Research on Home and Work Locations Based on Mobile Phone Data

Xusen Luo\textsuperscript{1*}, Yunyao Zhou\textsuperscript{1}, Yifu Yang\textsuperscript{1}, Shuyun Wu\textsuperscript{2}

\textsuperscript{1}School of Information Engineering, Wuhan University of Technology, Wuhan, Hubei, 430070, China
\textsuperscript{2}School of Economics and Management, Hunan Institute of Science and Technology, Yueyang, Hunan, 414000, China
\textsuperscript{*}Corresponding author’s e-mail: 1157872156@qq.com

Abstract. The home and work location information of residents is of great reference value for applications such as transport planning and urban development, since home and workplace are the most important places in people’s lives. The article introduces a method that uses mobile phone call detail records (CDR), which offers good potential for monitoring of the short-term mobility of populations, combined with city road network data, to identify respondents’ home and work locations. The method we developed contains three sections, which are data preprocessing, data fusion and location detection, and is tested with 6 months’ data collected in Hong Kong, including over 350 thousand residents. In the end, experiment results are compared with population register data and the results got in references, and it turns out that our method performs better with the average accuracy of 80%.

1. Introduction
In the dimension of time and space, human movements are highly regular, because of daily life activities. People usually spend most of their time at only a few of significant locations, such as home, workplace and a number of others. Moving between these locations also exhibit regularity. In urban area, distribution of significant locations and collective mobility of the residents pretty much define the spatial structure and activity patterns of the city. To many application domains including urban planning, transportation, and business site selection, that information is quite valuable [1].

Important places of a person can be identified by ranking the frequency and duration of visits at each location [1-2]. At now stage, mobile phone data is one of the most comprehensive source of location data for the requirement of tracking the individual throughout the day for a long period. For each-time exchanging information activities with base stations, the mobile network operator keeps a Call Detail Record (CDR) containing the type of call, phone numbers, time, duration, and also cell-tower ID that is associated with geolocation coordinates [1-4]. The major advantages of using CDR for collecting location data include large user base and ubiquitous usage. A survey reported that the smart mobile usage in Hong Kong, China is as high as 99.1\% of the population by the middle of 2019[5]. Comparing with CDR, other location-tracking methods including location-aware mobile applications and GPS equipped vehicles usually have many limitations of usage.

Recently, a number of studies have been conducted on the use of mobile phone data. Hoteit et al. studied on how to obtain instantaneous OD (Origin-Destination) matrix and transfer them to trajectories [6]. Based on the algorithm that transfers trajectories extracted from mobile phone data to
stay-point data, Garrido et al. studied on the ODs in different duration of the day (morning, afternoon and other times), then they analyzed the type of trip, such as home-to-work trip and home-to-market trip[7]. For the purpose of detecting home and work location, Lumpsus et al. used CDR data to analyze trajectories and made research on stay points [8]. Yan et al. first defined a fixed travel pattern, such as the home-work-shop-home circle, then tried to extract the most likely locations to fit it[9]. Chakirov et al. detected the home and work activities and their locations from public transport smart card data in Singapore using simple rule-based method as well as advanced discrete choice modelling approach [10].

However, these previous works may have some problems. Base station positioning has some errors, this may be even worse when base station has lower density in space, which means using positions of base stations to represent people’s is not accurate. For example, in some area, the distance between base stations may reach to several kilometers. What’s more, a common flaw in all these studies is the lack of a process to scale up the results from mobile phone users to the level of city dwellers as a whole, which is a crucial step in sampling studies [12-13].

Thus, to decrease base station positioning errors, we use CDR data combined with city road network data, fusing trajectories extracted from CDR with real road. Our study area is Hong Kong, in which 6-month CDR is collected including over 350 thousand residents. In summary, the contributions of our work are as follows

- An improved method for detecting trajectories from CDR and city road network data.
- An improved method for identifying home and work locations by considering regularity extracted from trajectories.
- Comparison between our experiment results and references’.

2. Methodology

This study mainly focuses on usage during workdays. In actual, trajectory is a group of special and temporal data, which consists of a series of discrete points. Typically, in two-dimensional geographic space, it can be represented as

\[ \{(x_1, y_1, t_1), (x_2, y_2, t_2), \ldots, (x_n, y_n, t_n) | (x_i, y_i) \in \text{flat space}, \]

\[ t_1 < t_2 < \ldots < t_n, i \in \text{positive integer} \]  

, then according to the definition of stay point, its constraint condition can be represented as

\[ \left( (x_j - x_i)^2 + (y_j - y_i)^2 \right)^{1/2} \leq \Delta D \]

\[ t_j - t_i \geq \Delta T \]  

, where \( \Delta D \) is distance threshold value and \( \Delta T \) is time threshold value.

In CDR, each record contains a cell-tower ID, which can be used to obtain the position of corresponding base station. Initially, we use the position of base station to indicate the people’s
position. Thus, in one day trajectory of person $i$ is represented as $sequ_i$.

$$sequ_i = \{l_{1i}, l_{2i}, \ldots, l_{ni}\}$$  \hspace{1cm} (3)

where $l_{ki}$ is the $k$th position. Then the next step is to obtain the real position of people, in order to decrease base station positioning errors.

$$sequ_i = \{l_{1i}, l_{2i}, \ldots, l_{ni}\} \Rightarrow sequ_i = \{rl_{1i}, rl_{2i}, \ldots, rl_{ni}\}$$  \hspace{1cm} (4)

The steps of improved method for detecting trajectories are as follows.

**Step 1.** Put all of the initial trajectories into the set $S$ and road into set $R$ and assign $R$ to $R_t$.

$$S = \{sequ_1, sequ_2, \ldots, sequ_k\} \quad \text{and} \quad R = \{R_1, R_2, \ldots, R_n\}$$  \hspace{1cm} (5)

**Step 2.** Select the first element from $S$ and assign it to $sq$, select the first element from $R_t$ and assign it to $R_{temp}$, then remove the first element both of $S$ and $R_t$.

**Step 3.** Calculate the center point $c$ of $sq$, select $n$ points of $R_{temp}$.

$$c = center(sq) = \left( \frac{\sum_{i=1}^{n} x_i}{n}, \frac{\sum_{i=1}^{n} y_i}{n} \right) = (x_c, y_c)$$  \hspace{1cm} (6)

**Step 4.** Calculate the distance deviation $\Delta d$.

$$\Delta d = \sum_{i=1}^{n} \frac{\pi}{180} \cdot R_e \cdot \left( \frac{(x_c - x_i)^2 + (y_c - y_i)^2}{n} \right)^{1/2}$$  \hspace{1cm} (7)

where $R_e$ is the radius of the earth.

**Step 5.** Assign the second element of $R_t$ to $R_{temp}$ and also remove it from $R_0$, then repeat step 3 and step 4 and cycle until the size of $R_t$ equals 0. Then calculate the minimum value of $\Delta d$, $\Delta d_{min}$, whose corresponding road is the real trajectory of $sq$, donated as $P_r$.

$$P = \{P_{r1}, P_{r2}, \ldots, P_{rn}\}$$  \hspace{1cm} (8)

**Step 6.** Return to step 2 and cycle until the size of $S$ equals to 0.

After that, we analyzed the regularity of these trajectory. It turns out that more than 70% residents’ movements are highly regular, each of which are always happened between two fixed locations during workdays. As a result, we can use the regularity to identify the home and work locations of residents.

---

Figure 2. The heat map of movements.  
Figure 3. The number of movements per hour per day.
To identify the home and work locations with the calculated trajectory data, steps of the method we developed are as follows.

**Step 1.** Group trajectory data by residents and ODs.

\[
groupDis(group(P, u), OD) = \left\{ \begin{array}{l}
    u_1, \{od_1, \{P_{11}, P_{22}, \ldots \} \}, \ldots \\
    u_2, \{od_2, \{P_{21}, P_{22}, \ldots \} \}, \ldots \\
    \ldots 
\end{array} \right. \tag{9}
\]

where \(group(P, u)\) means to group \(P\) by each resident, and \(groupDis()\) means to calculate the distance of origins and destinations of each two grouped trajectories, if the distance is less than threshold value, then divide the two trajectories into one group. In this paper, we select 800 meters as the threshold value.

**Step 2.** Calculate the number of each group of trajectories to get the tow groups with the maximum value of each resident during 6:00 to 10:00 and during 17:00 to 22:00.

**Step 3.** Compare the two groups. If their origins and destinations are inverse, the origin and destination of trajectory in the morning are home and work locations.

### 3. Experiment And Results

The proposed methods were applied to the CDR and Hong Kong city road network data with all of the phone numbers encrypted. As a result, home and workplaces can be identified for 268734 residents and identification rate is 75.08%. In the study, residents can be divided into two types, work type and no-work type. The work type’s movements have a high level of regularity, going to work from home in nearly each morning and returning home from workplace in almost each afternoon or evening during workdays. The no-work type’s daily movements are less regular, which makes it much harder to identify their home and work locations only using mobile phone data.

Then we compare our method with population register data and reference [6] and reference [8].

| Method                        | Identified Users | Percentage of Accuracy |
|-------------------------------|------------------|------------------------|
| Population Register Data      | 96.94%           | 35.95%                 |
| Reference [6]                 | 66.22%           | 62.03%                 |
| Reference [8]                 | 74.93%           | 74.11%                 |
| Our Method                    | 75.08%           | 80.12%                 |

### 4. Conclusions

This paper proposes methods for identifying home and work locations from mobile phone CDR and road network data. Firstly, to detect residents’ real trajectories, we use mobile phone CDR combined with road network data, aim at decrease base station positioning error. And next step is analyzing the regularity of these trajectories. Finally, we identify home and work locations using the regularity. As the result shows, population register data has the highest Identification rate and the lowest accuracy, which also takes the longest time. And our method performs well with the highest accuracy.

With the development of technology, the popularity of smart phones is increasing rapidly. There is a continuous source of data to enable planners and researchers to gain a better understanding of residents’ movements. Although the proposed method was firstly designed to solve the home and work locations identification problem, it has potential applications outside of the domains studied here. In the case that our method assumes that people go to work in public workdays, which is not appropriate, we will conduct a further study in the future.

### Acknowledgments

This research is supported by Shenzhen Institute of Advanced Technology Chinese Academy of Science and Wuhan University of Technology.
References

[1] S. Çolak, L. P. Alexander, B. G. Alvim, S. R. Mehndiratta, and M. C. González. (2015) Analyzing cell phone location data for urban travel: current methods, limitations, and opportunities. J. Transportation Research Record: Journal of the Transportation Research Board, 2526: 126–135.

[2] Z. Duan, L. Liu, S. Wang. (2011) MobilePulse: Dynamic profiling of land use pattern and OD matrix estimation from 10 million individual cell phone records in Shanghai. In: 19th International Conference on Geoinformatics. Shanghai. pp. 1–6.

[3] M. Yin, M. Sheehan, S. Feygin, J. F. Paiement, A. Pozdnoukhov. (2017) A Generative Model of Urban Activities from Cellular Data. J. IEEE Transactions on Intelligent Transportation Systems, 99: 1–15.

[4] M. S. Iqbal, C. F. Choudhury, P. Wang, M. C. González. (2014) Development of origin–destination matrices using mobile phone call data. J. Transportation Research Part C: Emerging Technologies, 40: 63–74.

[5] Hong Kong China news agency. (2019) The popularizing rate of smart phone in Hong Kong. http://www.ccpit.org/Contents/Channel_4113/2019/0709/1185439/content_1185439.htm.

[6] Sahar Hoteit, Stefano Secci, Stanislav Sobolevsky, Carlo Ratti, Guy Pujolle. (2017) Estimating human trajectories and hotspots through mobile phone data. J. Computer Networks, 64: 296–307.

[7] Graells Garrido E, Caro D, Parra D. (2018) Inferring trajectories of citizens using mobile phone data. J. epidata science, 7: 49.

[8] Tongsinoot Lumpsum, Muangsin Veera. (2017) Exploring Home and Work Locations in a City from Mobile Phone Data. In: 19th IEEE Intl Conference on High Performance Computing and Communications. Bangkok. pp. 56-59.

[9] Z. Yan, C. Parent, S. Spaccapietra, D. Chakraborty. (2010) A hybrid model and computing platform for spatio-semantic trajectories. In: 7th Extended Semantic Web Conference, ESWC. Heraklion. pp. 60–75.

[10] A. Chakirov, A. Erath. (2012) Activity identification and primary location modelling based on smart card payment data for public transport. In: 13th International Conference on Travel Behaviour Research. Jaipur. pp. 15-20.

[11] S. Phithakkitnukoon, T. Horanont, G. Di Lorenzo, R. Shibasaki, C. Ratti. (2010) Activity-aware Map: Identifying Human Daily Activity Pattern Using Mobile Phone Data. In: Proceedings of the First International Conference on Human Behavior Understanding. Berlin. pp. 14–25.

[12] P. Yang, T. Zhu, X. Wan, X. Wang. (2014) Identifying Significant Places Using Multi-day Call Detail Records. In: 2014 IEEE 26th International Conference on Tools with Artificial Intelligence. Limassol. pp. 360–366.

[13] Y. Zheng, Y. Liu, J. Yuan, X. Xie. (2011) Urban Computing with Taxicabs. In: Proceedings of the 13th International Conference on Ubiquitous Computing. New York. pp. 89–98.