision \times recall}{precision + recall} \qquad (7)$$

C. Results Summary

Table III summarizes the performance of these methods and the baselines. The experimental results show that: (1) LR outperforms SVM and DT in prediction accuracy and F-score measurement.

TABLE III. A performance comparison.

	SVM	DT	LR
Accuracy (%)	88.36	87.18	89.71
Precision (%)	76.13	67.70	77.45
Recall (%)	59.72	68.72	68.70
F1 (%)	66.87	68.12	72.72

D. Features Analysis

To further study the effectiveness of the features, we evaluate the performance of our framework with different feature categories. As shown in Fig. 4, we use DB and SB to represent features of daily behaviors and sustainable behaviors, respectively. The results show that the precision, recall and F-score are improved when DB and SB are used together. Such improvement demonstrates the effectiveness of sustainable behaviors features.

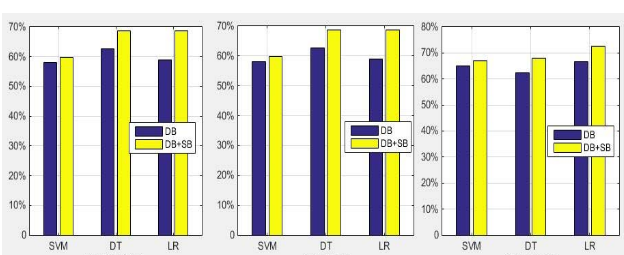

Figure 4. The impact of different feature categories

VI. Conclusion

This paper proposed a service to detect unlicensed taxis (UT) from a large number of vehicles in the city level. The service can be further used to identify active UT activities in densely-populated areas. Specifically, we first extracted 276 spatio-temporal features to capture the mobility characteristics of UT. Second, we investigated the detection accuracy of three machine learning algorithms (viz. SVM, DT, and LR). Finally, we employed a large-scale Vehicle License Plate Recognition (VLPR) dataset for model training and validation. Experimental results on a real-world vehicles dataset demonstrated the effectiveness of our proposed approach. In the future, the learning phase of the model can be further enhanced. Furthermore, it is also interesting to investigate where and when to seize UTs to assist traffic administrative enforcement department, with the consideration of other heterogeneous datasets, such as demographics, point of interests and so on.

Acknowledgment

The work was supported by grants from the Natural Science Foundation of China (61300232, 61300042); the Gansu Provincial Science and Technology Support Program (1504WKCA087); the China Postdoc Foundation (2015M580564); and Fundamental Research Funds for the Central Universities (lzujbky-2015-100, lzujbky-2016-br04, ZK1074).

References

[1] Illegal taxicab operation, Wikipedia, https://en.wikipedia.org/wiki/Illegal_taxicab_operation

[2] J. Yuan, Y. Zheng, and X. Xie, "Discovering regions of different functions in a city using human mobility and POIs," in Proc. of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 12), ACM press, August, 2012, pp. 186–194, doi: 10.1145/2339530.2339561.

[3] L. Hong, Y. Zheng, D. Yung, J. Shang, and L. Zou. "Detecting urban black holes based on human mobility data," in Proc. of the 23rd SIGSPATIAL International Conference on Advances in Geographic Information Systems (SIGSpatial 15), ACM press, November 2015, article no. 35, doi: 10.1145/2820783.2820811.

[4] Y. Ge, H. Xiong, C. Liu, and Z. Zhou, "A taxi driving fraud detection system," in Proc. of IEEE 11th International Conference on Data Mining (ICDM 11), IEEE press, December 2011, pp. 181-190, doi: 10.1109/ICDM.2011.18.

[5] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection: A survey," ACM computing surveys, vol. 41, no. 3, 2009, Article No. 15, doi: 10.1145/1541880.1541882.

[6] K.S. Kung, K. Greco, S. Sobolevsky, and C. Ratti, "Exploring Universal Patterns in Human Home-Work Commuting from Mobile Phone Data," PLoS ONE, vol. 9, no. 6, pp. e96180, doi: 10.1371/journal.pone.0096180.

[7] L. Gong, X. Liu, L. Wu, and Y. Liu, "Inferring trip purposes and uncovering travel patterns from taxi trajectory data," Cartography and Geographic Information Science, vol. 43, no. 2, 2016, pp. 103-114, doi: 10.1080/15230406.2015.1014424.

[8] G. Liu, Y. Wang, M.A. Orgun, X. Zheng, A. Liu, Z. Li, and K. Zheng, "Strong Social Component-Aware Trust Subnetwork Extraction in Contextual Social Networks," in Proc. IEEE International Conference on Web Services (ICWS 16), IEEE Press, June 2016, pp: 107-114, doi: 10.1109/ICWS.2016.22.

[9] Z. Zhang, G. Wang, B. Cao, and Y. Han, "Data Services for Carpooling Based on Large-Scale Traffic Data Analysis," Proc. of 2015 IEEE International Conference on Services Computing (SCC 15), IEEE Press, June 2015, pp. 672-679, doi: 10.1109/SCC.2015.96.

[10] W. Yuan, P. Deng, T. Taleb, J. Wan, and C. Bi, "An unlicensed taxi identification model based on big data analysis," IEEE Transactions on Intelligent Transportation Systems, vol. 17, no. 6, 2016, pp. 1703-1713, doi: 10.1109/TITS.2015.2498180.