A Novel Method for Activity Place Sensing Based on Behavior Pattern Mining Using Crowdsourcing Trajectory Data

Wei Yang[1], Tinghua Ai[1], Wei Lu[1], Tong Zhang[2]

[1]School of Resource and Environment Sciences, Wuhan University, Wuhan, China
Email: ywgismap@whu.edu.cn; tinghuaai@whu.edu.cn; whuluwei@whu.edu.cn

[2] State Key Laboratory of information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan, China
Email: zhangt@whu.edu.cn

1. Introduction

Human activity places such as sports place and english corner are closely related to daily life of urban residents. For example, jogging place as an important place for exercise and recreation can provide convenience for residents to relieve pressure, reduce weight, prevent obesity and enhance physical fitness (Cook 2016). Therefore, extracting the information of activity place and establishing corresponding geographic information services (e.g., inquiry, retrieval, recommendation, etc.) play an important role in smart city construction, urban infrastructure management and place-based GIS (Miller 2007, Liu 2015, Jenkins 2016). However, spatial information are mainly traditionally obtained from field surveys, remote sensing. These methods are good at obtaining urban static information (such as land use and land cover change), which are difficult to discover and extract activity place in a real time and low cost way due to activity places are time-varying and dynamic. Consequently, using a low cost and fast method to extract activity place information including spatial, human activity, semantic automatically is very significant.

In the era of big data, Volunteer Geography Information (VGI) data (e.g., trajectory, social media data) are available and open a new horizon for human mobility analysis, behaviour pattern mining, place sensing and urban studies (Hu 2015, Liu 2015). Particularly, spatio-temporal trajectory data contain rich information about activities and places, which has become a hot topic of behaviour pattern mining and activity place sensing (Yuan 2017, Yang 2018). Using massive trajectory data to extract activity place information, it is necessary to mine behavior patterns and identify behaviour activities at the individual level, and to discover and extract the place information at the collective aggregation level (Liu 2015, Yang 2018). Meanwhile, it needs to establish a general framework for place sensing based on trajectory behavior pattern mining, which can provide location service based on activity place information. To accomplish this, taking cycle sports place extraction from jogging GPS trace data as a case, this work proposed a novel method to extract activity place based on trajectory behavior pattern mining. Our main contributions are twofold: (1) From a methodological perspective, this paper contributes a mining framework for activity place sensing based on trajectory behavior pattern mining. (2) From an application perspective, this work proposes a new approach to mine cycle periodic behavior pattern and extract jogging sports place for location service, jogging place recommendation and jogging path planning.

2. Methodology

2.1 Overall architecture

We develop a mining framework which takes the crowdsourcing trajectory data as the input, and outputs the information of the activity place. Figure 1 shows the overall architecture of the proposed framework.

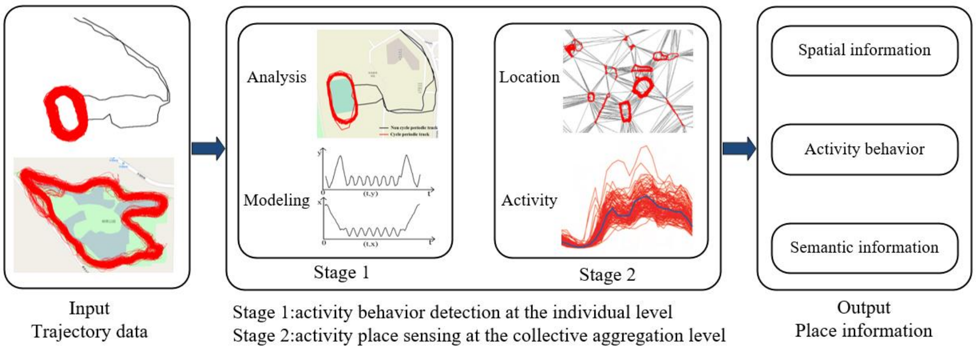


Figure 1. Overall architecture of the proposed framework.

2.2 Input: trajectory data

Based on the technology by which they are recorded, mobility data are available in different forms such as GPS, GNSS, Geo-social network and RFID. In this work, the jogging GPS track data which are obtained by smartphone sports App called Yuedong App (similar to Strava App).

2.3 Stage1: activity behavior modelling and detecting at the individual level

As to trajectory data contains various behaviors, it needs to establish a corresponding behavior model to distinguish the mining behavior from other behaviors and propose an efficient algorithm to extract the behavior sub-tracks by coupling the model. In this work, we modelled the cycle periodic behavior using movement parameters and proposed trajectory distance matrix search (TDMS) algorithm extract cycle periodic tracks from jogging traces coupling the periodic behavior model. This stage includes two steps: First, using movement parameters to model the cycle periodic behavior.

Definition 1: Cycle periodic item: Moving objects do cycle periodic movement repeatedly in the activity region, which is called cycle activity, such as cycle running. A cycle activity is composed of multiple cycle behavior. A cycle behavior corresponds to a cycle sub-track, which is called cycle periodic item and is represented by $SubTraj=\{p_i, p_{i+1},\ldots, p_j\}$, the p_i is GPS point, as shown in Figure 2.

Definition 2: Cycle periodic pattern: Cycle trajectory is composed of a set of cycle periodic items, which is represented by $CycleTraj=\{SubTraj_1, SubTraj_2,\ldots, SubTraj_n\}$. Periodic items are appeared frequently with an approximately equal path distance, time period according to the time sequence, and this pattern is called cycle periodic pattern, as shown in Figure 2.

The cycle periodic behavior is analyzed and modelled using the following trajectory movement parameters:

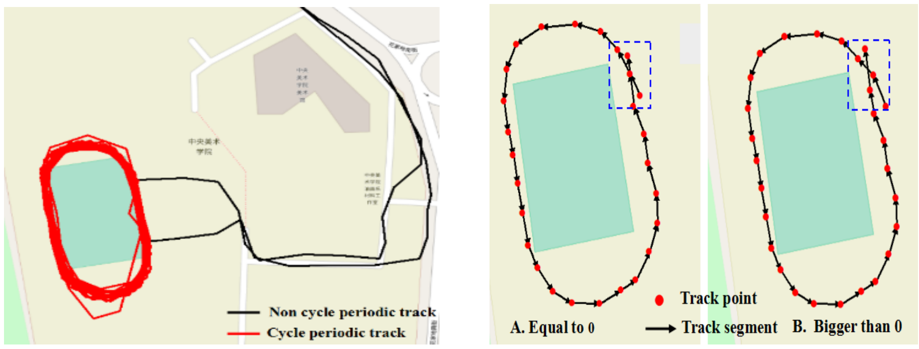
Definition 3: Trajectory path distance: the path distance of a sub-track, which is composed of a sequence of GPS points $T= \{p_m, p_{m+1}, \ldots, p_n\}$, and is denoted by:

$$\mathrm{TrajPathDist}(T) = \sum_{k=m}^{k-1}\left(Dist(p_k, p_{k+1})\right) \quad (1)$$

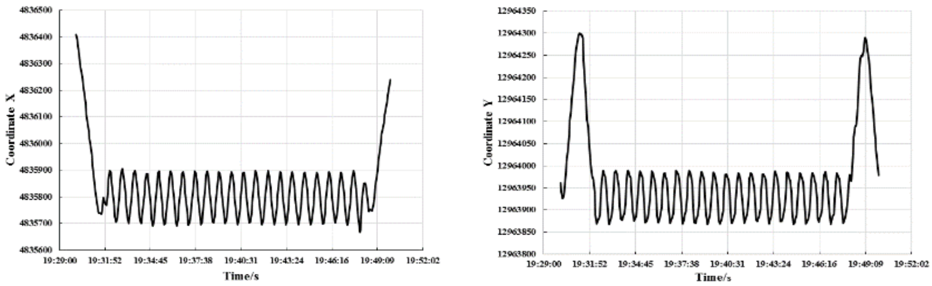
Definition 4: Trajectory direction distance: the direction distance of sub-track $T = \{p_m, p_{m+1}, \ldots, p_n\}$ equals the distance between the first point and the last point in the sub-trajectory and is denoted by:

$$\mathrm{TrajDirDist}(T) = Dist(p_m, p_n) \quad (2)$$

For the cycle periodic item, the corresponding trajectory direction distance is far less than the trajectory path distance. In Figure 2a, the trajectory direction distance is equal to 0 when the starting point of periodic item trace coincides with the end point; conversely, the trajectory direction distance is bigger than 0 and it is less than the distance of the two adjacent sampling track points. Therefore, the direction distance of periodic item trace is less than the distance threshold *maxDirect*, and the *maxDirect* value is the average sampling distance of trajectory. The periodic item has different trajectory path distance according to the different sports place, and the trajectory path distance is bigger than *minPath* (the *minpath* value is 400m in this work), and the path distance difference between adjacent periodic items should be less than 2.*maxDirect*. A cycle behavior should be last for a certain time. If the time duration of a cycle periodic item trace is less than time duration threshold *minTime* (the *minTime* is 6min), the trace segment is filtered as trajectory noise. In additional, the coordinates of the raw track line in Figure 2b are decomposed into two-dimensional time series data ((t, x) and (t, y)). The decomposed curve is composed of peaks and troughs, and the corresponding sub-track between adjacent peaks or troughs is the cycle behavior. So, the cycle periodic behavior detection is converted to the time series data sequence of periodic trajectory coordinate identification and extraction.



(a) Cycle periodic pattern (b) Cycle periodic item sub-trajectory



(c) Time series data (t, x) (d) Time series data (t, y)

Figure 2. Analyzing and modelling cycle periodic behavior.

Second, based on the cycle periodic behavior model, the trajectory distance matrix search algorithm is proposed to extract cycle periodic tracks from each jogging traces. And the main steps of TDMS algorithm are as follows:

Step1, parameter values are defined. The parameter values of *minPath*, *maxDirect* and *minTime* are determined according to the above stated.

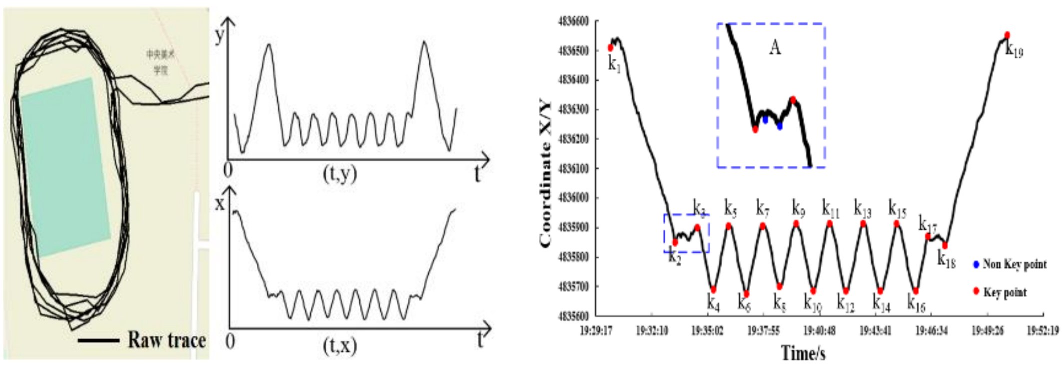
Step2, trajectory coordinates are decomposed into time series. For each track line, the track point coordinates (x, y, t) are decomposed into two dimensional time series data (t, x) and (t, y), as shown in Figure 3a.

Step3, key points extraction. The starting, end, maximum and minimum point of the time series data are taken as the key points. The extracted key points from (t, x) and (t, y) time series are merged to get all the key point set $KP=\{k_1, k_2, \dots, k_n\}$. The original track points corresponding to the key points are extracted according to time sequence as the periodic pattern searching point set $SP=\{s_1, s_2, \dots, s_n\}$.

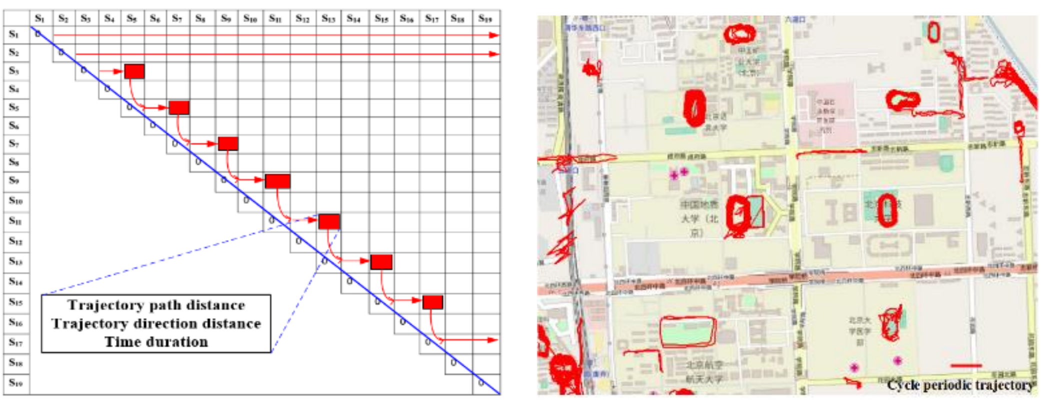
Step4, trajectory distance matrix is constructed. According to time sequence, the trajectory distance triangular matrix is constructed based on the SP, and each matrix unit records the trajectory path distance, the trajectory direction distance and time duration, as shown in Figure 3c.

Step5, cycle periodic sub-tracks are searched. The trajectory distance matrix is searched by scanning the rows. If the three feature values of a matrix unit meet the feature of the cycle periodic item, the sub-track corresponding to the row and column of the matrix unit can be extracted as the cycle periodic item sub-trace, and then the cycle periodic item sub-traces are searched sequentially taking the row of the matrix in which the end point of this sub-traces located as start searching point, as shown in Figure 3c; If the cycle periodic item sub-trace is searched, the distance matrix is searched by depth-first search according to above steps; The trajectory line is processed and the cycle periodic trajectory is outputted until it cannot search the cycle periodic sub-trace, as shown in Figure 3d.

Step6, collective cycle periodic trajectories extraction. The collective cycle periodic trajectories are obtained after processing the all track lines in the database according to the above steps of algorithm, and the algorithm is stopped.



(a) Conversion of track line to time series (b) Finding the KP and SP

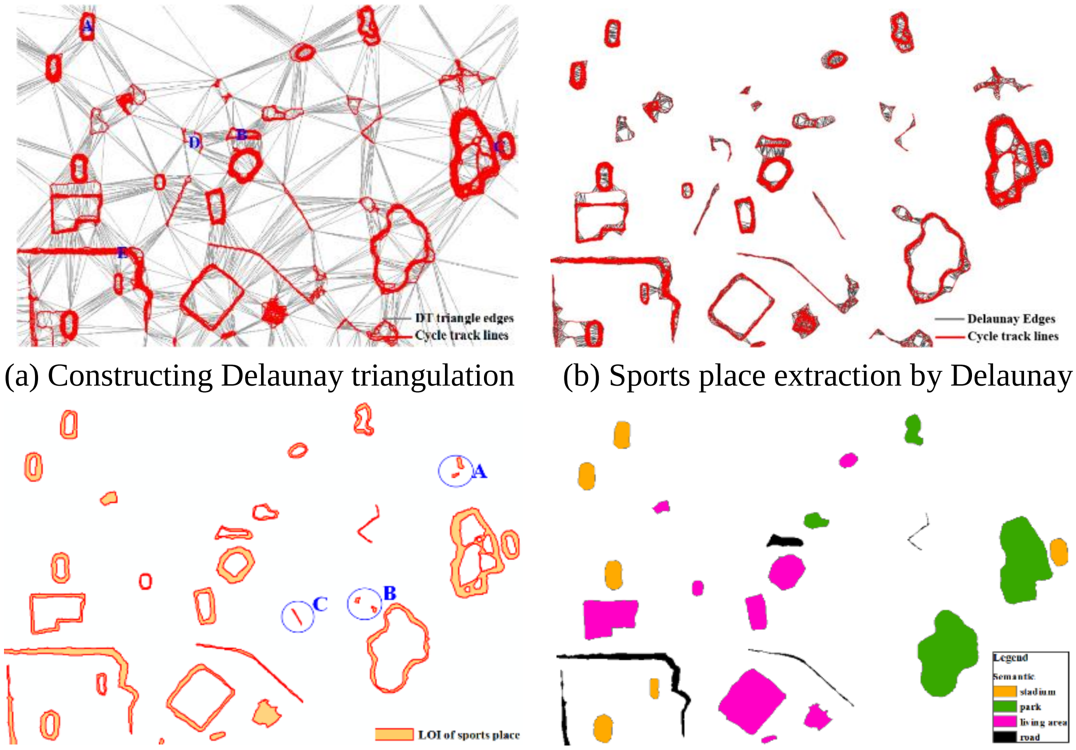


(c) Cycle periodic item search (d) The extraction results

Figure 3. Detecting cycle periodic behavior and extracting periodic trajectory by the TDMS algorithm.

2.4 Stage2: activity place sensing at collective aggregation level

This stage includes two steps: first, the spatial information of cycle sports places is extracted from the collective cycle periodic trajectories by the Delaunay triangulation model (Yang 2018). Then the land use semantic information of the cycle sports places is extracted by the reverse geocoding method. The process of sport place information extraction and the results of the extraction sport place information are as shown in Figure 4. Second, constructing sports place profiles using collective cycle periodic trajectories. The first approach clusters sports places based on when their activity occurs, i.e., how trips made at the sports place are distributed over time. The second approach makes it possible to identify groups of joggers in each sports place that have similar behavior pattern aggregated into weekly profiles.



(a) Constructing Delaunay triangulation (b) Sports place extraction by Delaunay

(c) LOI of sports place extraction (d) Semantic classification of place (AOI)

Figure 4. The process of sport place information extraction.

2.5 Output: spatial service information based on active places

The output information including spatial information of sports place, semantic information and collective activity information. We can use the output activity place information to construct a geography information location service based on place.

3. Experimental and analysis

3.1 Study Area and Study Data

To verify the validity of the proposed approach, jogging trace data in Beijing, China were tested, as shown in Figure 5. Jogging GPS track data is recorded by the sports App installed on smartphone. The 15786 trajectories of the 3623 smartphone sport

App users in 1 month are used to conduct experiment. The jogging trajectory dataset contained the following fields: user ID, latitude, longitude, elevation, time, sports form, and energy. The trajectory sampling interval is 1s-10s, and the average sampling intervals were 4s-5s.

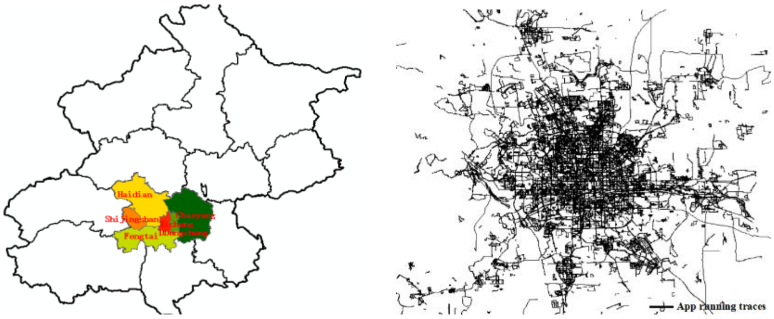
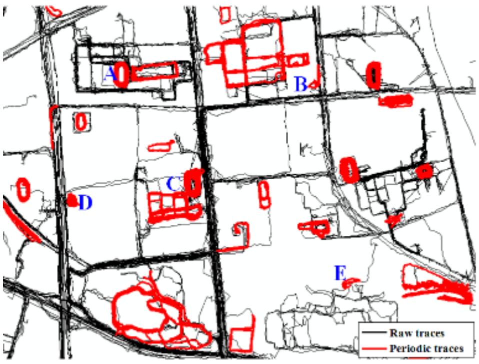
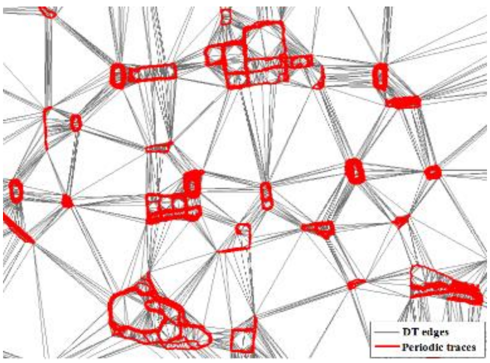


Figure 5. Experimental area and mobile App running track data.

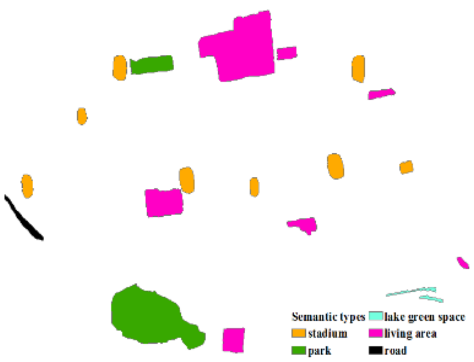
3.2 Experimental results of the extraction sports place

After conventional trajectory data pre-processing, the jogging trace lines are generated according to user ID and time. We use TDMS algorithm to detect cycle periodic behavior and extract cycle trajectory from individual jogging track line, and a total of 8731 cycle trajectories are extracted (partial results as shown in Figure 6a). Then Delaunay triangulation is constructed using collective periodic track lines (Figure 6b), and a total of 573 sports places are extracted (Figure 6c). Finally, we extracted 255 living communities, 139 stadiums, 77 roads, 68 parks, 15 recreational green spaces, 7 squares and 5 lakes green spaces after semantic classification (Figure 6d). The experimental result shows that living areas, stadiums and parks become the main sports places for urban residents jogging leisure activity.





(a) Cycle periodic trajectory extraction (b) Constructing DT using tracks



(c) Semantic classification of sports places (d) The overall results

Figure 6. The process of experimental and experimental results.

4. Conclusion

This research developed a novel framework for activity place sensing based on behavior pattern mining using crowdsourcing trajectory data. Taking sports place extraction as a case, 1 month smartphone App jogging traces in Beijing were used to verify the validity of the novel method.

Acknowledgements

This research was supported by the National Key Research and Development Program of China (Grant No. 2017YFB0503500), and the National Natural Science Foundation of China (Grant No. 41531180).

References

Cook S, Shaw J and Simpson P, 2016, Jography: Exploring meanings, experiences and spatialities of recreational road-running. *Mobilities*, 11(5): 744-769.

Miller H, 2007, Place-based versus people-based geographic information science. *Geography Compass*, 1(3): 503-535.

Liu Y, Liu X and Gao S, 2015, Social sensing: A new approach to understanding our socioeconomic environments, *Annals of the Association of American Geographers*, 105(2): 512-530.

Jenkins A, Croitoru A and Crooks A, 2016, Crowdsourcing a collective sense of place. *PloS one*, 11(4):1-13

Hu Y, Gao S and Janowicz K, 2015, Extracting and understanding urban areas of interest using geotagged photos. *Computers, Environment and Urban Systems*, 54: 240-254.

Yang W, Ai T and Lu W, 2018, A Method for Extracting Road Boundary Information from Crowdsourcing Vehicle GPS Trajectories. *Sensors*, 18, 1261.

Yang W, Ai T, 2018, POI information enhancement using crowdsourcing vehicle trace data and social media data: A Case Study of Gas Station. ISPRS Int. J. Geo-Inf. 7, 178.

Yuan G, Zhao J and Xia S, 2017, Multi-granularity periodic activity discovery for moving objects. *International Journal of Geographical Information Science*, 31(3): 435-462.