Discovering People's Life Patterns from Anonymized WiFi Scanlists

Sha Zhao[1], Zhe Zhao[1], Yifan Zhao[1], Runhe Huang[2], Shijian Li[1], Gang Pan[1,3]
[1]Department of Computer Science, Zhejiang University, China
[2]Faculty of Computer and Information Sciences, Hosei University, Japan
[3]Cyber Innovation Joint Research Center, Zhejiang University, China
[1]{szhao, yifanzhao, zhezhao, shijianli, gpan}@zju.edu.cn
[2]rhuang@hosei.ac.jp

Abstract—**The prevalence of smart phones equipped with various sensors enables pervasive capturing users' mobility data (GPS, GSM network, WiFi, etc.), which contains approximate whereabouts of users. In order to protect users' privacy some mobility data is anonymized, which is challenging for discovering individual information implicated in the data. In this study, we are attempting to discover people's life patterns, which capture individual's life regularity and living style, from the anonymized WiFi scanlists. We transform the life pattern discovery problem into an unsupervised problem by extracting stay place, discovering trajectory patterns, etc. Particularly, we design a user feature space in which we use frequent trajectory patterns to represent each user as a feature vector. Thus, the life pattern discovery problem can be solved by finding clusters of users in the user feature space. The proposed approach is verified using the Device Analyzer data, which contains records of smart phone usage of more than 17,000 volunteering participants. Our work is a promising step towards automatically mining people's life patterns from anonymized mobility data of smart phones.**

Keywords—anonymized WiFi scanlists; stay place; trajectory pattern; user feature space; life pattern

I. INTRODUCTION

Nowadays, more than 1.5 billion people worldwide have been covered by smart phones, which are becoming people's essential belongings. Frequent use of smart phones generates much personal historical location data. The development in location-acquisition technologies, such as GPS, GSM network, WiFi MAC AP (access point), enables the capture of users' location data by smart phones. Some applications are developed to collect the captured data, such as Lausanne data collection campaign (LDCC) for Mobile Data Challenge [1]. A continuous collection of individual history location data for a long duration will bring detailed records about users' mobility. The ubiquitous and large-scale individual history data derived from mobile phone data bring us challenges as well as opportunities to discover valuable knowledge from the raw data.

Considering the closed relationship between people's daily life and geographic locations, information on one's life, such as significant places, mobility pattern and daily routine, can be discovered from his/her location history. The discovery of individual life has a manifold of application scenarios. On one hand, the discovered knowledge based on location data can help users understand their way of living objectively and extensively. On the other hand, the valuable knowledge can be used in some location-based services, such as location recommender system, context-based computing system, smart environment, and personalized services. Besides, the discovered knowledge can be used in building human model, such as SmartShadow [2]. Furthermore, studying mobility patterns or daily routine of multiple individuals is important for mining valuable knowledge about social development trends.

Several work has been made for mining mobility regularity and living habits using location history data, for example, discovering semantically meaningful and important places using radio trace of GPS, WiFi and GSM cellular [3, 4], discovering places of interest in everyday life by integrating GPS, WiFi, GSM and accelerometer sensors [5], predicting the next place using some context features (time, accelerometer, application Bluetooth and WLAN, call log, system) [6] and GPS data [7], and mining individual mobility pattern based on location history derived from cellular network data [8] and CDRs (Call Details Records) [9]. Some location data can cause exposure of users' privacy, such as GPS with longitude and latitude. In order to protect users' privacy, some data is anonymized when being collected, which is challenging for discovering individual information indicated in the data.

Compared with GPS, GSM network, WiFi logs of smart phones are easily to be collected, including WiFi scanlists of a smartphone, AP MAC addresses, etc. The WiFi APs are periodically scanned. It can roughly indicate the physical location of the phone. Thus, the WiFi scanlists accompanying one's moving reveal rich information on his/her mobility patterns and lifestyle, such as "how often users visit some significant places", "how regular a trajectory happens", "how many places contained in a frequent trajectory". We define the phenomenon mentioned above as people's life pattern, which captures individual life regularity and living style.

In this paper, we are attempting to discover people's life patterns from anonymized WiFi scanlists. Firstly, we extract stay places from WiFi scanlists, where users stay for a significant amount of time. Secondly, PrefixSpan is employed to discover trajectory patterns from individual trajectories. Sequently, we design a user mobility space by using frequent trajectory patterns to represent each user as a feature vector. Considering the similarity of trajectory patterns shared by users, we discover life patterns by finding clusters in the user feature space. Thus, the life pattern discovery problem can be transformed to a clustering problem, which can be solved many classical clustering methods. In this paper, the DBSCAN (Density-Based Spatial Clustering of Applications with Noise) is employed. We have a real-world dataset of WiFi scanlists collected by Device Analyzer [10] for the experiment evaluation.

Based on the steps mentioned above, the contributions of our research are three-fold:

- In this paper, we discover people's life pattern from anonymized WiFi scanlists. The anonymous data makes the problem more challenging to be solved, such as in expressing users' stay places in physical meaning and constructing individual trajectory patterns. Our work is a promising step towards automatically mining people's life patterns from anonymized mobility data captured by smart phones.
- To solve the problem, we design a user feature space in which we use frequent trajectory patterns to represent each user as a feature vector. This kind of user representation can express each user's life regularity and living style, which is helpful for discovering common characteristics of trajectory patterns in the user feature space.
- To solve the problem, we transform the life pattern discovery problem into an easy-to-solve form: finding clusters of users in the user feature space. Our approach includes stay place extraction, trajectory pattern discovery and life pattern discovery. At last, we use a real-world dataset of anonymized WiFi scanlists collected by Device Analyzer to evaluate our approach. The life pattern we discover is more representative and general, because the WiFi dataset we use is not limited to sampled users and location areas, covering more than 17,000 users from over 150 countries.

The remainder of the paper is organized as follows. In Section II, the related work is reviewed. Section III analyzes several aspects regarding to individual life pattern, formalizes the relevant definition, and states the problem we are attempting to solve. In Section IV, our proposed approach is described by steps. The experimental evaluations and discussions are provided in Section V. Our conclusion and outlook on the future work are given in Section VI.

II. RELEATED WORK

The increasing availability of individual location history data has brought about relevant researches in the last decade, such as extracting significant places, mining mobility pattern, discovering daily routine or daily living activities, and so on. It has been shown that users' significant places can be extracted from mobile phone data [4, 11, 12, 13]. For example, the authors of [11] extracted origin and destination regions for each trip of each user. In [4], by using anonymized cellular network the authors identified generally important locations and discerned semantically meaningful locations such as home and work. In [12], the authors identified important places, such as home and work, from the sequence of visited places and accelerometer samples. [13] addressed the problem of automatic place labeling based on co-occurrence of WiFi AP and GPS data. The frequently visited places can be recognized reliably (over 80 percent).

There are also several works on mining users' mobility pattern from mobile phone data [14, 15, 16] and taxi GPS data [17, 18, 19]. Based on the location information derived from cell towers, the work of [15] has discovered that human trajectories show a high degree of temporal and spatial regularity. It was discovered that the individual travel patterns collapse into a single spatial probability distribution: humans follow simple reproducible patterns. The authors of [14] analyzed each individual in the CDRs to capture their home/ work locations, investigated the formation of segregated communities based on users' home and work locations, and estimated people flows within the city within a day time scale. In [16], the authors used cellular networks of real-world cell phone data to analyze human mobility in city-wide level. Their analysis of mobile profiles of cell phone users exposes a significant tail in a user's location-time distribution: A total of 15% of a cell phone user's time is spent on average in locations that each appears with less than 1% of total time.

In addition, several studies are focusing on discovering users' daily routine or daily living activities [20, 21, 22]. For instance, the researchers of [21] proposed a framework to automatically discover location-driven routines from the day in the life of a person without any supervision. The discovered daily routines include "going to work at 10am", "leaving work at night", "being at home in the mornings and evenings while being out in the afternoon" and so on. In[22], the authors mined individual life pattern based on GPS data collected by GPS devices. The discovered life pattern, which captures individual's life style and regularity, focuses on significant places of individual life and considers diverse properties to combine the significant places. In [20], trajectory patterns of users, namely GTS (Geographic-Temporal-Semantic) patterns, were discovered to capture frequent movements by taking into account users' geographic-triggered intentions, temporal-triggered intentions and semantic-triggered intentions. On the basis of GTS patterns, the authors predicted next location of users more precisely than other location prediction techniques. In this paper, we aim at automatically discovering individual life pattern from anonymized WiFi scanlists of smart phones.

Some mobility data used in aforementioned works are sensitive to users' privacy, such as GPS data with longitude and latitude used in [22]. Mobility data is among the most sensitive data currently being collected and personal information could be inferred about an individual knowing only his/her mobility trace [23]. Considering the privacy protection and easy availability of WiFi logs of smart phones, we are attempting to discover people's life patterns from anonymized WiFi scanlists. The anonymous data make it more difficult to be solved, such as in expressing users' stay places in physical meaning and constructing individual trajectory patterns.

III. PROBLEM STATEMENT

This paper addresses mining one's underlying life patterns from mobility information indicated by anonymized WiFi scanlists. A WiFi scanlist can be denoted by a set of WiFi access points: $L=\{w_i\}$, where w_i is the hashed (thus anonymized) MAC address of a WiFi AP. A set of WiFi scanlists can be denoted by $W=\{(L_i,t_i)\}$, where (L_i,t_i) is a WiFi scanlist L_i scanned at the time t_i.

The WiFi scanlist is recorded around every five minutes. It can roughly indicate the physical location of the phone at the time. According to the scanlists, one's stay places can be detected, which has been visited for a significant amount of time. In order to mine life pattern, we only take one's stay places into consideration and ignore those places where the user just passes by. The stay places, such as work office, home or school, can reflect individual's activities.

A *stay place* p represents a geographic region in which a user stays for a while without departure. The record (t_i, p_i) means that a user is staying at place p_i at the scan time t_i. A *stay place* may occur many times in a period of time. The occurrence of a *stay place* is computed using the number of days in which it occurs. For instance, a stay place occurs in 80 days in a period of 100 days and its occurrence is 0.8. The tuple (p_i,o_i) represents stay place p_i with occurrence of o_i.

DEFINITION 1. An individual's trajectory over a time interval can be depicted by connecting the stay places in order according to the scan time. An individual trajectory over a time interval is a sequence of stay places, i.e.,

$$t: p_1 \rightarrow p_2 \rightarrow \ldots \rightarrow p_n \tag{1}$$

where p_i is a stay place.

Given that life pattern is related with scale of temporal granularity, individual trajectory is partitioned into subsequences according to the specific granularity like "day", "week", etc. In this way, for a period, individual trajectories can be denoted as $D=\{t_1,t_2,t_3,\ldots,t_n\}$. Each t_i in D corresponds the trajectory of one day, or one week or etc., according to the granularity. All of individual trajectories with specific granularity share common sequential patterns because people tend to stay more frequently at their regular places.

DEFINITION 2. A trajectory pattern over a determined time interval is derived as a sequence of common places shared by all trajectories with a specific temporal granularity. Each pattern P is associated with a support value s, which denotes the percentage of temporal observation units (temporal granularity) when P is satisfied. Hence, a trajectory pattern can be represented as

$$(P,s) \tag{2}$$

For instance, the trajectory pattern P_1 occurs in 80 days in the total time of 100 days and it can be represented as $(P_1,0.8)\cdot$ The number of sequential stay places in a trajectory pattern is defined as the length l of the pattern. The length of one's trajectory patterns may be different, such as length-1, length-2, length-3, and so on.

The support value s and length l of one individual's trajectory patterns reflect his/her life style and regularity, such as "how many frequently visited places", "at what order the user visits the places", "how often a trajectory happens", and so on .

DEFINITION 3. People's life pattern can be represented by the common characteristics of life regularity and living style shared by multiple users with similar trajectory patterns. Thus, the people's life pattern can be represented as

$$P_{life}\text{: } Similarity(u_1,u_2,\cdots,u_i,\cdots,u_n) \tag{3}$$

where u_i represents a user.

In the real world, there exits similarity in trajectory patterns in the length and support value among multiple users. That is, some users have similar life regularity and living style. The common characteristics of life regularity and living style shared by similar users represent a type of life pattern. Thus, Life pattern can be discovered from multiple users with similar trajectory patterns.

PROBLEM Given a set of anonymized WiFi scanlists $W=\{(L_i,t_i)\}$ of smart phones, find the similarity of users' life regularity and living style so as to automatically discover people's life patterns P_{life}.

IV. THE PROPOSED APPROACH

In order to mine people's life patterns, it is necessary for us to firstly extract stay places from anonymized WiFi scanlists. We depict individual trajectories with a specific temporal granularity by connecting stay places in sequence. Trajectory patterns can be discovered from trajectories. Frequent patterns are selected to represent each user as a feature vector. An unsupervised method is applied to cluster users so as to look at individual life pattern. Figure 1 shows the overview of our approach.

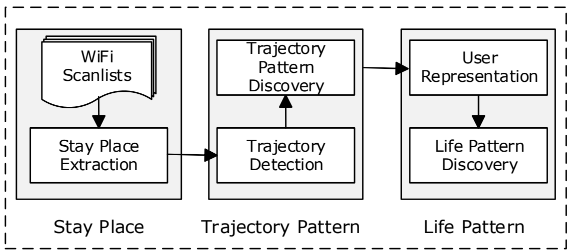


Figure 1. Overview of our approach

A. *Stay place extraction*

A *stay place* p represents a geographic region in which a user stays for a significant amount of time. Life pattern emphasizes on *stay places* while ignores the transition between these places. A *stay place* p_i can be detected when consecutive WiFi scans contain some same AP MAC addresses. For stay place detection, a big challenge is how to deal with intermittent MACs during users' moving. In order to reduce the impact on extracting stay places from these intermittent MACs, a WiFi AP will be filtered if its stay duration less than 10 minutes. Scanlists in the same cluster represent a *stay place* p. The tuple (t_i, p_i) records that a user is staying at place p_i at the scan time t_i. The UIM clustering algorithm [24] is employed to cluster WiFi scanlists into a set of stay places.

The execution of UIM algorithm will be introduced in details. First, scanlists frequently appearing together are selected and defined as good set of scanlists. Sequently, the similarity between the scanlists in the good set is computed and similarity graph G_S is constructed. Then, a candidate cluster set C_C is obtained by clustering the vertex of the G_S. Finally, candidate clusters are merged based on the similarity measures to obtain the set C_F of final clusters. Each cluster in C_F represents a stay place.

B. *Trajectory patterns discovery*

A temporal sequence of stay places composes an individual history trajectory, indicating where he/she visits at what order. To mine people's life pattern in daily life, the temporal observation unit for individual trajectory is set to one day. That is, one's moving in each day can be represented by a daily trajectory. A trajectory t_i in individual trajectories $D = \{t_1, t_2, t_3, \ldots, t_n\}$ is for one day. Figure 2 gives an example of one individual's trajectory in one day.

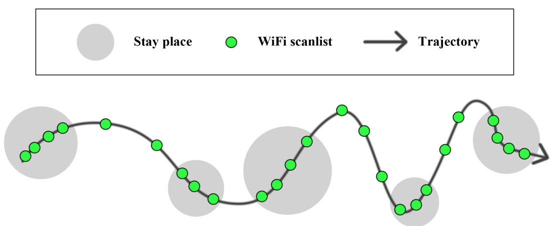


Figure 2. An example for a trajectory in one day

The trajectory occurring in every day may be different in the order or the number of the visited places. However, there exists regularity among one individual's trajectories considering that people tend to stay more frequently at their regular places. That is, there exist common patterns shared by all daily trajectories. For example, one individual frequently visit three places in his/her daily life. In addition, he/she visits these three places in a specific order. Figure 3 shows an illustration of the trajectory patterns (p_i: stay place).

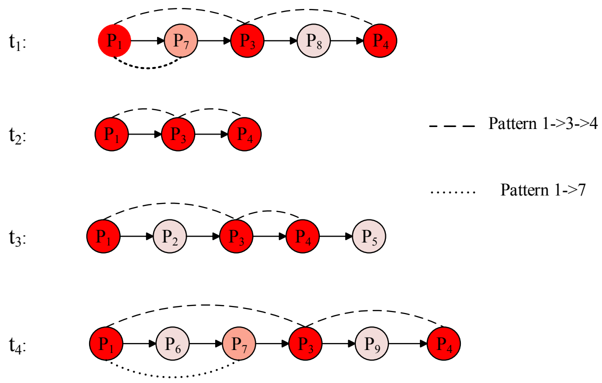


Figure 3. One individual's four trajectories in four days.

In Figure 3, there are four trajectories of one individual in four days. Although the four trajectories are different in the order and number of the stay places, there exists regularity. This user visits Place 1, 3, and 4 most frequently, visits Place 7 more frequently and rarely visits Place 2, 5, 6, 8 and 9. In addition, he/she regularly visits places in the sequential order of 1->3->4 and 1->7. In addition, 1->3->4 occurs more frequently than 1->7. The regularity reflected by one individual's multiple trajectories in different days represents individual trajectory patterns, such as "how many places the user visit frequently", "at what order the user visits places", "how often a trajectory happens", "how many places contained in a frequent trajectory", and so on.

Individual trajectory pattern is a sequential pattern shared by all of trajectories during a period. It is necessary for us to mine the frequent trajectory patterns from individual trajectories in order to discover people's life patterns. Having formalized trajectory pattern, the task of discovering frequent trajectory patterns (P, s) from trajectories D can be described as: given trajectory sequences and a support threshold s, to discover all subsequences as patterns with support value (percentage of days) greater than s.

There are many developed sequential pattern mining methods, such as GSP [25], FreeSpan [26] and SPADE [27]. In this paper, PrefixSpan [28] is employed to discover individual trajectory patterns (P, s) from the trajectories D.

C. *Life pattern discovery*

Trajectory patterns reflect individual life regularity and living style. In the real world, some users have similar trajectory patterns. The common characteristics of life regularity and living style shared by similar users represent a type of life pattern. In order to discover people's life patterns from multiple users' trajectory patterns, we use frequent trajectory patterns to represent each user. Intuitively, frequent trajectory patterns correspond to more general life pattern. The support value and length of the selected patterns can reflect users' life regularity and living style. Frequent trajectory patterns with different length are selected to represent each user as a feature vector. The value of each dimension is the support value of each pattern. Each user is formally represented as

$$u = (P_1, P_2, \cdots, P_i, \cdots, P_n) \tag{4}$$

where P_i represents a selected frequent trajectory pattern.

The user representation with trajectory patterns reflects the characteristics of users' life regularity and living style. The different length reflects how many places users frequently visit, and the different support value reflects how often the patterns occur. In addition, the sequence of the stay places involved in the selected frequent patterns represent at what order users visit regular places. Figure 4 shows an example for user representation. The user is represented as a 5-demension feature vector using 5 selected frequent patterns. And the value of each dimension is 0.89, 0.67, 0.34, 0.53 and 0.44 respectively.

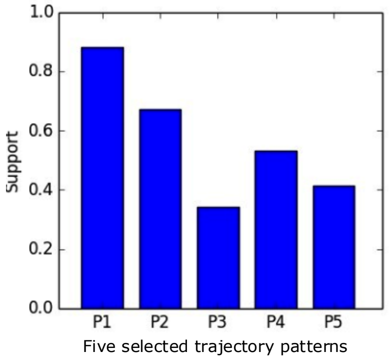


Figure 4. An example for user representation

After user representation, users with similar life patterns form a cluster in the mobility feature space. The common characteristics in life regularity and living style shared by users in the same cluster represent a type of life pattern. Thus, the task of life pattern discovery can be transformed to a clustering problem: to find clusters of users in the mobility feature space and then discover people's life patterns. Here, an unsupervised approach of DBSCAN (Density-Based Spatial Clustering of Applications with Noise) [29]is applied to cluster users so as to automatically discover individual life pattern.

V. Experiments and Analyses

A. *Data*

The Device Analyzer dataset is a collection of smart phone usage of more than 17,000 volunteering participants from more than 150 countries, who have installed a copy of Device Analyzer from the Android market and consented to their data being collected [10]. The data set has been made publicly available for the first time in the context of the UbiComp/ISWC 2014 Programming Competition. For our work, Device Analyzer captures a rich, highly detailed time-series log of approximately different 300 events, including CDR, app usage, alarm, battery and so on[1]. For our work, we focus on the data about WiFi records.

The WiFi logs contain the information about currently visible WiFi networks, including the AP MAC address that the device is currently connected to, state, scanlists, etc. Device analyzer periodically records WiFi scanlists of a smartphone including anonymized AP MAC address. It scans the WiFi APs around every five minutes. Each record in a scanlist contains the scanned time, anonymized AP MAC address and the related SSID. In this study, we focus on the anonymized AP MAC address.

The duration of individuals' WiFi logs varies from a day to nearly three years since they join the data collection program at different time. In this paper, we consider users who have contributed data over more than three months. Thus, there are 2500 users in total to be selected for experiments.

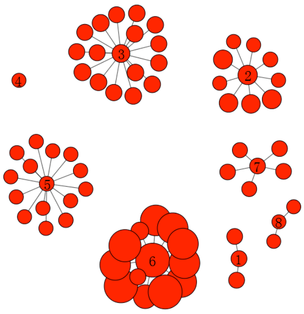


Figure 5. Extracted stay places of User A by UIM algorithm

B. *Result of stay place extraction*

WiFi scanlists were clustered into a set of stay places by the UIM clustering method [24]. In order to show the clustering result, we take a user for example, named User A with duration of around 300 days. The clustering result of User A is shown in Figure 5. Each node represents a WiFi scanlist. The node size means average stay duration. The numbers in the figure represent the ID of each cluster. There are 8 mined *stay places* for User A, where he/she stays more than 10 minutes. Taking User A's mobility record on Feb. 11^{th}, 2013 for example, Figure 6 shows User A stays at which stay place at what time. User A stays in Place6 overnight and in the morning leaves Place6 for Place2, 3, 5 and 8, where he/she spends most time in the daytime.

Beginning	Ending	Stay place ID
2013/2/11 00:02	2013/2/11 08:12	6
2013/2/11 08:32	2013/2/11 10:02	2
2013/2/11 10:22	2013/2/11 10:47	2
2013/2/11 11:48	2013/2/11 12:13	3
2013/2/11 13:29	2013/2/11 14:04	2
2013/2/11 14:09	2013/2/11 15:40	8
2013/2/11 16:04	2013/2/11 17:28	3
2013/2/11 18:02	2013/2/11 19:27	3
2013/2/11 19:46	2013/2/11 23:57	6

Figure 6. A sample of User A's mobility records

C. *Result of trajectory pattern discovery*

According to the research in [22], the patterns of workdays are more representative. We focused on users' trajectories and trajectory patterns on weekdays. For 2500 users in our experiment, 154,616 trajectories with temporal granularity of one day are detected. Based on trajectories,

[1] http://deviceanalyzer.cl.cam.ac.uk/keyValuePairs.htm

trajectory patterns with a support value (percentage of days) were discovered by applying PrefixSpan. The trajectory pattern with the greatest support value for each user usually is one-place (length-1) pattern, which consists of the most frequently visited place. The longer patterns are usually with lower support values.

Figure 7 shows User A's trajectories in February of 2013 and the trajectory patterns with support greater than 0.5. In Figure 7 a), trajectories are different while some of them share the same pattern. Place6 occurs in most trajectories and the pattern of 6 accounts for the largest proportion (s =0.987). In Figure 7 b), the trajectory patterns of 6-6, 6-2, 2-6 and 6-2-6 have higher support values (s =0.885, 0.832, 0.765, and 0.757 respectively), which show that User A frequently moves between Place6 and Place2.

Date	Trajectory
2013/2/1	6-3-7-3-2-5-5-2-2
2013/2/7	1-6
2013/2/8	6-3-7-3-3-2-2-6
2013/2/11	6-2-2-3-2-8-3-3-6
2013/2/12	6-3-3-3-6
2013/2/14	6-2-8-2-3-2-5-3-6
2013/2/15	6-3-7-5-5-2-6
2013/2/18	6-2-5-2-5
2013/2/20	2-5-5-5-2-3-3-3-6
2013/2/21	6-2-5-2-1-5-2
2013/2/25	6-2-2-2-3- 6
2013/2/26	6-3-3-3-6
2013/2/27	6-2-3-5-2-3-6
2013/2/28	6-2-2-5-5-2-3-1-3-6

a)

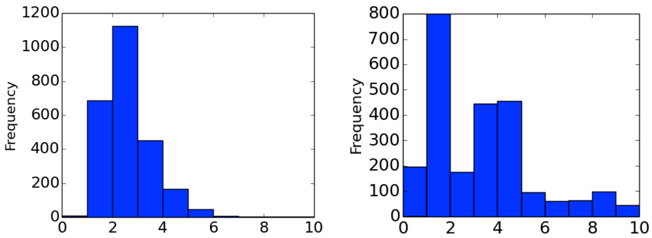
Pattern	Support
6	0.987
6-6	0.885
2	0.841
6-2	0.832
2-6	0.765
6-2-6	0.757
3	0.642
2-2	0.642
6-3	0.637
6-2-2	0.637
3-6	0.593
6-3-6	0.588
2-2-6	0.58
6-2-2-6	0.575

b)

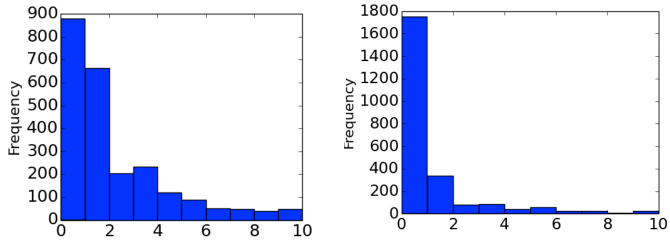
Figure 7. For User A, a) trajectories in February of 2013; b) trajectory patterns with support greater than 0.5.

D. *Result of life pattern discovery*

We used frequent trajectory patterns on weekdays with greater support than 0.2 for user representation and life pattern discovery. For trajectory patterns with the same length, the number and support value differ among users. We plotted the histogram of users' trajectory patterns ($s \geq 0.2$) in Figure 8. And the length is respectively 1, 2, 3 and 4 (Figure 8). For length-1 pattern, most users (about 1100 users) have 3 lenght-1 patterns ($s \geq 0.2$). Most users (about 800 users) have 1 length-2 pattern ($s \geq 0.2$). About 900 users have 1 length-3 trajectory pattern ($s \geq 0.2$). About 1800 users have 1 length-4 pattern ($s \geq 0.2$), but the support values of these length-4 patterns are much lower.



a) One-place patterns (length-1) **b)** Two-place pattern (length-2)



c) Three-place pattern (length-3) **d)** Four-place pattern (length-4)

Figure 8. The histogram of users' trajectory patterns ($s \geq 0.2$)

Based on the histogram in Figure 8, 5 most frequent patterns($s \geq 0.2$) were selected to represent each user as a 5-dimension vector: 3 one-place (length-1) patterns, 1 two-place (length-2) pattern and 1 three-place (length-3) pattern. The value of each dimension is the support value of the related pattern or zero if there is no relevant trajectory pattern with support value greater than 0.2. The features were firstly ranked in an ascending order according to the length of the pattern, then ranked in a descending order for the same length according to the support value.

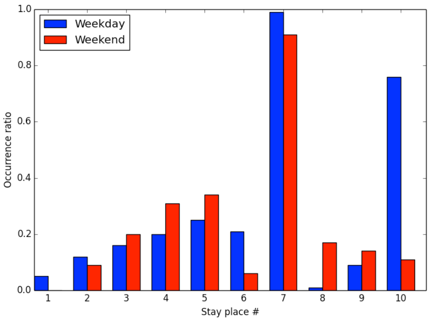
DBSCAN (Eps=0.1 in Euclidean distance, MinPts=10) was applied to cluster users in the feature space. For the selected 2500 users, there are 44.7%, 27.7% and 27.6% of users in 3 clusters respectively. To better understand users' life pattern of each cluster, we analyze the centroid of 3 clusters, shown in Figure 9. For each centroid user, the occurrence of each stay place on weekdays and weekends was compared. In the meantime, we analyzed frequent trajectory patterns with support value greater than 0.2.

For the largest cluster shown in Figure 9 a), it can be seen that the centroid user has two more frequent stay places, Place7 (s =0.988) and Place10 (s =0.765). His/ her regular trajectory patterns consist of the two places in different orders, such as 10-7, 7-10-7, etc. Combining the occurrence of each place on weekdays and weekends, Place7 has highest occurence both on weekdays and weekends while Place10 has higher occurence on weekdays but very low occurence on weekends. Life pattern of this user may be leaving Place7 for Place10 and going back Place7 after a period of time. This kind of life pattern is suitable for office staff, who usually leave home for work places and go back home after work on weekdays.

For the second largest cluster shown in Figure 9 b), the centroid user has more frequently visited places, Place8 (s =0.6), Place4 (s =0.543), and Place11 (s =0.343) while the occurence of each place is not so high compared with the centroid of the largest cluster and the smallest cluster. This user spends most of days in Place8 on weekdays and weekends, goes to Place4 more frequently on weekdays, and goes to Place11 more frequently on weekends.

For the smallest cluster shown in Figure 9 c), it can be seen that the centroid user has only one stay place with higher occurence, Place3 (s =0.793) with highest occurence on both weekdays and weekends. His/her frequent pattern is 3-3 (s =0.678) and 3-3-3 (s =0.552). He/she often goes to another Place5 (s =0.241) and goes here more frequently on weekends. This user spends most of time in Place3 and occasionally goes out for Place5 and then goes back to Place3 again.

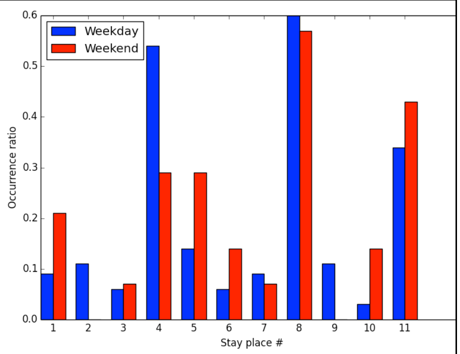
7	0.988
10	0.765
7-7	0.765
10-7	0.659
7-10	0.553
7-7-7	0.471
7-10-7	0.435
10-7-7	0.424
10-10	0.341
10-10-7	0.294
7-10-7-7	0.294
5	0.247
7-10-10	0.235
6	0.212
7-10-10-7	0.212



Trajectory Patterns | Support | The occurrence of stay place

a) Centroid of the largest cluster

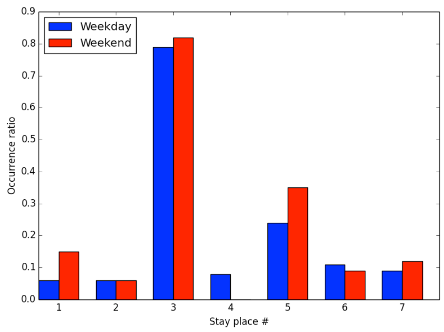
8	0.6
8-8	0.571
4	0.543
4-8	0.543
8-4	0.514
8-4-8	0.514
11	0.343
8-8-8	0.343
4-8-8	0.314
8-4-8-8	0.314
11-11	0.229



Trajectory Patterns | Support | The occurrence of stay place

b) Centroid of the second largest cluster

3	0.793
3-3	0.678
3-3-3	0.552
5	0.241
3-5	0.241
3-3-3-3	0.23
3-5-3	0.207
5-3	0.207



Trajectory Patterns | Support | The occurrence of stay place

c) Centroid of the smallest cluster

Figure 9. The occurrence of stay places and trajectory patterns ($s \geq 0.2$) of each centroid user in three clusters

VI. CONCLUSIONS

Based on the large-scale mobile phone dataset collected by Device Analyzer, we mined individual life pattern from anonymized WiFi scanlists. Firstly, stay places from the raw data were extracted and each user's mobility record was obtained. Individual trajectory was constructed by a sequence of stay places. From the individual trajectories, frequent trajectory patterns of each individual were mined. The support value and length of trajectory patterns reflects users' life regularity and living style. Using frequent trajectory patterns to represent each user as a feature vector, we designed a user feature space. The users with similar trajectory patterns form a cluster in the mobility feature space. Thus, the life pattern discovery problem was smoothly transformed into an unsupervised problem: finding cluster of users in the mobility feature space. By using unsupervised method of DBSCAN, users were clustered into 3 clusters, which represent 3 types of people's life pattern. In order to better understand each type of life patterns, we analyzed the characteristics of centroid users' mobility. Our work is a promising step towards automatically mining individual life pattern from anonymized mobility data captured by smart phones.

Although we mined people's life patterns from the raw data, the life patterns have no semantic information because of the groundtruth limitation. In the future work, we will make our efforts on mining semantic information of stay places by combining some context features, such as time tag.

ACKNOWLEDGMENTS

We would like to thank the researchers in Cambridge University who develops Device Analyzer for sharing their dataset. This work is supported by Program for New Century Excellent Talents in University (NCET-13-0521), the Fundamental Research Funds for the Central Universities, and the National Key Technology R&D Program (2012BAH25B01, 2012BAH94F01, and 2012BAH94F03).

REFERENCES

[1] Kiukkonen, N., J. Blom, O. Dousse, D. Gatica-Perez, and J. Laurila, "Towards rich mobile phone datasets: Lausanne data collection campaign". ICPS 2010.

[2] Zhaohui Wu, G.P., SmartShadow: Models and Methods for Pervasive Computing. Springer. 2013.

[3] Donnie H.Kim, J.H., Ramesh Govindan, Deborah Estrin, "Discovering Semantically Meaningful Places from Pervasive RF-Beacons". UbiComp 2009.

[4] Isaacman, S., R. Becker, and e. al., "Identifying important places in people's lives from cellular network data". IEEE Pervasive Computing, 2011: p. 133-151.

[5] Raul Montoliu, J.B., Daniel Gatica-Perez, "Discovering places of Interest in everyday life from smartphone data". Multimedia Tools and Applications, 2013. 62(1): p. 179-207.

[6] Zhongqi Lu, Y.Z., Vincent W. Zheng, Qiang Yang, "Next Place Prediction by Learning with Multiple Models". The Mobile Data Challenge 2012 Workshop by Nokia 2012.

[7] Anna Monreale, F.P., Roberto Trasarti, "WhereNext: a Location Predictor on Trajectory Pattern Mining". KDD 2009.

[8] Richard Becker, R.C., karrie Hanson, Sibren Isaacman, Ji Meng Loh, Margaret Martonosi, James Rowland, Simon Urbanek, Alexander Varshavsky, Chris Volinsky, "Human mobility

characterization from cellular network data". Communications of the ACM, 2013. 56(1): p. 74-82.
[9] Francesco Calabrese, M.D., Giusy Di Lorenzo, Joseph Ferreira Jr, Carlo Ratti, "Understanding individual mobility patterns from urban sensing data: A mobile phone trace example.". Transportation research part C: emerging technologies, 2013. 26: p. 301-313.
[10] Wagner, D.T., A. Rice, and A.R. Beresford, "Device Analyzer: Understanding smartphone usage". MOBIQUITOUS 2013.
[11] Francesco Calabrese, L.L., Carlo Ratti, "Estimating Origin-Destination Flows Using Mobile Phone Location Data". IEEE Pervasive Computing, 2011. 10(4): p. 36-44.
[12] Ying Zhu, Y.S., Yu Wang, "Nokia Mobile Data Challenge: Predicting Semantic Place and Next Place via Mobile Data". Mobile Data Challenge by Nokia Workshop 2012.
[13] Trinh Minh Tri Do, D.G.-P., "The Places of Our Lives: Visiting Patterns and Automatic Labeling from Longitudinal Smartphone Data". IEEE Transactions on Mobile Computing, 2014. 13(3).
[14] Fahad Alhasoun, A.A., Kael Greco, Riccardo Campari, Anas Alfaris, Carlo Ratti, "The City Browser: Utilizing Massive Call Data to Infer City Mobility Dynamics". UbiComp 2014.
[15] Marta C. Gonza´lez, C.s.A.H., Albert-La´szlo´ Baraba´si, "Understanding Individual Human Mobility Patterns". Nature, 2008. 453(5): p. 779-782.
[16] Murat Ali Bayir, M.D., Nathan Eagle, "Mobility profiler: A framework for discovering mobility profiles of cell phone users". Pervasive and Mobile Computing 2010. 6(4): p. 435-454.
[17] Gang Pan, G.Q., Wangsheng Zhang, Laurence T. Yang, Shijian Li, Zhaohui Wu, "Trace Analysis and Mining for Smart Cities: Issues, Methods, and Applications". IEEE Communications Magazine, 2013. 51: p. 120-126.
[18] Pablo Samuel Castro, D.Z., Chao Chen, Shijian Li, Gang Pan, "From Taxi GPS Traces to Social and Community Dynamics: A Survey". ACM Computing Surveys, 2013. 46(2).
[19] Xiaolong Li, G.P., Zhaohui Wu, Guande Qi, Shijian Li, Daqing Zhang, Wangsheng Zhang, Zonghui Wang, "Prediction of Urban Human Mobility Using Large-Scale Taxi Traces and Its Applications". Frontiers of Computer Science, 2012. 6(1): p. 111-121.
[20] Josh Jia-Ching Ying, W.-C.L., Vincent S. Tseng, "Mining Geographic-Temporal-Semantic Patterns in Trajectories for Location Prediction". ACM Transactions on Intelligent Systems and Technology 2013. 5(1): p. 2(1-33).
[21] Katayoun Farrahi, D.G.-P., "What Did You Do Today? Discovering Daily Routines from Large-Scale Data". Mobile Data Challenge by Nokia Workshop 2012.
[22] Ye, Y., Y. Zheng, Y. Chen, J. Feng, and X. Xie, "Mining individual life pattern based on location history". MDM 2009.
[23] Yves-Alexandre de Montjoye, C.s.A.H., Michel Verleysen,Vincent D. Blondel, "Unique in the Crowd: The privacy bounds of human mobility". Scientific Reports, 2013. 3.
[24] Vu, L., Q. Do, and K. Nahrstedt, "Jyotish: A novel framework for constructing predictive model of people movement from joint wifi/bluetooth trace". PerCom 2011.
[25] Ramakrishnan Srikant, R.A., "Mining Sequential Patterns: Generalizations and Performance Improvements". Extending Database Technology 1996.
[26] Jiawei Han, J.P., Behzad Mortazavi-Asl, Qiming Chen, Umeshwar Dayal, Mei-Chun Hsu, "FreeSpan: Frequent Pattern-Projected Sequential Pattern Mining". KDD 2000.
[27] Zaki, M.J., "SPADE: An Efficient Algorithm for Mining Frequent Sequences". Machine Learning, 2001. 40: p. 31-60.
[28] Pei, J., J. Han, and e. al., "Mining sequential patterns by pattern-growth: The prefixspan approach". TKDE, 2004. 16(11): p. 1424-1440.
[29] Ester, M., H.-P. Kriegel, J. Sander, and X. Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise". KDD 1996.