Catch Me If You Can: Detecting Pickpocket Suspects from Large-Scale Transit Records

Bowen Du
State Key Lab of Software Development Environment
Beihang University
Beijing, China
dubowen@buaa.edu.cn

Chuanren Liu
Decision Sciences and MIS
LeBow College of Business
Drexel University
Philadelphia, US
chuanren.liu@drexel.edu

Wenjun Zhou
Business Analytics & Statistics
Haslam College of Business
University of Tennessee
Knoxville, US
wzhou4@utk.edu

Zhenshan Hou
State Key Lab of Software Development Environment
Beihang University
Beijing, China
zhh@buaa.edu.cn

Hui Xiong*
Management Science and Information Systems
Rutgers University
New Jersey, US
hxiong@rutgers.edu

ABSTRACT

Massive data collected by automated fare collection (AFC) systems provide opportunities for studying both personal traveling behaviors and collective mobility patterns in the urban area. Existing studies on the AFC data have primarily focused on identifying passengers' movement patterns. In this paper, however, we creatively leveraged such data for identifying thieves in the public transit systems. Indeed, stopping pickpockets in the public transit systems has been critical for improving passenger satisfaction and public safety. However, it is challenging to tell thieves from regular passengers in practice. To this end, we developed a suspect detection and surveillance system, which can identify pickpocket suspects based on their daily transit records. Specifically, we first extracted a number of features from each passenger's daily activities in the transit systems. Then, we took a two-step approach that exploits the strengths of unsupervised outlier detection and supervised classification models to identify thieves, who exhibit abnormal traveling behaviors. Experimental results demonstrated the effectiveness of our method. We also developed a prototype system with a user-friendly interface for the security personnel.

CCS Concepts

•Information systems → Spatial-temporal systems; Data mining; •Computing methodologies → Anomaly detection;

*Corresponding Author

KDD '16, August 13-17, 2016, San Francisco, CA, USA

DOI: http://dx.doi.org/10.1145/2939672.2939687

Keywords

Automated Fare Collection; Travel Behaviors; Mobility Patterns; Public Safety; Anomaly Detection.

1. INTRODUCTION

Passengers in the public transit systems have been the main target for pickpockets. In many cities, thefts happen frequently in public transit systems, because passengers tend to pay less attention to their belongings when they are in a rush or in a crowded environment. For example, during the first nine months of 2014, 350 pickpockets were caught in the subway system and 490 were caught on buses[1] in Beijing, China. Many other big cities in the world, such as Barcelona, Prague, Rome, and Paris, are also reported to suffer from the pickpocket problem [2], which has led to public safety concerns [23, 7]. Indeed, it is challenging to detect theft activities committed by cunning thieves who know how to escape without being disclosed. Despite the substantial cost in manpower and resources, many thieves are still at large. It is critical to provide a smart surveillance and tracking tool for the security personnel of the transit systems.

With rapid advances in information technology and data processing capacities, transactional records collected by automated fare collection (AFC) systems [20] have become valuable for understanding passengers' mobility patterns and the urban dynamics [6, 3, 20, 29, 18]. However, most of the existing studies focused on identifying regular, collective mobility patterns, such as commute flows and transit networks. Our study is the first to focus on identifying thieves based on AFC data. In fact, it is possible to detect thieves using AFC records because behavioral differences are coined in the mobility footprints, which can help to separate suspects from regular passengers. Examples of such behaviors, which can make suspects notable, include traveling for an extended length of time, making unnecessary transfers, and/or wandering on certain routes while making random stops.

[1] http://www.bjgaj.gov.cn/web/detail_zxftDetail_397242.html

[2] http://abcnews.go.com/Travel/top-10-pickpocket-cities-watch-wallet-avoid-thieves/story?id=11769828

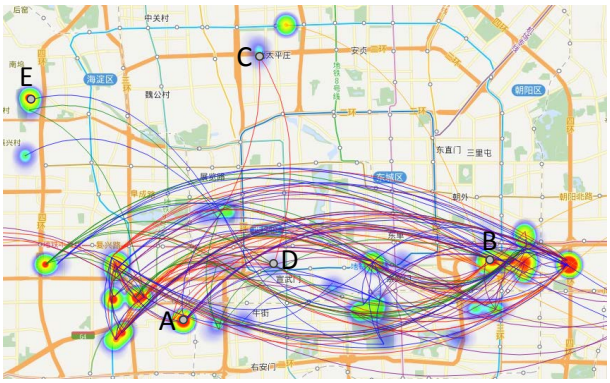


Figure 1: Trajectories of passengers.

However, detecting thieves based on AFC records is not a simple outlier detection problem. Figure 1 shows the difference between a known thief and an outlier. We can see a number of trajectories between hot regions A and B. By careful examination, we can see that most passengers move from one region to another using near-optimal configuration (*e.g.*, shortest time/distance, or a minimal number of transfers). However, a passenger (a known suspect) who took the path $A \rightarrow C \rightarrow D \rightarrow B$ looks suspicious because there is no need to make transfers at C and D in order to reach B. Based on the above observation, passengers who exhibit such abnormal behaviors will be selected for further examination. In contrast, another passenger who travels from E to B is an outlier, since few passengers take the same path. However, this passenger is likely just a regular passenger who originates from a less crowded area.

In summary, to identify thieves from AFC records, we are faced with a number of inherent challenges.

- The first challenge is how to identify useful features to distinguish thieves from regular passengers. These features should not only help us understand the behaviors of pickpockets, but also help us build a suspect detection and tracking system for supporting the security personnel.
- Second, using regular outlier detection methods tends to result in a large number of false positives. In particular, not every trip made by a regular passenger looks normal. Regular commuters may occasionally make trips to visit friends or places of interest, and some of such trips may look suspicious by how much they deviate from regular behaviors.
- Third, a large number of AFC records are being collected from millions of passengers, only a tiny fraction of which are pickpockets. Identifying such a small group of people in such a large-scale dataset is like looking for a needle in the haystack.
- Finally, we also need to effectively transform our knowledge based on model development into a decision support system, so that real-time, personalized deployment recommendations could be made to help to guide security personnel to perform their work more efficiently.

To this end, in this paper, a comprehensive approach is taken to meet the above challenges. Specifically, we first construct a feature representation for profiling passengers. Furthermore, we establish a two-step framework to separate normal movement patterns from irregular behaviors, and eventually, distinguish thieves from regular passengers. Finally, we leverage real-world datasets from multiple sources for model training and validation, and implement a prototype system for end users.

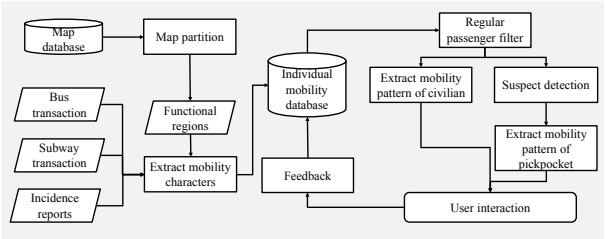


Figure 2: The framework.

Figure 2 shows the overall architecture of our framework. We first partition the city area into regions with functional categories. Then, the mobility characteristics of passengers are extracted from transit records and incident reports. Moreover, we build an individual mobility database to store the profile of each passenger. Next, we implement our framework by regular passenger filtering and suspect detection. The system is efficient and interactive, with both mobile and desktop clients. Finally, the user feedback information, such as newly confirmed thieves, will be entered as ground truth for future model training.

The remaining of this paper is organized as follows. Section 2 provides an overview the AFC datasets, based on which we performed the study. A detailed description of features that we extract to characterize mobility profiles of passengers is presented in Section 3. A two-step framework of the suspect identification system is proposed in Section 4. Experimental results are summarized in Section 5, and an overview of the deployed system is presented in Section 6. Finally, we summarize related work in Section 7, and draw conclusions in Section 8.

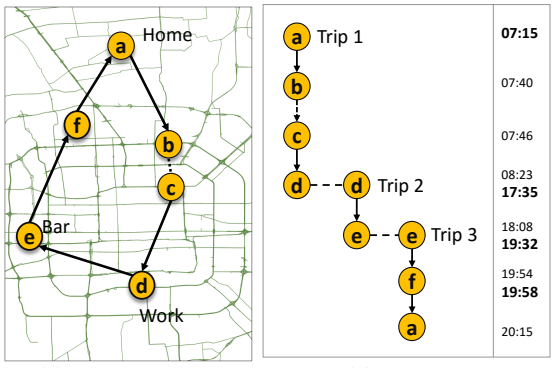
2. DATA DESCRIPTION

The data for our study have been collected from multiple sources. These include transit records, geographical information, and theft incident reports. In this section, we provide an overview of the data.

2.1 Transit Records

Our study is based on a large-scale transit records dataset collected from a public transit system that includes buses and subways. Passengers utilizing the transit service are charged by the distance they travel. A rechargeable smart card is issued to each passenger, who has to swipe the card when they board or exit a vehicle. The AFC system then calculates the fare according to the stations of boarding and exiting. As a result, each raw AFC record consists of the smart card ID, the route number, the event (*i.e.*, boarding or exiting), the station, and the time stamp. We transformed the data so that each transit record consists of one boarding and one exiting event of the same ID. There are about 1.7 billion records between April and June in 2014. After removing replicates and extremely infrequent riders, we are left with over 1.6 billion records that involve approximately 6 million passengers (for three months).

In order to describe the data and subsequent feature extraction process clearly, here we clarify two concepts, *transit records* and *trips*, using a concrete example. Figure 3 illustrates an example passenger's activities in a typical day. Part (a) is the actual trajectory on the city's map; Part (b) splits the trajectory into three separate trips; and Part (c) demonstrates the corresponding transit records in our data. In particular, let us say that the passenger holds a smart card with an ID of 4322. He started by taking Route 52 at Station a, which is next to his home, at 7:15 a.m. Having to make a transfer, he exited Route 52 at Station b at 7:40 a.m., and walked across the street to take Route 26 at Station c (7:46 a.m.). Then he got off at Station d, which is next to his workplace, at 8:23 a.m. He completed Trip 1 since the next time he was in the transit system was more than 30 minutes (*i.e.*, our empirical cutoff) later. Similarly, the transit in the afternoon from d to e was considered Trip 2; and the transit from e back home at a, making a transfer at f, was considered Trip 3. As a result, we collected the five transit records that describe three trips.



(a) Example Activities (b) Trips

Smart Card ID	Route Number	Boarding Station	Boarding Time	Exiting Station	Exiting Time
4322	Route 52	a	07:15	b	07:40
4322	Route 26	c	07:46	d	08:23
4322	Route 11	d	17:35	e	18:08
4322	Route 11	e	19:32	f	19:54
4322	Route 16	f	19:58	a	20:15

(c) Transit Records

Figure 3: An example of trips and transit records.

Intuitively, a transit record corresponds to one segment of transit between a pair of consecutive boarding and exiting events of the same passenger. Even though this segment of transit may pass a number of stations, the passenger has never left the vehicle during the time. In contrast, a trip consists of one or more such segments, which connect places of interest on the two ends, where the passenger stays for extended periods of time. A trip may include connections or transfers, as long as those breaks are relatively short in time. Formally, we provide the following definitions.

DEFINITION 1 (TRANSIT RECORD). *A transit record tr contains the following information:*

- tr_{route}*: the bus/subway route number;*
- tr_{sboard}, tr_{tboard}*: the boarding station and time; and*
- tr_{sexit}, tr_{texit}*: the exiting station and time.*

As a result, for each transit record, we were able to compute the travel distance tr_{dist}, travel time tr_{time} and number of stops tr_{stops} during the transit.

DEFINITION 2 (TRIP). *A trip Tr is a sequence of transit records $Tr = (tr^1, tr^2, \cdots, tr^n)$, where the passenger's origin location is $Tr_{origin} = tr^1_{vboard}$ and the destination is $Tr_{dest} = tr^n_{sexit}$.*

In practice, we construct one trip record if and only if the time gap between two consecutive transit records is no more than 30 minutes. The trip's time duration is calculated as $Tr_{time} = tr^n_{texit} - tr^1_{tboard}$.

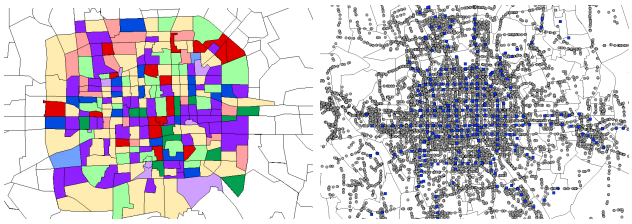
2.2 Geographical Information

To project the transit records to the city map, we made use of two datasets of important geographical information. Namely, the points of interests (POI) data and the public transit network information.

Table 1: Categories of Functional Regions

Category	Examples	Frequency
Home	Apartment buildings	28,731
Work	Government or office buildings	71,364
Education	Schools, training centers	3,527
Food	Restaurants and dining	56,906
Shopping	Shopping malls and outlets	24,310
Entertainment	Museums, theaters, clubs	18,223
Scenic Spot	Parks, sports fields	2,362
Transportation	Airports, transit centers	15,287
Healthcare	Hospitals, pharmacy	8,685
Car services	Car sales, repairs	1,781

With the POI data, we followed Yuan's work [28] to segment the urban area into small regions by major road networks, and then categorize each region into one of ten functions, as listed in Table 1. These regions are then color coded and visualized in Figure 4(a).



(a) Functional Regions (b) Stations

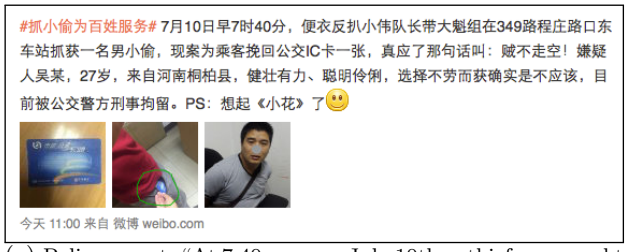
Figure 4: Geographical information.

The public transit network dataset provides basic information of each route of buses and subways. Each route is a sequence of stations, and the raw data include the route number, the sequence ID in the route, the name of the station, and the geo-location expressed as latitude and longitude. As shown in Figure 4(b), in total, we have 44,524 bus stations (points in gray) covered by 896 bus routes, and 320 subway stations (points in blue) covered by 18 subway routes. To remove the redundancy and better model the human mobility pattern, we merge stations located at the same road intersection.

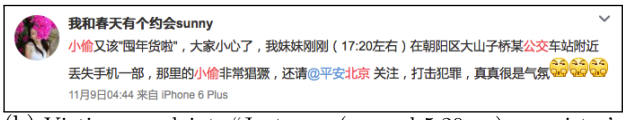
2.3 Incident Reports

Confirmed pickpocket incidents are publicly announced via Sina Weibo, the primary social network services in China.

It is considered public data since posts are all visible to everyone, just like Twitter in the United States. We included two types of pickpocket reports: official announcements, as announced by the police[3], and personal complaints, as posted by the victims. Figure 5 provides an example of each type of report. We can see that the date, time, and location of the theft events are normally identified in the posts, which has helped us to link such events to other sources of data. We found 10,529 records during our study period.



(a) Police report: "At 7:40 a.m. on July 10th, a thief was caught at Route 349 East Chengzhuanglukou Station."



(b) Victim complaint: "Just now (around 5:20pm), my sister's phone was stolen at the Dashanzi Bridge Bus Station."

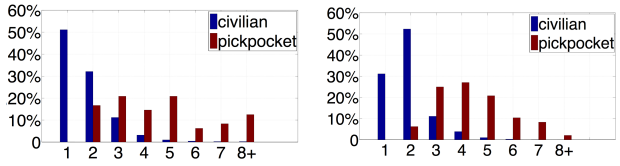
Figure 5: Example incident reports on Weibo.

3. MOBILITY CHARACTERISTICS

In this section, we will describe the features we extracted from passengers' AFC records, which are potentially useful for characterizing public transit mobility patters, and thus, will eventually be used for distinguishing pickpocket suspects from regular passengers. As shown in Table 2, our features are grouped into three categories: daily behaviors, social comparisons, and historical behaviors.

3.1 Travel Time and Frequency

The daily travel time is defined as the total duration spent by each passenger in the public transit system, and the daily riding frequency is defined as the number of transit records traveled by each passenger per day.



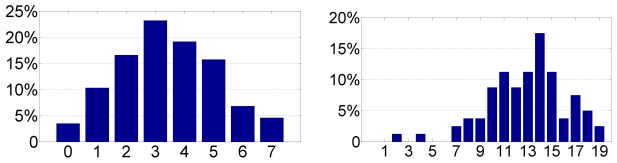
(a) Travel time. (b) Riding frequency.

Figure 6: Distributions of travel time and the number of transit records.

Indeed, picking pockets is a hard work: a thief has to spend quite long time in the crowded buses, subways, or near the transit stations to find potential victims and better crime moments. Also, in order to have more theft opportunities, a thief (pickpocket) tends to travel between bus/subway stations in random ways without specific destinations. Such abnormal behaviors lead to abnormal daily travel time and riding frequency. Figure 6 plots the distribution of daily travel time and riding frequency, respectively, of passengers traveling around the city with smart cards. We can see that more than 80% passengers finish their travels in 2 hours and within 2 transit records per day. In comparison, the identified thieves often spend more than 3 hours of daily travel time, and their daily riding frequency is also larger.

3.2 Short Rides

A short ride is a transit record tr with less than 3 stops. Regular passengers normally prefer fewer transfers in each trip. Therefore, if a transit has to be made, each transit record will pass several bus/subway stops. In contrast, pickpockets often switch routes within few stops to avoid attracting fellow passengers' attention and being recognized.



(a) (b)

Figure 7: Distributions of short distance trips.

Figure 7 are the distribution of the daily number of short rides for all passengers with at least 7 and 19 transit records, respectively. The x-axis is the number of short rides and the y-axis is the percentage of passengers. For the passengers with at least 7 transit records in Figure 7(a), the distribution approximates Gaussian with mean around 3. In Figure 7(b), for the passengers with at least 19 transit records, the distribution peak is shifted to the right. It shows that the frequency of short rides increases with increasing number of transit records.

3.3 Functional Transitions

A high-level view of the human mobility patterns can be summarized by transition among regions, where each region covers multiple stations. For example, the morning commuting trips can be abstracted as 'leaving from residence region, then transiting at transfer stations, and arriving at the workplace region at last'. Other examples include 'shopping trips' like `residence` → `shopping facilities` → `residence`, or 'sightseeing trips' like `residence` → `scenic spot` → `scenic spot` → `residence`.

Indeed, such sequential information is very useful for pattern discovery and predictive modeling [14, 12, 30]. To discriminate regular passengers and pickpockets, we observed that there are typical sequential patterns followed by the regular passengers. However, pickpockets tend not to follow the typical patterns, and wander randomly among the functional regions.

As a result, we define features such as the number of boarding stations and the number of boarding regions. Then, we count the transition frequency between any pair of function categories for the daily trip of each passenger.

[3] http://weibo.com/571100476

Table 2: List of extracted features.

Category	Feature Description	Mean	Median
Current	Travel time (hours)	1.25	0.976
Behaviors	Riding frequency	3.93	3
	Number of trips	2.38	2
	Number of short rides	0.81	0
	Number of boarding stations	3.72	3
	Number of regions	3.06	2
	Total number of functional transitions (See Subsection 3.3)	2.26	2
	Number of rides on the most frequent route	2.63	2
	Maximum number of visits of a functional region	2.06	2
	Number of wandering concentration spots (See Subsection 3.4)	2.17	2
Social	Time gap of trips (hours) (See Subsection 3.5)	0.49	0.43
Comparisons	Time gap of region transitions	0.37	0
Historical	Daily travel time, median	1.14	0.89
Behaviors	Daily riding frequency, median	3.71	3
(See Subsection 3.6)	Number of trips, median	2.55	2
	Number of short rides, median	0.83	0
	Number of boarding stations, median	2.89	2
	Number of regions, median	3.16	2
	Functional transitions, median	2.88	2
	Daily travel time(hour), standard deviation	0.36	0.35
	Daily riding frequency, standard deviation	0.79	0.57
	Number of trips, standard deviation	0.89	0.64
	Number of short rides, standard deviation	0.42	0.31
	Number of boarding stations, standard deviation	1.81	1.13
	Number of regions, standard deviation	1.43	0.92
	Functional transitions, standard deviation	1.28	1.17
	Number of days detected as suspect	0.00	0

3.4 Frequently Visited Regions

In practice, the majority of human mobility patterns are regular movements between a small set of locations that the passenger is familiar with. Pickpockets often spend a significant portion of the time within few routes or regions if they intend for opportunities. In particular, once a thief has committed the crime or lost the target, he or she would likely come back to a familiar station for the next target. Thieves know profoundly well about and wander around these areas. Therefore, we can measure such wandering behaviors by counting the maximum number of times a route was taken, or the maximum number of visits made to a region.

Moreover, the wandering behaviors lead to concentration of the passenger's boarding or exiting locations. We use the number of clusters as another feature, which measures the wandering concentration.

3.5 Deviation from the Social Norm

This group of features measures the difference between the individual behaviors and the typical behaviors of the population, so we call them social features. According to our empirical study, two very informative social features are the time gap of trips and the time gap of region transitions. Actually, given the same origin and destination, the trip variation of the majority of the population (i.e., regular passengers) is low. For example, most of the trips will be finished within a specific amount of time given the trip origin and destination, while pickpocket suspects may spend more time in the transit system during the trip.

Thus, for each pair of origin o and destination d, we find the travel time of all trips between this pair. We then convert the time gap significance of a given trip by the quantile of its travel time with respect to the population. Similarly, we also define the time gap significance of region transitions.

3.6 Historical Behaviors

We compute the statistics (e.g., median and standard deviation) of the daily features observed in the last seven days for each passenger, to quantify their historical behaviors. We use the median instead of the mean, because median is more robust in the presence of outliers, especially if the sample size is small. In our case, we only used data from the past seven days, and median can effectively avoid the impact of non-routine passenger behaviors. The standard deviation indicates the degree of variation of the daily behaviors of individual passengers. Regular passengers following routine trajectories normally generate statistics with less variations. Moreover, we also record the number of days (out of the recent seven days) when the passenger was detected as potential pickpocket suspects.

4. SUSPECT IDENTIFICATION

This section presents the key component of our thief footmark detection system. Specifically, in order to distinguish pickpockets from the regular passengers with high accuracy and low false-positives, we develop a two-step framework. The first step adopts anomaly detection techniques to identify thieves as well as some suspects from the passenger population. The second step further distinguishes the pickpockets from the suspects, by mapping the defined features into a high dimensional space and computing the optimal decision surface in support vector machines (SVM).

Suppose there are totally N passengers included in the dataset $\{(x_j, y_j) \mid j = 1, ..., N\}$, where $x_j \in \mathbb{R}^q$ is the extracted features associated with the j-th passenger, and $y_j \in \{0, 1\}$ is the classification label, such that $y_j = 1$ if and only if the passenger is a determined as a pickpocket suspect. Our objective is to develop a predictive model

$$f : x \mapsto y = f(x), \tag{1}$$

which can be used to identify pickpockets from the regular passenger population.

4.1 A Two-Step Framework

In this study, we develop a two-step framework for identifying pickpocket suspects. On one hand, since the dominant majority of the passengers are regular ones, it is non-trivial to use a classification algorithm. Specifically, the percentage of confirmed pickpockets is extremely low in the passenger population. Simple heuristics like over-sampling and under-sampling would only be helpful for handling moderate class imbalance, but not imbalance as extreme as ours. Building robust machine learning models for such unbalanced data is still an active research area in the literature [21, 4, 16, 17]. On the other hand, utilizing anomaly detection algorithms, which are typically unsupervised, not only cannot scale well, but also may lead to significant false-positives, since many regular passengers who occasionally perform irregular activities can be misclassified as suspects.

To address these challenges, we develop a two-step framework by unifying the unsupervised anomaly detection and supervised classification in a novel way. We show that the two steps can effectively utilize the supervised information, overcome the issue of unbalanced data distribution, and reinforce the learning performances. The overall framework can predict the pickpockets with very low false-positives. In the two-step framework, we approximate the predictive function $f(\cdot)$ as $f(x) = g(x)h(x)$, and equivalently:

$$f(x) = \begin{cases} 0 & \text{if } g(x) = 0 \\ h(x) & \text{if } g(x) = 1 \end{cases} \tag{2}$$

This relationship can also be demonstrated in Figure 8.

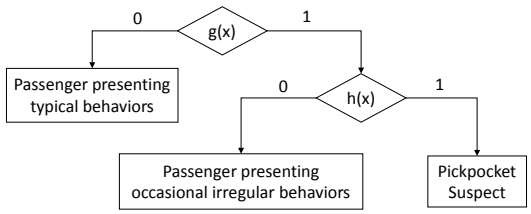


Figure 8: A two-step approach for suspect detection.

In other words, we first use function $g(\cdot)$ (the first step) to filter out regular passengers whose mobility patterns are typical in the majority of the passenger population. If the passenger associated with feature vector x is not filtered out (i.e., $g(x) = 1$), we then use function $h(\cdot)$ (the second step) to detect whether the passenger is a pickpocket suspect. In the following of this section, we develop the two steps (function $g(\cdot)$ and $h(\cdot)$) in our framework.

4.2 Regular Passenger Filtering

The first step in our framework is the regular passenger filtering by anomaly detection algorithm. Its objective is to exclude regular passengers without any suspicious behaviors from later modeling steps. Therefore, we intentionally allow some false-positives in the anomaly detection result.

Many general purpose anomaly detection algorithms can be used to implement the filtering function $g(\cdot)$. In this paper, we use the One-Class SVM (Support Vector Machine) [25] method due to its superior detection accuracy, computing efficiency, and modeling flexibility. One-Class SVM computes non-linear decision boundaries to detect outliers, using appropriate kernel functions and soft margins. The kernel function $\kappa(\cdot, \cdot)$ is defined as

$$\kappa(x_1, x_2) = \langle \phi(x_1), \phi(x_2) \rangle.$$

The function $\phi(\cdot)$ maps the original feature into a high-dimensional kernel space where the optimal decision boundary exists:

$$\widehat{g}(x) = \langle w, \phi(x) \rangle + \rho,$$

and then

$$g(x) = \begin{cases} 1 & \widehat{g}(x) \geq 0 \\ 0 & \widehat{g}(x) < 0 \end{cases} \tag{3}$$

The optimization objective of the One-Class SVM is to:

$$\min_{w,\rho} \frac{1}{2}\|w\|^2 + C\sum_{n=1}^{N} \epsilon_i - \rho$$

subject to $\widehat{g}(x_i) = \langle w, \phi(x_i) \rangle + \rho \leq \epsilon_i$ and $\epsilon_i \geq 0$ for all collected passengers $n = 1, 2, \cdots, N$. The parameter C controls the fraction of anomalies (e.g., x such that $g(x) = 1$) after the filtering.

It can be shown that the optimization process requires $\kappa(\cdot, \cdot)$ instead of an implicit formulation of $\phi(\cdot)$. We use the widely used Gaussian kernel:

$$\kappa(x_1, x_2) = e^{-\|x_1 - x_2\|^2/h}, \tag{4}$$

and learn the best parameter h (bandwidth) by cross-validation.

4.3 Suspect Detection

The second step in our framework is the suspect detection to eventually identify the suspect pickpockets. After the regular passengers filtered by the first step, now we have subjects including the real suspects as well as the false-positives not filtered out by the One-Class SVM. However, by controlling the parameter C of One-Class SVM, the number of the false-positives are limited and comparable with the number of suspects. Now the second step further distinguish these two subsets with supervised information verified by the social media.

Specifically, suppose there are in total M subjects after filtering in the dataset $\{(x_j, y_j) \mid j = 1, \cdots, M\}$, where for the training purpose we have $y_j = 1$ if and only if the passenger associated with feature vector x_j is a verified pickpocket, otherwise $y_j = 0$. Now, to train the suspect detection model, we again use the support vector machine (SVM), where we use the same feature mapping function $\phi(\cdot)$ and kernel function $\kappa(\cdot, \cdot)$ defined in Subsection 4.2. To classify the real suspects and false-positives, the optimal decision hyperplane is: $\widehat{h}(x) = \langle w, \phi(x) \rangle + \rho$ and then $h(x) = \begin{cases} 1 & \widehat{h}(x) \geq 0 \\ 0 & \widehat{h}(x) < 0 \end{cases}$. To compute the optimal w and ρ in $h(\cdot)$, we optimize:

$$\begin{aligned} \min_{w,\rho,\xi} \ & \frac{1}{2}\|w\|^2 + C\sum_{j=m}^{M} \xi_m \\ \text{s.t. } & \widehat{h}(x) \geq +1 - \xi_i, \forall y_j = 1 \\ & \widehat{h}(x) \leq -1 + \xi_i, \forall y_j = 0 \\ & \xi_j > 0, \forall j \end{aligned} \tag{5}$$

Again, C is trade-off parameter and the computing depends only on the kernel function define in Equation 4.

5. EXPERIMENTAL RESULTS

In this section, we present experimental results with our proposed framework. First, we describe the experimental environments and provide implementation details. Then, in comparison with several baseline methods, we demonstrate the effectiveness of our framework. We will also evaluate the usefulness of feature extracted in Table 2. Finally, we show practical usage of our methods in a real-world system.

5.1 Experiment Settings

Datasets. We conduct our experiments on real-world datasets containing over 1.6 billion transit records, as discussed in Section 2. We split the data into historical training set and evaluation test set. Specifically, training set covers three months (from April to June, 2014) and test set comes from the following two weeks (in July 2014). Also, from the training set, we filter out passengers whose maximum number of daily records is no more than three.

Platform. All experiments were conducted on a Windows Server 2012 64-bit system (4-CPU, each with 2.6GHz with Quad-Core, and 128G main memory). All algorithms and our real-world system were implemented with Java.

Baselines. The method is compared with a variety of competing methods grouped into the following categories:

Classification methods (CM). The classification methods, including logistic regression (LR), decision trees (DT), and support vector machines (SVM), are straightforwardly fitted with the training set and evaluated with the test set. Since the proportion of positive instances are extremely low, the classification problem is unbalanced and it's expected to observe high Type II Error. In experiments, we under-sample the negative instances to balance the data and improve the results. For each method, we repeat the sampling 10 times and report the averaged results.

Anomaly detection (AD). Anomaly detection methods, such as one-class SVM (OCSVM) and local outlier factor (LOF), seem more appropriate for our setting. Among them, LOF is unsupervised, finding outliers by measuring the local deviation of a given data point with respect to its neighbors. OC-SVM can be fitted in a supervised manner, with only the negative instances in the training set, to identify the suspects.

Two-step (TS) methods. As aforementioned, our approach is a TS method, consisting of unsupervised one-class SVM and supervised two-class SVM. For comparison, we also experiment using the LOF as the first step to identify potential suspects, and further use a classification method (*e.g.*, LR or DT) the second step, to filter out the false positives.

For all the methods with parameters, we optimize the parameters with 10-fold cross-validation by further dividing the training set into 80% for model fitting and 20% for parameter validation.

Evaluation metrics. We use *precision, recall, and F-score* computed with test set to evaluate the performances of different methods. The precision is the number of correctly identified positives divided by the number of identified positives instances. The recall is the number of correctly identified positives divided by the number of all positive instances in the test set. Then, the F-score is defined as:

$$\text{F-score} = 2 \times \frac{precision \times recall}{precision + recall}.$$

5.2 Results Summary

Table 3 summarizes the performances of our method and the baselines listed above.

Table 3: A Performance Comparison.

Algorithm	Precision	Recall	F-score	Run Time(s)
CM Methods				
DT	0.002	0.451	0.004	44.81
LR	0.003	0.476	0.006	36.72
SVM	0.005	0.512	0.009	21.31
AD Methods				
LOF	0.004	0.560	0.009	300+
OCSVM	0.015	0.583	0.029	39.67
TS Methods				
LOF+DT	0.011	0.780	0.022	301.18+
LOF+LR	0.016	0.829	0.031	301.16+
OCSVM+DT	0.053	0.878	0.099	41.19
TS-SVM	**0.071**	**0.927**	**0.133**	**41.05**

We have several interesting observations which confirm our research motivation. First, the precisions of all one-step methods are very low, especially for classification methods including DT, LR, and SVM. The AD methods perform somehow better, but still lower than our two-step framework. In contrast, all the two-step combinations significantly improve the precisions, among which, our TS-SVM performs best. This observation shows that the two-step approach can effectively reduce the false-positives. Second, two-step methods also perform better in terms of other metrics. For example, the recall of our TS-SVM is consistently above 90%, by finding the detection/classification boundaries in the non-linear kernel space. Finally, given the excellent recall of TS-SVM, we contend that it's "ground-truth" precision can be higher than the reported 7%. The reason is that not all the suspects have been caught or reported.

5.3 Feature Analysis

To further study the discriminative power of the features, we evaluate the performance of our framework (TS-SVM) with different feature combinations.

As shown in Figure 9, we use Db, Sc, and Hb to represent the *daily behavior*, *social comparison*, and *historical behavior* features, respectively. Most significantly, the precision of the daily behavior features is improved by the social comparison and further the historical behaviors. Such improvements can also be observed for other metrics in the table.

We also empirically compare the modeling performances on weekdays and weekends. As expected, since the human mobilities in weekends are more complicated, the detection accuracy of our method is slightly lower on weekends.

6. DEPLOYMENT AND INSIGHTS

With the automatic feature extraction and two-step suspect detection model described above, we developed a decision support system for the security personnel to easily spot

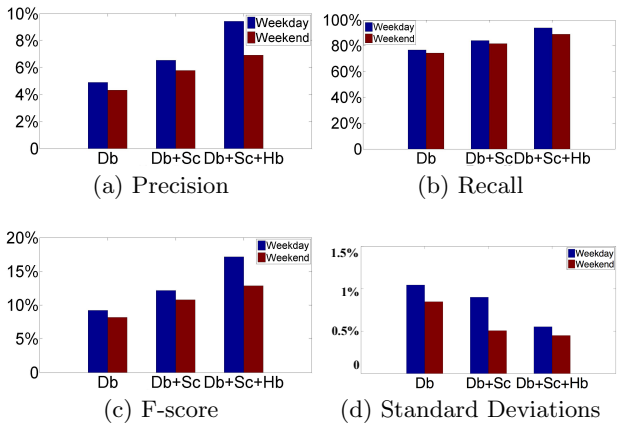


Figure 9: The impact of feature combinations.

pickpocket hotspots, and crack down on the suspects efficiently at the crime scene. In this section, we demonstrate how our system works in practice.

6.1 A Prototype System

The prototype system was implemented using bootstrap[4], Java and SparkQL. Figure 10 is a screenshot of the graphical user interface (GUI), which can be viewed on a computing terminal or a mobile device. The GUI has the following five basic components, which allow users to view suspect analytics at different levels of details.

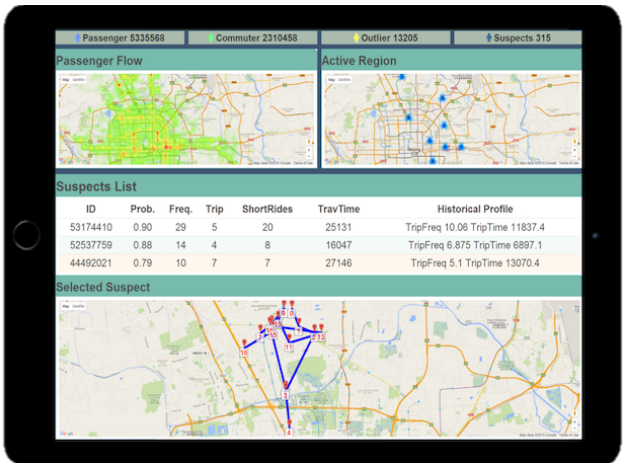


Figure 10: A screenshot of the prototype system.

Statistics. Summary statistics about the transit system status are provided at the top of the screen, which includes the total numbers of passengers, commuters, outliers, and suspects, respectively. The user is allowed to specify the time window for these statistics, in terms of the number of days, in the settings.

Passenger Flows. The density of passenger flows have a high correlation with pickpocket activities. The live state of passenger flow is shown with a heat map, where the density of passenger flow of each station is expressed by blending the color between green and red. (Redder means higher density.) This map helps users find areas with more traffic and thus are more vulnerable to theft.

Active Regions. Active regions of suspects at the city level is visualized in the "active regions" map. These active regions are indicated by the blue flashing circles, which are found by calculating the centroids of a DBSCAN algorithm. The user can inspect a specific area by zooming in.

Suspect List. Passengers identified as suspects will be listed on the "suspect list." Profiles of these suspect, such as smart card ID number, total travel time, riding frequency, the number of trips and the number of short rides, will be displayed by default. Quantile score against the social norm, historical profile information, and system determined likelihood of a suspect, are also available. The user is allowed to choose which features to display and to sort with.

Selected Suspect. When a suspect on the list is selected, his or her trajectory, as represented by linking the boarding and exiting stations, will be displayed in the "selected suspect" panel.

The database which stores identified suspect records is updated every day with newly collected transactions data. The two-step suspect detection model is trained offline, so that detectives can get an instant result when they interact with the system.

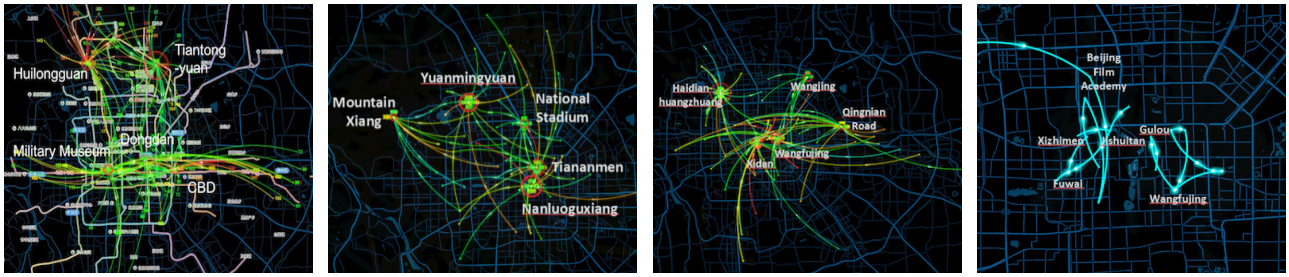
6.2 Examples of Passenger Behaviors

As mentioned in the "passenger flow" function above, we visualize the passenger movement patterns on the city map, which will be used for analyzing the behaviors of different types of passengers.

Figure 11 provides examples of representative movement patterns in different passenger groups in Beijing on a typical day from 8:00 a.m. to 11:00 a.m. Recall that each curve in the figure represents the transition between a pair of origin and destination, and the color represents traffic density between regions (red=high, green=low). In addition, the more curves through the same node means the more routes pass through that region.

The overall passenger flow, as shown in Figure 11(a), provides users a bird's view to know which regions have dense traffic at the city level. We can see that the Huilongguan area, Tiantongyuan area, Military Museum, CBD area, and the Dongdan area have the highest density. In order to understand the transit purposes, we classify passengers by the major categories of functional regions they visit. Besides Figure 11(a), which covers primarily regular commuters (*i.e.*, those who visit residence, workplace and transit regions), Figure 11(b) shows that visitors frequently visit Yuanmingyuan, Tiananmen, and Nanluoguxiang, whereas Figure 11(c) shows that shoppers tend to visit regions like Wangfujing and Xidan. Most of the normal travels are one-way or round-trip with clear directions. In contrast, pickpockets behave quite differently. As visualized in Figure 11(d), pickpockets tend to present a wandering pattern without a clear destination. They tend to make frequent, random stops with short travel segments. They also like to visit a variety of functional regions, such as transit hubs (*e.g.*, Xizhimen), shopping regions(*e.g.*, Wangfujing), and scenic spots (*e.g.*,

[4] Bootstrap (https://wrapbootstrap.com) is a web client framework.



(a) all passengers (b) visitors (c) shoppers (d) thieves

Figure 11: Movement patterns of different type of passengers.

Gulou), whereas most regular passengers only visit one functional region during a short period of time.

7. RELATED WORK

As urban sensing data, such as GPS traces, call details records, and smart card logs, grow ubiquitous, research efforts devoted to analyzing such data has resulted in a number of works in recent years. In this section, we provide a brief review of the related work, categorized into two groups.

7.1 Passengers Activity Patterns

The first group of literature focuses on finding patterns in passenger activity records. Such knowledge can be useful in a variety of applications, and plays a vital role in effectively finding and satisfying passenger needs. Examples include assessing the performance of the transit network, identifying and optimizing problematic or flawed bus routes, improving the accuracy of passenger flow forecast between two regions, and making service adjustments that accommodate variations in ridership on different days. In particular, using AFC data, [3] estimates the crowdedness of various stations in the transportation network. [22] measures the variability of transit behaviors on different days of the week. In addition, different studies have investigated unique characteristics of traveling patterns of the elderly [27], students, and adults [22], which provide interesting insights for understanding behavioral differences of sub-populations.

It has been suggested that human mobility patterns follow a high degree of spatial and temporal regularity, and are thus highly predictable [9, 26]. Existing studies in discovering trip patterns typically aim at discovering movement patterns by finding frequently visited places of regular passengers, who travel with the same sequence of places at the similar time of day. For example, [5] identifies spatiotemporal patterns from GPS traces of taxis for night bus route planning. [19] tries to reflect the common routing preference of the past passengers by finding the most frequent path of a certain time period. [6] discovers and explains movement patterns of a set of moving objects (e.g. traffic management, birds migration, disease spreading).

7.2 Abnormal Traveling Behavior Detection

Existing studies on detecting anomalies in urban sensing data can be divided into two categories: those based on locations, and those on trajectories.

Along the line of location-based anomaly detection, [28] presents a framework that learns the context of different functional regions in a city, which provides the basis of our feature extraction approach. Besides, [15] attempts to discover casual relationships among spatiotemporal outliers. [24] mines representative terms from social media that people posted when location-relevant events happen in the city, such as accidents and protests. [10] discovers black-hole or volcano patterns in human mobility data in a city, which can quickly identify gathering events, such as football matches and concerts. Detection of such anomalies can help sensing abnormal events, and provide input for intelligent decision support, such as smoothing the traffic flow [10].

The main goal of trajectory based anomaly detection is to find a small percentage of individuals, whose movement traces are different from the general population. One example is to identify fraudulent taxi driving behaviors. A large number of studies have investigated trajectory based anomaly detection using data mining techniques, such as graph based [6], clustering based [20, 3, 1], local/context-aware based [2, 13], dimension reduction based [11] and evidence based (*e.g.*, using Dempster-Shafer theory [8]). While the trajectory of pickpockets with features that are implicit, previously unknown, and potentially useful from large datasets, pickpocket suspect detection based on AFC records is a novel problem that has not been considered in the literature, and is quite challenging.

8. CONCLUSION

In this paper, we developed a suspect detection and tracking system by mining large-scale transit records. The system can help identify pickpocket suspects' and enable active surveillance in high-risk areas. Specifically, we first constructed a feature representation for profiling passengers. Then, we established a novel two-step framework to distinguish regular passengers from pickpocket suspects. Finally, we leverage real-world datasets from multiple sources for model training and validation, and implement a prototype system for end users. Experimental results on real-world data showed the effectiveness of our proposed approach.

9. ACKNOWLEDGMENTS

This research was supported in part by National Natural Science Foundation of China (No. 51408018), National Natural Science Foundation of China (No. 71329201), National High Technology Research and Development Program (863, 2013AA01A601).

References

[1] Paul Bouman, Evelien Van der Hurk, Leo Kroon, Ting Li, and Peter Vervest. Detecting activity patterns from smart card data. In *BNAIC*, 2013.

[2] Markus M. Breunig, Hans-Peter Kriegel, Raymond T. Ng, and Jörg Sander. LOF: Identifying density-based local outliers. *SIGMOD Rec.*, 29(2):93–104, May 2000.

[3] Irina Ceapa, Chris Smith, and Licia Capra. Avoiding the crowds: understanding tube station congestion patterns from trip data. In *UrbComp*, pages 134–141, 2012.

[4] Olivier Chapelle, Eren Manavoglu, and Romer Rosales. Simple and scalable response prediction for display advertising. *ACM Trans. Intell. Syst. Technol.*, 5(4):61:1–61:34, 2014.

[5] Chao Chen, Daqing Zhang, Zhi-Hua Zhou, Nan Li, Tülin Atmaca, and Shijian Li. B-planner: Night bus route planning using large-scale taxi gps traces. In *PerCom*, pages 225–233, 2013.

[6] Ticiana L. Coelho da Silva, José A. F. de Macêdo, and Marco A. Casanova. Discovering frequent mobility patterns on moving object data. In *MobiGIS*, pages 60–67, 2014.

[7] Marcus Felson and Ronald V Clarke. Opportunity makes the thief: Practical theory for crime prevention. Report 98, Policing and Reducing Crime Unit: Police Research Series, 1998.

[8] Yong Ge, Hui Xiong, Chuanren Liu, and Zhi-Hua Zhou. A taxi driving fraud detection system. In *ICDM*, pages 181–190, 2011.

[9] Marta C Gonzalez, Cesar A Hidalgo, and Albert-László Barabási. Understanding individual human mobility patterns. *Nature*, 453(7196):779–782, 2008.

[10] Liang Hong, Yu Zheng, Duncan Yung, Jingbo Shang, and Lei Zou. Detecting urban black holes based on human mobility data. In *GIS*, 2015.

[11] Shan Jiang, Joseph Ferreira Jr, and Marta C Gonzalez. Discovering urban spatial-temporal structure from human activity patterns. In *UrbComp*, pages 95–102, 2012.

[12] Chuanren Liu, Fei Wang, Jianying Hu, and Hui Xiong. Temporal phenotyping from longitudinal electronic health records: A graph based framework. In *KDD*, pages 705–714. ACM, 2015.

[13] Chuanren Liu, Hui Xiong, Yong Ge, Wei Geng, and Matt Perkins. A stochastic model for context-aware anomaly detection in indoor location traces. In *ICDM*, pages 449–458, 2012.

[14] Chuanren Liu, Kai Zhang, Hui Xiong, Guofei Jiang, and Qiang Yang. Temporal skeletonization on sequential data: Patterns, categorization, and visualization. *IEEE Transactions on Knowledge and Data Engineering*, 28(1):211–223, Jan 2016.

[15] Wei Liu, Yu Zheng, Sanjay Chawla, Jing Yuan, and Xie Xing. Discovering spatio-temporal causal interactions in traffic data streams. In *KDD*, pages 1010–1018, 2011.

[16] Xianglong Liu, Cheng Deng, Bo Lang, Dacheng Tao, and Xuelong Li. Query-adaptive reciprocal hash tables for nearest neighbor search. *IEEE Transactions on Image Processing*, 25(2):907–919, 2016.

[17] Xianglong Liu, Yadong Mu, Bo Lang, and Shih-Fu Chang. Mixed image-keyword query adaptive hashing over multilabel images. *ACM Transactions on Multimedia Computing, Communications, and Applications*, 10(2):22:1–22:21, 2014.

[18] Yanchi Liu, Chuanren Liu, Jing Yuan, Lian Duan, Yanjie Fu, Hui Xiong, Songhua Xu, and Junjie Wu. Intelligent bus routing with heterogeneous human mobility patterns. *Knowledge and Information Systems*, Forthcoming (Accepted as of Feb 2016).

[19] Wuman Luo, Haoyu Tan, Lei Chen, and Lionel M Ni. Finding time period-based most frequent path in big trajectory data. In *SIGMOD*, pages 713–724, 2013.

[20] Xiaolei Ma, Yao-Jan Wu, Yinhai Wang, Feng Chen, and Jianfeng Liu. Mining smart card data for transit riders travel patterns. *Transportation Research Part C*, 36:1–12, 2013.

[21] H. Brendan McMahan, Gary Holt, D. Sculley, Michael Young, Dietmar Ebner, Julian Grady, Lan Nie, Todd Phillips, Eugene Davydov, Daniel Golovin, Sharat Chikkerur, Dan Liu, Martin Wattenberg, Arnar Mar Hrafnkelsson, Tom Boulos, and Jeremy Kubica. Ad click prediction: A view from the trenches. In *KDD*, pages 1222–1230, 2013.

[22] Catherine Morency, Martin Trépanier, and Bruno Agard. Analysing the variability of transit users behaviour with smart card data. In *ITSC*, pages 44–49, 2006.

[23] Graeme R Newman and Megan M McNally. Identity theft literature review. Report 210459, United States Department of Justice, July 2005.

[24] Bei Pan, Yu Zheng, David Wilkie, and Cyrus Shahabi. Crowd sensing of traffic anomalies based on human mobility and social media. In *GIS*, pages 344–353, 2013.

[25] Bernhard Schölkopf, John C. Platt, John C. Shawe-Taylor, Alex J. Smola, and Robert C. Williamson. Estimating the support of a high-dimensional distribution. *Neural Computation*, 13(7):1443–1471, 2001.

[26] Chaoming Song, Zehui Qu, Nicholas Blumm, and Albert-László Barabási. Limits of predictability in human mobility. *Science*, 327(5968):1018–1021, 2010.

[27] Myung Joon Sung. Analysis of travel patterns of the elderly using transit smart card data. In *TRB*, volume 11-2357, 2011.

[28] Jing Yuan, Yu Zheng, and Xing Xie. Discovering regions of different functions in a city using human mobility and pois. In *KDD*, pages 186–194, 2012.

[29] Kai Zheng, Yu Zheng, Nicholas Jing Yuan, Shuo Shang, and Xiaofang Zhou. Online discovery of gathering patterns over trajectories. *IEEE Transactions on Knowledge and Data Engineering*, 26(8):1974–1988, 2014.

[30] Hengshu Zhu, Chuanren Liu, Yong Ge, Hui Xiong, and Enhong Chen. Popularity modeling for mobile apps: A sequential approach. *Cybernetics, IEEE Transactions on*, 45(7):1303–1314, 2015.