Mobility Prediction Based on POI-Clustered Data

Haoyuan Chen[1,2], Yali Fan[1,2], Jing Jiang[1,2], and Xiang Chen[1,2,3](✉)

[1] School of Electronics and Information Technology, Sun Yat-sen University, Guangzhou 511400, China
chenxiang@mail.sysu.edu.cn
[2] Key Lab of EDA, Research Institute of Tsinghua University in Shenzhen (RITS), Shenzhen 518075, China
[3] Starway Communications Inc., Guangzhou 510663, China

Abstract. Predicting users' mobility trajectories is significant for service providers, such as recommendation systems for tourist routing, emergency warning, etc. However, the former researchers predict the next location merely by observing the past individual trajectories, which usually performs poor in the accuracy of trace prediction. In this paper, POIs (Points of Interest) information is used to adjust the weight parameters of the predicted results, and the rationality and precision would be improved. The cellular towers are firstly classified into seven types of functional area through POIs. Then the target user's next possible functional area could be speculated, which acts as a supervision of the ultimate prediction outcome. We use the DP (Dirichlet Process) mixture model to identify similarity between different users and predict users' locations by leveraging these similar users. As is shown in the results, the methods proposed above are highly adaptive and precise when being utilized to predict users' mobility trajectories.

Keywords: Point of interest · Clustering
Mobility trajectory prediction

1 Introduction

It has been a hot topic of study for human behaviors and locations. Some government workers initially used census data to roughly locate the scope of citizens' movement [1], which cost a lot of time and money. As a result, its practicability is imprecise. Some researchers used the phone call records to obtain individual locations [2]. This approach, disappointingly, is seldom put into practice because of high sparsity between two neighboring phone calls, which results in low accuracy of users' location prediction during the time without making phone calls. The current study mainly focuses on cellular network data [3], which possesses the following characteristics: (1) frequent accesses within minutes; (2) improved spatial granularity as a result of the development of cellular

L. Meng and Y. Zhang (Eds.): MLICOM 2018, LNICST 251, pp. 60–72, 2018.
https://doi.org/10.1007/978-3-030-00557-3_7

infrastructure; (3) easily available records which are saved once when interactions between phones and bases occur. Therefore, analyzing cellular network data or wireless network data facilitates the study of individual trajectories.

Many researchers have conducted researches on user trajectories based on cellular network data or mobile network data. For instance, Bagrow et al. studied the collective response to emergencies [4,5]. Shibasaki et al. predicted the collective movement of people in rare crowding incidents [6]. Shimosaka et al. took advantage of Bilinear Poisson regression model to predict population movement in cities [7]. Some methods of trajectory prediction are roughly similar [8–11], which merely use the target user's personal history trajectories to predict the user's next most likely position. Long-time phone data from users is applied to obtain the high accuracy. However, it is hard to collect such big amount of data in reality, let alone users may conceal their data due to privacy protection, which results in insufficient data quantity and low prediction accuracy. McInerney et al. [12] mainly paid attention to resemblance in users' temporal patterns with the help of Bayesian mobility model. Jeong et al. [13] proposed an improved algorithm named Cluster-Aided Mobility Predictor (CAMP), which predicts the next most likely position of a target user using the historical trajectories of all users within a certain range. As an advance processing, this algorithm relies on the clustering techniques to discover the similarity between the user behavior profiles from the training dat. The limitation of this algorithm is that the data source is single because it only used the mobile trajectories of users drawing out from cellular network data. Meanwhile, the method of judgment is also single when selecting the best prediction outcome. Therefore, another dimension's data may be needed in order to aid the prediction.

Points of interest (POI) gathered from the map reflects certain socioeconomic activity and functional attributes, such as restaurants, playgrounds, schools, etc. The function type [3] of a certain cell can be obtained by dealing with the POI information in the cell. Furthermore, human behaviors are also closely related to these cells' functions. Recently, researchers have conducted numerous studies on POI information. [14] and [15] provided personal recommendations to users based on the POIs of the places users have visited. Carmo et al. [16] solved the problem of overlapping symbols in POI visualization. Therefore, the mobility trajectories of the user among base stations can be interpreted as the mobility trajectories among functional areas, which indicates that this dimension of data regarding functional areas is suitable to assist the trajectory prediction. Combining the prediction of the user's next functional area with the prediction of the next base station, the rationality and accuracy of the prediction result can be promoted.

Our contributions consist of two aspects:

- To the best of our knowledge, it is the first time to use POIs to categorize bases and study users' trajectories in terms of base station types. This step is crucial in our algorithm since it resolves a challenging problem. Since different users' set of locations vary a lot, i.e., the range of one person's movement have little overlap with that of another person's movement, discovering parallelism

among users is difficult. However, if we map bases into 7 functional types of bases, the similarity of human mobility pattern is easier to identify.
- In addition, it is the first time that the CAMP algorithm is applied to users' mobility trajectories among bases as well as trajectories in terms of base types. Judging users' locations according the prediction of the user's next functional area and that of the next base station, the prediction result could be more accurate and credible.

The rest of this paper is structured as follows. In Sect. 2, the collection of POI data and processing methods of attaining corresponding function type of each base station are introduced. In Sect. 3, we transform users' trajectories by mapping the base station into base station types and predict the next possible base station type of a certain user through DP mixture model. The prediction results and discussions are given in Sect. 4. After showing our prediction results and discussing our ideas for future work in Sects. 5, the related works from three aspects are displayed in Sect. 5. We conclude this paper in Sect. 6.

2 POI Clustering

In this section, the collection of POI data and processing methods of attaining corresponding function type of each base station will be explained in detail.

2.1 TF-IDF Processing

TF-IDF, a statistical method, is utilized to reflect the importance of a word in a document. The importance of a word is measured from two aspects. On one hand, the importance of a word is in proportion to the frequency of the word appearing in the document. On the other hand, it is inversely proportional to the frequency of the word appearing in the corpus. TF (Term Frequency), refers to the frequency of occurrence of a word in all words in a file; IDF (Inverse Document Frequency) is the reverse file frequency, which means the logarithm of ratio between the total number of files and the number of files containing the specific word. The idea of IDF is to accentuate the significance of specific words to categorize documents, and to reduce the importance of commonly used terms in documents. In other words, if a word frequently appears in a document, the term may be a commonly used word and is slight when distinguishing different documents. If a word frequently appears only in one or a few documents, then the word is likely to be a jargon, in other words, a "label" in a specific field. Therefore, IDF is a measure of the universal importance of words and reflects the effectiveness of words to distinguish documents.

The value of TF-IDF is the product of TF and IDF, which can be perceived as the adjustment of TF by taking IDF as the weight. The purpose of this method is to highlight important words and deemphasize the secondary words.

POI data can reflect the function of an area and can be attained from API mapping service providers. However, an area may contain multiple types of POIs which makes it confusing. Therefore, preprocessing the data is highly necessary. Among the POI data utilized in this study, there are mainly 21 kinds of POIs, including: food, hotel, shopping, entertainment, sports, schools, attractions, tourism development, finance, office buildings, companies, shopping malls, factories, industrial areas, science and technology parks, economic development areas, high-tech development areas, residential areas, living services, townships and villages. Then these kinds of POIs can be grouped into 7 categories based on the classic functional area classification [17,18]: residential, recreational, commercial, industrial, educational, scenic and suburban areas as showing in Table 1.

Table 1. The POI categories and taxonomies

Number	Function	Type
1	Residential	Residential areas, living services
2	Recreational	Food, hotel, shopping, entertainment, sports
3	Commercial	Finance, office buildings, companies, shopping malls
4	Industrial	Factories, industrial areas, science and technology parks, economic development zones, high-tech development zones
5	Educational	Schools
6	Scenic	Attractions, tourism development
7	Suburban	Townships and villages

In order to correctly measure the importance of a POI within a cell area, TF-IDF is performed to process the classified POI information. In the calculation process, for a given area unit $a \in A$, where A refers to the set of all regional units, the number of POIs in each POI category is counted, and then we can calculate the POI vectors: $[TF-IDF_1^a, TF-IDF_2^a, \ldots, TF-IDF_i^a, \ldots, TF-IDF_7^a]$, where $TF-IDF_i^a$ represents the TF-IDF value of the i-th POI in region a. It can be calculated as follows:

$$IDF_i^a = log(\frac{A}{R}), \tag{1}$$

$$TF-IDF_i^a = n_i^a \cdot \frac{IDF_i^a}{N^a}, \tag{2}$$

where R represents the number of regional units that contain the i-th POI in A; n_i^a is the number of POIs contained in the i-th POI category in unit a; N^a refers to the total number of POIs in unit a.

The corresponding TF-IDF Vector for each area is obtained, which contains a total of 7 attribute values, and each value represents the importance of its corresponding function in this area.

2.2 K-Means Clustering

The K-Means method is one of the most widely used partition-based clustering algorithms. Its basic idea is as follows: firstly, a K value is selected, which represents the number of clustering centroid points; secondly, the data points in space are allocated to the same category when their nearest centroids are the same according to Euclidean Distance; thirdly, according to the clustering result, the position of the centroid of each cluster is updated, and the data points in the space are redivided to generate K new clusters. The iteration continues until the centroid position is no longer changed.

After clustering TF-IDF Vectors of all cell area units, then we map the area units into seven functional types, we can get "base station type" of each cell area unit. The clustering results are shown in Fig. 1. As can be observed from the figure, the number of area units marked as "entertainment" is the largest, accounting for 41.78% of all area units; the total proportion of area units marked as "residential" and "entertainment" is 68.84%. This clustering result indicates that the functions of the region we study are biased toward housing and entertainment.

Next, the base stations appearing in the cellular network data can be mapped to the 7 base station types. The track of the users based on the base station types can be generated, which may reflect the daily mobility patterns of users. For example, the track of a user based on the base station types might be: industrial area - commercial area - entertainment area - residential area. For all users, their trajectories for base station type may be analogous.

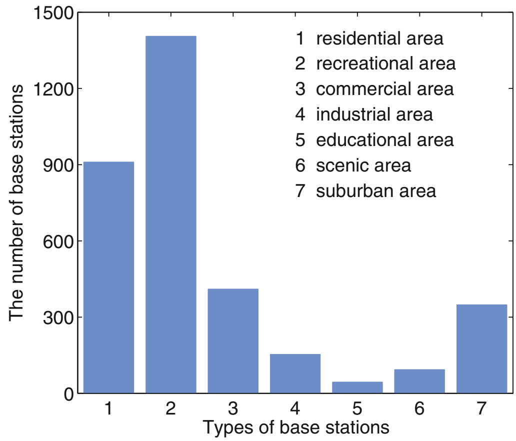

Fig. 1. The results of clustering POI of base stations

3 Trajectory Prediction Method Based on Dirichlet Process Mixture Model

Our goal is to predict the next location for a certain user at a given time on the basis of POI data as well as all users' past trajectories. The set of users is denoted by U. Assuming that the set of locations in all trajectories is $\mathcal{L}$, and its size is L. The trajectory of a user u is denoted by $x^u = (x_1^u, \ldots, x_{n^u}^u)$, where x_t^u means the t-th location that user u visits and n^u means the total length of the track. $x_{n^u}^u$ is where u currently locates. Since we are studying the patterns of users' behavior moving from one base to another, we impose $x_t^u \neq x_{t+1}^u$. In other words, two consecutive locations on the user's trajectory must be different.

Next additional notations are introduced as follows,

$$n_{i,j}^u = \sum\nolimits_{t=1}^{n^u-1} \mathbb{I}(x_t^u = i, x_{t+1}^u = j), \tag{3}$$

where $n_{i,j}^u$ represents the number of transitions for user u from location i to location j.

Similarly,

$$n_i^u = \sum\nolimits_{t=1}^{n^u} \mathbb{I}(x_t^u = i). \tag{4}$$

We assume that a user's trajectory is a order-1 Markov chain. The user's trajectory is drawn from the transition kernel $\theta^u = (\theta_{i,j}^u)_{i,j \in L} \in [0,1]^{(L \times L)}$ where $\theta_{i,j}^u$ means the probability that user u shift from i to j. Thus, the probability of observing trajectory of user u is as follows,

$$P_{\theta^u}(x^u) := \prod\nolimits_{t=1}^{n^u-1} \theta_{x_t^u, x_{t+1}^u}^u. \tag{5}$$

We assume that the transition kernels of different users are independently generated from the same distribution μ. In other words, users' trajectories are drawn from the hierarchical model: for all $u \in U$, $\theta^u \sim \mu$, $X^u \sim P_{\theta^u}$.

Our model is applicable to predict locations of users with rather short trajectories since similarities between different users' mobile habits are taken into account. As we know, a few analogous kernels may generate many different users' trajectories. That is, the distribution μ might be composed of a few clusters. Our goal is to find these user clusters and to predict individual trajectory based on all users' trajectories from the same cluster. Our data includes trajectories of 1000 users, while the total number of base stations is more than 3000. The number of users is smaller than that of base stations. In order to discover existing parallelism among users, we transform users' trajectories by mapping the base station into base station types. We can better characterize the similarity between different users.

Bayesian nonparametric inference [19] is adopted in this paper since the number of clusters is not obtainable before clustering. The number of clusters is flexible which could increase as input data grow. Thus, the number of clusters is a posteriori parameter which is updated in the computation procedure.

In this model, we approximate the distribution μ of transition kernels by computing g iteratively. When computing g, Dirichlet Process (DP) mixture model which is applied in the CAMP algorithm proposed by Jaeseong et al. is adopted. Refer to the work of Jeong et al. [13] for a more exhausted description of the CAMP algorithm. The procedure of the algorithm is outlined in Algorithm 1.

The algorithm of Gibbs Sampler is used to attain independent samples of the allocation of users to clusters. The algorithm of Update DP algorithm is used to update two parameters of DP mixture model, α and G_0, according to the results of Gibbs Sampler [20]. These two algorithms are shown as Algorithms 2 and 3 respectively.

4 Simulation Results and Discussions

Our experimental dataset is the trajectories of 1000 users during one week from a metropolis in China provided by the Ministry of Education-China Mobile Research Fund "DPI & Pipeline Big Data". There are totally 3363 base stations appearing in the trjectories of these 1000 users. According to the locations of these base stations, we first ultilize Voronoi polygons to obtain their coverage area as is shown in Fig. 2 (left). We refer to the base station coverage area obtained in this way as the Voronoi area of the base station. Specifically, for any Voronoi area, the Euclidean distance of any point in it from its base station is always closer than that from other base stations. Then we calculate the area

---

Algorithm 1. CAMP

Input: x^U, K, B, M
Output: $\hat{\theta}^u, \hat{x}^u$

Step 1: Updates of G_0 and α
function
    $G_0^1 \leftarrow Uniform\left(\Theta\right), \alpha_1 \leftarrow 1$
    for $k = 0 \ldots K-1$ **do**
        for $b = 1 \ldots B$ **do**
            $c^{U,b,k} \leftarrow GibbsSampler\left(x^U,\ G_0^k,\ \alpha_k,\ M\right)$
        end for
    end for
end function
Step 2: Last sampling and prediction
function
    for $b = 1 \ldots B$ **do**
        $c^{U,b,k} \leftarrow GibbsSampler\left(x^U,\ G_0^k,\ \alpha_k,\ M\right)$
    end for
    Compute $\hat{\theta}^u$ and $\hat{x}^u$
    $\theta^u = \frac{1}{\mathrm{B}}\sum_{b=1}^{B} E_g[\overline{\theta}^{c^{u,b,K}} | x^{c^{u,b,K}}] = \frac{1}{\mathrm{B}}\sum_{b=1}^{B} \frac{\int_\theta \theta\cdot P_\theta(x^{c^{u,b,K}})G_0^K(d\theta)}{\int_\theta P_\theta(x^{c^{u,b,K}})G_0^K(d\theta)}$
    $\hat{x}^u = arg\max_j \hat{\theta}^u_{x^u_{n^u},j}$
end function

Algorithm 2. Gibbs Sampler

Input: $x^U, G_0,\ \alpha,\ M$
Output: c^U

function
  $\forall \mathrm{u} \in \mathrm{U}, c^u \leftarrow c_1,\ n_{c_1,\ -u} \leftarrow |U| - 1; N \leftarrow 1;\ c^U = \{c_1\}$
  for $i = 1 \ldots M$ **do**
    for u $\in$ U **do**
      $c^u \leftarrow c^u\ \{u\}$
      $\beta_{new} \leftarrow z \frac{\alpha}{\alpha+|U|-1} \int_\theta P_\theta(x^u) G_0(d\theta)$
      $\beta_c \leftarrow z \frac{n_{c,-u}}{\alpha+|U|-1} \int_\theta P_\theta(x^u) G_0(d\theta|x^c), \forall c \in c^{U\setminus\{u\}}, G_0(d\theta|x^c) = \frac{\int \theta_{i,j} P_\theta(x^u)\mu(d\theta)}{\int P_\theta(x^u)\mu(d\theta)}$
      In the above expressions, z is a normalizing constant selected to satisfy:
      $\beta_{new} + \sum \mathrm{c} \in c^{U\setminus\{u\}} \beta_c = 1;$
      With probability β_{new} do:
      $c_{N+1} \leftarrow \{u\}; c^U \leftarrow c^U \cup \{c_{(N+1)}\}; N \leftarrow N + 1;$
      With probability β_c do:
      $c^u \leftarrow c; c \leftarrow \mathrm{c} \cup \{\mathrm{u}\};\ n_{c,\ -v} \leftarrow n_{c,\ -v} + 1, \forall v \neq u.$
    end for
  end for
end function

Algorithm 3. Update DP

Input: $x^U,\ G_0^k,\ \{\mathrm{c}^{\mathrm{U,b,k}}\}_{b=1,\ldots,B}$
Output: $G_0^{k+1},\ \alpha_{k+1}$

Compute $G_0^{k+1}(.)$ and α_{k+1} as follows
function
  $G_0^{k+1}(.) = \frac{1}{B} \sum_{b=1}^{B} \sum_{c \in c^{\mathrm{U,b,k}}} \frac{n_{c,b,k}}{|U|} G_0^k(.|x^c)$

  $\alpha_{k+1} = \arg\min_{\alpha \in R} \left| \sum_{i=1}^{|U|} \frac{\alpha}{\alpha+i-1} - \frac{1}{B} \sum_{b=1}^{B} N_b \right|$

end function
where $n_{c,b,k}$ is the size of cluster $c \in c^{(U,b,k)}$, and N_b is the total number of (non-empty) clusters in $\mathrm{c}^{\mathrm{U,b,k}}$.

covered by each base station and also get the corresponding Cumulative Distribution Function (CDF)) curve as is shown in Fig. 2 (right). From Fig. 2 (right), we can see that 50% base stations cover less than 0.78 square kilometers. By analyzing the user trajectories, we attain the transition kernels for each user. Then the probability distribution that all users move among different base stations is obtained as is shown in Fig. 3. As can be observed, both residential area and entertainment area are visited most frequently when individuals shift among base stations. In particular, both educational and scenic areas have a high degree of relevance to entertainment area. When predicting the base station type of the next location of 596 users by cluster-aided mobility prediction algorithm, 310 users are predicted accurately and the accuracy rate reaches 52%. These results are higher than previous research on mobility prediction.

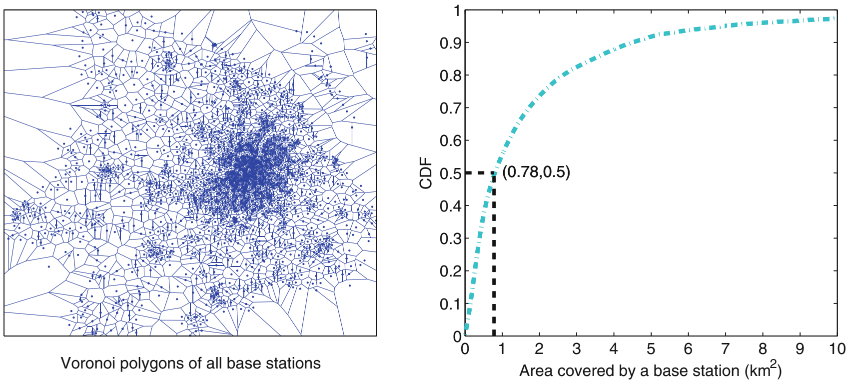

Fig. 2. All base stations' coverage area obtained through Voronoi diagram (left) and the corresponding CDF curve (right)

Issue concerning large population is one of the most challenging problems for the society today. Recently, the accelerated urbanization has exacerbated urban problems. Researches about urban population may bring about new insights for urban problems. It is known that the distribution of population in cities within 24 h may vary greatly. It is of great significance to foresee the aggregation of the population concerned in advance since the corresponding early warning mechanism [21]. which has been put forward will be more effective.

Xu et al. used the cellular network data to estimate Real-time population [3]. Their idea is to use the number of users connected to base stations sampled at given time. Then a specific model between the sampled population and the actual population is established. In the model put forward, the source of the sample data is exactly the number of connections detected by the base station at that moment. However, if a user's device is not connected to the base station at the sampling time, but actually they are in the base station area, they are omitted. The location of users who does not interact with base stations may be estimated, taking advantage of the trajectory prediction method proposed in this paper. Based on our research, historical cellular data for all users and POI information can assist the prediction on where a particular user may appear in the future. Therefore, for those who are leaved out at the sampling time, we can predict their most likely position. Fusing this forecast data with the sample data together as the observed data can increase its proximity with the actual population in the sample data, and the accuracy of the population distribution estimation may be promoted. Given that all of these data sources are real-time, predicting the population's dynamic distribution is highly practical.

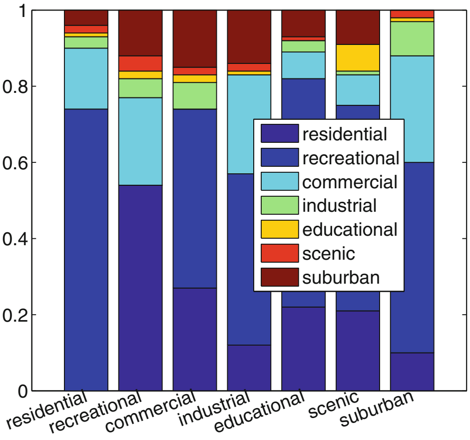

Fig. 3. The transition probability distribution of users from different types of bases

5 Related Work

With the development of the communication industry and the proliferation of smartphones, an abundance of temporal and spatial information regarding users' location are accessible to researches and are intensively studied by researches, which have brought about fruitful insights on human mobility behavior. Related work from three aspects will be introduced as followings.

5.1 Various Data Types Generated from Smartphones

A variety of data types can be generated from a smartphone, e.g. GPS traces, connection records generated from Apps, cellular data collected by base stations, etc. Fang et al. detected popular user mobility patterns by transforming the GPS trajectories into POI trajectories [22]. Wirz et al. predicted crowd density and crowd velocity before serious crowding evens occurred with a dataset of volunteers who reported their surrounding environment periodically utilizing smartphones [23]. Noulas et al. successfully predicted certain users' activities by formulating their communication patterns using their check-in data generated from a special App [24]. Calabrese et al. proposed a new real-time system to estimate urban traffic using a dataset of network bandwidth usage records [25]. These data above needs to be collected from certain volunteers through Apps or GPS sensors on their smartphones, which costs a great deal of time and money, resulting in small number of attendees and low reliability of human mobility patterns. By contrast, the real cellular data could be collected by base stations without installing additional hardware or software on smartphones and they can provide almost continuous trajectories of enough users with large amounts of information, which makes human mobility patterns exploiting credibly possible.

5.2 Researches Based on Trajectories of Users

As more user trajectories are accessible to researches, exploration of human's mobility pattern and recommendation of interesting places based on their trajectories are boosted. Isaacman et al. [26] manage to discover important places of a person according to the individual trajectory. Fan et al. [6] speculate users' collective movement when encountering unusual crowd incident. Zheng et al. [27] built a tree-based hierarchical graph (TBHG) model to recommend interesting locations to users. Yoon et al. [28] propose an algorithm that takes advantage of multiple users' trajectories in a city to provide itinerary tour routing for those who are not familiar with the city.

5.3 Intensive Study on Human Mobility Prediction

As for the methods of trajectory prediction, Zonoozi et al. [29] developed a mathematical formulation to track mobile movement systematically. Calabrese et al. [30] put forward an idea that utilizing geographical features of collective behavior such as land use, POI could be effective regarding mobility prediction.

6 Conclusion

In this paper, POI information of the cellular towers is taken into consideration to adjust the weight parameters of the predicted results, and the rationality and precision are improved significantly. The POIs information is utilized to divide cellular towers into seven categories, which makes it achievable to predict the target users' next possible function area. The DP mixture model is applied to identify similarity between different users, and the resemblance of the users is used to predict the users' locations precisely. The experimental results verify the performance of our proposed scheme, and the proposed method is highly adaptive and precise when being utilized to predict users' mobility trajectories.

Acknowledgements. The work is supported in part by the NSFC (No. 61501527), Science, Technology and Innovation Commission of Shenzhen Municipality (No. JCYJ20170816151823313), Guangdong Innovative and Entrepreneurial Research Team Program (No. 2013D014), State's Key Project of Research and Development Plan (No. 2016YFE0122900-3), the Fundamental Research Funds for the Central Universities, Guangdong Science and Technology Project (No. 2016B010126003), and 2016 Major Project of Collaborative Innovation in Guang-zhou (No. 201604046008).

References

1. Perangkaan, M.J.: 1980 population and housing census of Malaysia. General report of the population census, vol. 1, Kuala Lumpur, Malaysia (2005)
2. Deville, P., et al.: Dynamic population mapping using mobile phone data. Proc. Nat. Acad. Sci. U.S.A. **111**(45), 15888–15893 (2014)

3. Xu, F., Zhang, P., Li, Y.: Context-aware real-time population estimation for metropolis. In: 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing, Heidelberg, pp. 1064–1075. ACM (2016)
4. Bagrow, J., Wang, D., Barabasi, A.L.: Collective response of human populations to large-scale emergencies. PLoS ONE **6**(3), e17680 (2011)
5. Bengtsson, L., Xin, L., Thorson, A., Garfield, R., Schreeb, J.V.: Improved response to disasters and outbreaks by tracking population movements with mobile phone network data: a post-earthquake geospatial study in Haiti. PLoS Med. **8**(8), e1001083 (2011)
6. Fan, Z., Song, X., Shibasaki, R., Adachi, R.: CityMomentum: an online approach for crowd behavior prediction at a citywide level. In: 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing, Osaka, pp. 559–569. ACM (2015)
7. Shimosaka, M., Maeda, K., Tsukiji, T., Tsubouchi, K.: Forecasting urban dynamics with mobility logs by bilinear poisson regression. In: 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing, Osaka, pp. 535–546. ACM (2015)
8. Chon, Y., Shin, H., Talipov, E., Cha, H.: Evaluating mobility models for temporal prediction with high-granularity mobility data. In: 2012 IEEE International Conference on Pervasive Computing and Communications, Lugano, vol. 25, pp. 206–212. IEEE (2012)
9. Song, L., Kotz, D., Jain, R., He, X.: Evaluating next-cell predictors with extensive Wi-Fi mobility data. IEEE Trans. Mob. Comput. **5**(12), 1633–1649 (2006)
10. Scellato, S., Musolesi, M., Mascolo, C., Latora, V., Campbell, A.T.: NextPlace: a spatio-temporal prediction framework for pervasive systems. In: Lyons, K., Hightower, J., Huang, E.M. (eds.) Pervasive 2011. LNCS, vol. 6696, pp. 152–169. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-21726-5_10
11. Jacquet, P., Szpankowski, W., Apostol, I.: An universal predictor based on pattern matching: preliminary results. IEEE Trans. Inf. Theory **48**(6), 1462–1472 (2002)
12. Mcinerney, J., Zheng, J., Rogers, A., Jennings, N.R.: Modelling heterogeneous location habits in human populations for location prediction under data sparsity. In: ACM International Joint Conference on Pervasive and Ubiquitous Computing, Zurich, pp. 469–478. ACM (2013)
13. Jeong, J., Leconte, M., Proutiere, A.: Cluster-aided mobility predictions. In: IEEE INFOCOM 2016-The 35th Annual IEEE International Conference on Computer Communications, San Francisco, vol. 18, pp. 1–9. IEEE (2016)
14. Yao, Z., Fu, Y., Liu, B., Liu, Y., Xiong, H.: POI recommendation: a temporal matching between POI popularity and user regularity. In: 2016 IEEE 16th International Conference on Data Mining, Barcelona, pp. 549–558. IEEE (2017)
15. Zhang, Z., Liu, Y., Chen, H., Liu, Q.: POI recommendation with geographical and multitag influences. In: International Conference on Behavioral, Economic and Socio-Cultural Computing (BESC), Durham, pp. 1–6. IEEE (2017)
16. Carmo, M.B., Afonso, A.P., Ferreira, A., Claudio, A.P., Silva, G.: PoI awareness, relevance and aggregation for augmented reality. In: 2016 IEEE 20th International Conference on Information Visualisation, Lisbon, pp. 300–305. IEEE (2016)
17. Yuan, J., Zheng, Y., Xie, X.: Discovering regions of different functions in a city using human mobility and POIs. In: 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Beijing, pp. 186–194. ACM (2012)
18. Mimno, D., Mccallum, A.: Topic models conditioned on arbitrary features with dirichlet-multinomial regression. In: 2008 24th Conference on Uncertainty in Artificial Intelligence, arXiv. 1206.3278v1, pp. 411–418 (2012)

19. Ferguson, T.S.: A Bayesian analysis of some nonparametric problems. Ann. Stat. **1**(2), 209–230 (1973)
20. Gershman, S.J., Blei, D.M.: A tutorial on Bayesian nonparametric models. J. Math. Psychol. **56**(1), 1–12 (2012)
21. Gonzalez, M.C., Hidalgo, C.A., Barabasi, A.L.: Understanding individual human mobility patterns. Nature **453**(7196), 779–782 (2008)
22. Fang, Z., Ma, C., Wang, X., Qu, J.: Mining popular mobility patterns from user GPS trajectories. In: 2016 9th International Conference on Service Science, Chongqing, pp. 180–181. IEEE (2016)
23. Wirz, M., Franke, T., Roggen, D., Mitleton-Kelly, E., Lukowicz, P., Troster, G.: Inferring crowd conditions from pedestrians' location traces for real-time crowd monitoring during city-scale mass gatherings. In: 2012 IEEE 21st International Workshop on Enabling Technologies: Infrastructure for Collaborative Enterprises, Hammamet, pp. 367–372. IEEE (2012)
24. Noulas, A., Mascolo, C., Frias-Martinez, E.: Exploiting foursquare and cellular data to infer user activity in urban environments. In: 2013 IEEE 14th International Conference on Mobile Data Management, Milan, vol. 1, pp. 167–176. IEEE (2013)
25. Calabrese, F., Colonna, M., Lovisolo, P., Parata, D., Ratti, C.: Real-time urban monitoring using cell phones: a case study in Rome. IEEE Trans. Intell. Transp. Syst. **12**(1), 141–151 (2011)
26. Isaacman, S., et al.: Identifying important places in people's lives from cellular network data. In: Lyons, K., Hightower, J., Huang, E.M. (eds.) Pervasive 2011. LNCS, vol. 6696, pp. 133–151. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-21726-5_9
27. Zheng, Y., Zhang, L., Xie, X., Ma, W.Y.: Mining interesting locations and travel sequences from GPS trajectories. In: 18th International Conference on World Wide Web, Madrid, pp. 791–800. ACM (2009)
28. Yoon, H., Zheng, Y., Xie, X., Woo, W.: Smart itinerary recommendation based on user-generated GPS trajectories. In: Yu, Z., Liscano, R., Chen, G., Zhang, D., Zhou, X. (eds.) UIC 2010. LNCS, vol. 6406, pp. 19–34. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-16355-5_5
29. Zonoozi, M.M., Dassanayake, P.: User mobility modeling and characterization of mobility patterns. IEEE J. Sel. Areas Commun. **15**(7), 1239–1252 (1997)
30. Calabrese, F., Lorenzo, G.D., Ratti, C.: Human mobility prediction based on individual and collective geographical preferences. In: 2010 13th International IEEE Conference on Intelligent Transportation Systems, Funchal, pp. 312–317. IEEE (2010)