From Trajectories to Path Network: An Endpoints-Based GPS Trajectory Partition and Clustering Framework

Hua Yuan[1,*], Yu Qian[1], Baojun Ma[2], and Qiang Wei[3]

[1] School of Management and Economics, University of Electronic Science and Technology of China, 610054 Chengdu, China
{yuanhua,qiany}@uestc.edu.cn
[2] School of Economics and Management, Beijing University of Posts and Telecommunications, 100876 Beijing, China
mabaojun@bupt.edu.cn
[3] School of Economics and Management, Tsinghua University, 100084 Beijing, China
weiq@sem.tsinghua.edu.cn

Abstract. In this paper, we aim to mine the interesting locations and the frequent travel sequences in a given geo-spatial region. Along this line, a new partition method is proposed to divide the trajectories into a set of line segments and the geographical-similar endpoints are clustered into groups to detect the fixed territories. Also, a path network is generated to show the linkage relations between these fixed territories. The proposed method can be used to detect frequent movement paths as well as fixed territories from GPS trajectories efficiantly.

Keywords: GPS trajectory, stationary sub-trajectory, clustering, path network.

1 Introduction

With the widespread usage of miniaturized GPS devices, recording the trace data of moving objects become extremely easy and useful work while it provides enormous business opportunity in geography navigation and recommendation system[1]. During the past years, a bunch of research has been performed based on individual location history represented by GPS trajectories. These works include detecting individual locations [2], recognizing user-specific activities[3,4] and predicting traveler's movement [2]. The trajectory pattern mining problem was introduced in [5]. Following this work, some important efforts had been devoted, such as[6,7,8].

Finding some key information, such as characteristic points[9,10] and representative line[10], is the feasible method to help people extracting useful information from GPS trajectory. However, it is hard to get such information because the recorded GPS data are always non-uniformity, sparse, lost and inconsistency with the endpoints in cases that the users turn on and off the GPS-enabled devices casually. In this paper, we present an endpoints-based GPS trajectory partition and clustering framework to mine users' interested locations and frequent travel sequences in a given geo-spatial region.

* Corresponding author.

F. Li et al. (Eds.): WAIM 2014, LNCS 8485, pp. 740–743, 2014.

2 The Method

2.1 Stationary Sub-trajectory

Let $TR = \{g_1 g_2 ... g_i g_{i+1} ... g_n\}$ denote the *trajectory* of the object moving from g_1 to g_n. For any sub-trajectory $STR = \{g_s g_{s+1} ...... g_t\} \subseteq TR$, $\overrightarrow{g_s g_{s+1}}$ is the *first move action*. Now, suppose the projection of $g_i, (i = s+1, ..., t)$ on $\overrightarrow{g_s g_{s+1}}$ is g_i', then the distance between g_i and g_i' is called *position disturbance* (denoted by $d_{g_i} = |g_i g_i'|$).

A sub-trajectory STR is a *stationary sub-trajectory* (SST) if the position disturbance of each GPS point $g_i, s+1 < i \leq t$ changes not rapidly with respect to the *first move action* $\overrightarrow{g_s g_{s+1}}$ while the moving behavior of g_{t+1} changes rapidly.

2.2 Trajectory Partitioning Method

In fact, the key issue for partitioning a TR into SSTs is to find out all the *characteristic points*[10] in it. In this section, we propose a new trajectory partitioning algorithm which aims at finding the points where the behavior of a trajectory changes rapidly. The main idea is to check the value of $d_{g_i}, i = 2, ..., n$ with respect to the present *first move action* (Algorithm 1).

Algorithm 1. Trajectory Partitioning Algorithm

```
1: Input: Trajectory TR = {g1 g2 g3 ......gn}, threshold d0;
2: Output: A set of characteristic points CP;
3: g1 → CP; i = 1;
4: repeat
5:     first move action= →(gi gi+1);
6:     for j = i + 2 to n do
7:         if d_gj ≥ d0 then
8:             gj−1 → CP; i = j − 1;
9:         end if
10:    end for
11: until gn → CP.
12: return CP.
```

2.3 Characteristic Points Clustering and Path Network

Given N trajectories, the clustering processes can be specified as follows: Firstly, we partition each $TR_i = \{g_{1_i}, ..., g_{n_i}\}, i = 1, ..., N$, into m_i SSTs. Then, we introduce the general clustering method to cluster all the points in $CP = \cup_{i=1}^{N} CP_i$ into l clusters: $C_1, ..., C_l$, based on the Euclidean distance of paired points. The element number of each cluster is $|C_i|, i = 1, ..., l$. Finally, we calculate the centroid point c_i of cluster $C_i, (i = 1, ..., l)$ and use these centroid point to represent each cluster.

The edge between two nodes c_i and c_j $(i \neq j)$ can be used to represent all the possible SSTs whose start point in C_i (or C_j) and end point in C_j (or C_i). In another word, it approximates to the common path between area c_i to c_j. We can construct an undirected *path network* by connecting these fixed territories.

3 Experimental Results

3.1 Efficiency and Preciseness

The following experiments are based on the real GPS datasets[1]. The results about efficiency and preciseness are shown in Figure 1. It indicates that: 1) The efficiency of our method ($d_0 = 1$) is more faster than that of the method presented by [10]. 2) The computation speed will become more faster with increasing value of d_0. 3) The curves in Figure 1 are almost straight lines, these are experimental evidences about the truth that both these two methods have computation complexity of $O(n)$.

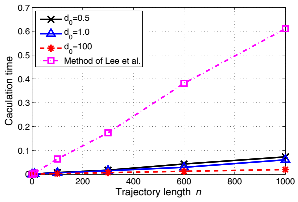


Fig. 1. Comparison of Efficiency.

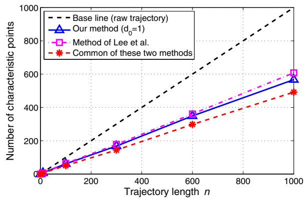


Fig. 2. Comparison of Preciseness.

Figure 2 shows the results about preciseness, the black dash line is the base of raw trajectory. The pink line is the number of characteristic points generated with the method presented by Lee et al., and the blue line is that with our method. The red line presents the number of characteristic points found by these two methods commonly. We can see that, 1) Both the preciseness of these two methods would become worse along with the increase of trajectory length, n. 2) The method presented by Lee et al. has advantage in preciseness because it keeps more GPS points as characteristic points.

3.2 Path Network

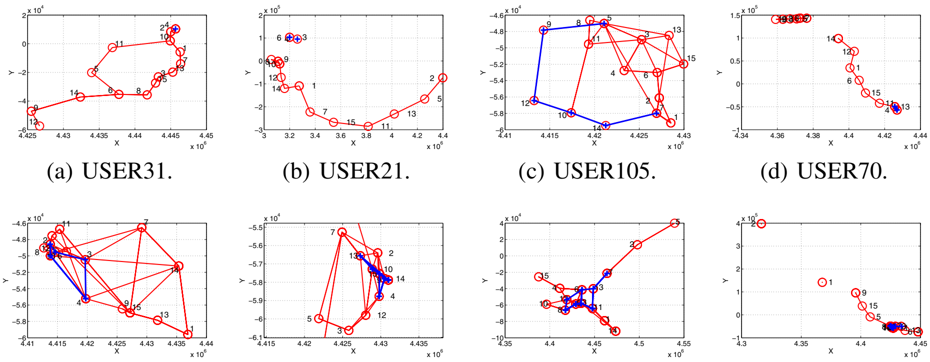
In the following, we choose randomly 8 users' data to conduct the experiments. Their path networks are generated as in Figure 3. Each network is a spatial map of the typical motions for a specific user, and it contains valuable information about: the frequent pathes, the fixed territories, and the movement correlations among representative points.

4 Conclusion

In this work, we try to trim GPS trajectories into *fixed territories* and *frequent path* to generate a spatial map of typical motions, i.e., *path network*, by taking into account of users' historic travel experiences as well as the correlation between locations.

The methods proposed in this work are efficient for mining information hidden in the trajectory data, especially the frequent path, fixed territories and movement intention, which can provide business opportunity in geography information service.

[1] http://research.microsoft.com/en-us/projects/urbancomputing/.



(a) USER31. (b) USER21. (c) USER105. (d) USER70.

(e) USER155. (f) USER5. (g) USER24. (h) USER13.

Fig. 3. Path networks generated for the eight selected users.

Acknowledgements. This work is supported by the NSFC(Nos. 71271044/U1233118/71102055), the SRFDP (20100185120024) and the Fundamental Research Funds for the Central Universities (No. 2014RC0601).

References

1. Zheng, Y., Zhou, X. (eds.): Computing with Spatial Trajectories. Springer, Berlin (2011)
2. Ashbrook, D., Starner, T.: Using gps to learn significant locations and predict movement across multiple users. Personal and Ubiquitous Computing 7(5), 275–286 (2003)
3. Patterson, D.J., Liao, L., Fox, D., Kautz, H.: Inferring high-level behavior from low-level sensors. In: Dey, A.K., Schmidt, A., McCarthy, J.F. (eds.) UbiComp 2003. LNCS, vol. 2864, pp. 73–89. Springer, Heidelberg (2003)
4. Zheng, Y., Xie, X.: Learning location correlation from gps trajectories. In: Proceedings of the MDM 2010, pp. 27–32 (2010)
5. Giannotti, F., Nanni, M., Pinelli, F., Pedreschi, D.: Trajectory pattern mining. In: Proceedings of the 13th ACM SIGKDD, pp. 330–339 (2007)
6. Monreale, A., Pinelli, F., Trasarti, R., Giannotti, F.: Wherenext: a location predictor on trajectory pattern mining. In: Proceedings of the KDD 2009, pp. 637–646 (2009)
7. Ye, Y., Zheng, Y., Chen, Y., Feng, J., Xie, X.: Mining individual life pattern based on location history. In: Proceedings of the MDM 2009, pp. 1–10 (2009)
8. Wei, L.-Y., Zheng, Y., Peng, W.-C.: Constructing popular routes from uncertain trajectories. In: Proceedings of the 18th SIGKDD, pp. 195–203 (2012)
9. Lee, J.-G., Han, J., Li, X., Gonzalez, H.: Traclass: trajectory classification using hierarchical region-based and trajectory-based clustering. Proceedings of the VLDB Endowment 1(1), 1081–1094 (2008)
10. Lee, J.-G., Han, J., Whang, K.-Y.: Trajectory clustering: a partition-and-group framework. In: Proceedings of the 2007 ACM SIGMOD (2007)