Significant Place Discovery from Taxi Trajectory Data

THI THI SHEIN[†1] MAKOTO IMAMURA[†2]
SUTHEERA PUNTHEERANURAK[†1]

Abstract: Extracting meaningful knowledge from trajectory data is a popular topic in the area of trajectory data mining. Due to the development of location acquisition technology, we can collect trajectory data of moving objects (i.e., people, animal, and vehicles). In this research, we will detect a significant place from the vehicle's trajectory data. The significant place represents an abnormal event (i.e. congested area) on the road. Firstly, we extract stay points by applying StayPointDetection algorithm from each vehicle movement trajectory. Each stay point means a geographical region where each moving object stays together for a certain duration. From these stay points, we discover significant places by using the density-based clustering method. The experiments will evaluate based on real taxi data.

Keywords: Clustering, Stay point extraction, Taxi trajectory data, Trajectory data mining

1. Introduction

With the rapid development of location recording devices such as GPS, GSM networks, RFID (Radio Frequency Identification), a large amount of trajectory data has been generated from many moving objects such as people, animals, and vehicles. From these trajectories, we can retrieve useful knowledge that is contributed to many areas such as, traffic monitoring system, road congestion prediction, location recommendation, and moving object behavior analysis.

The trajectory is the historical location record of a moving object along with the time. The trajectory data, also called spatial-temporal data, represented by a series of chronologically ordered points with corresponding timestamps in spatiotemporal space, i.e., Traj_k= {$< p_1, t_1 >, < p_2, t_2 >, \ldots, < p_n, t_n >$} , where each element $< p_i, t_i >$ indicates a moving object is at location p_i at timestamp t_i.

Currently, since most of the countries used the vehicles as the main transportation system, they have still faced traffic congestion problem. If we detect traffic congestion in early, we can avoid the congested road in advance. Thus, many researchers proposed different approaches to detect traffic congestion. However, there still have to improve in accuracy and efficiency.

Therefore, this research focuses on the discovery of significant places from the trajectories of taxicabs. A significant place denotes an unusual event that occurs for a certain time period (i.e., traffic-congested area). In order to detect significant places, stay points were firstly extracted. Then, significant places were discovered by clustering these stay points.

2. Related Work

Luong et al. [1] proposed a density-based clustering method, MKEIDBSCAN, to handle the noise in detecting significant places. Xiao et al. [2] proposed an algorithm for detecting the stay point to remove undesired points. Zheng et al. [3] proposed the gathering pattern to detect the congregation of vehicles over the road. They discovered the gathering in the form of offline (i.e., discovery from the whole database) and online (i.e., incremental discovery structure). Yu et al. [4] proposed a stay point detection algorithm. And then, they used DBSCAN and K-means clustering to detect congested areas. In this research, they only compared the accuracy of stay point extraction methods and still needed to improve accuracy and efficiency in the detection of a congested area.

3. Significant Place Discovery from Taxi Trajectory Data

In our approach, there have two main phases to discover significant places: 1) Stay point extraction 2) Significant place discovery by clustering the extracted stay points. Fig.1 illustrates the process of significant place discovery.

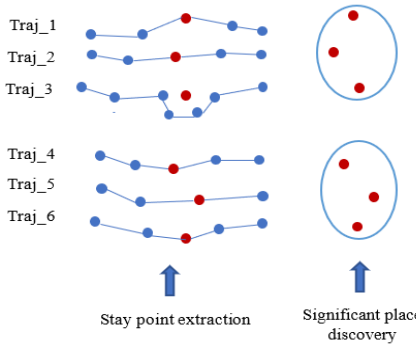


Fig.1 Example of significant place discovery

3.1 Stay Point Extraction

The aim of the stay point extraction from the trajectory is to improve the efficiency and accuracy in the detection of a significant place.

Definition1: (Stay point) A stay point, ***s***, denotes as a geographic region where the vehicle stayed within a given time threshold, θ_t, and a given distance threshold, θ_d. It can be represented as, $\boldsymbol{s} = <latitude, longitude, t_{arrive}, t_{leave}>$.

Fig.2 illustrates the extraction of stay point from a trajectory. There are two situations that happened the stay points. The first

†1 Department of Computer Engineering, Faculty of Engineering, King Mongkut's Institute of Technology Ladkrabang
†2 Tokai University

one is a vehicle that locates at the car parking or stops on the road according to traffic light, like *stay point 1*(P3) in Fig.2. On the other hand, a vehicle is on the state of slow moving according to the vehicle crowded on the road. In this case, a series of points (p5, p6, p7, p8) are involved within a particular spatial region. In such a situation, we need to extract the *stay point 2* by averaging the points as a common spatial region. The pseudo-code of StayPointDetection algorithm was described in [5].

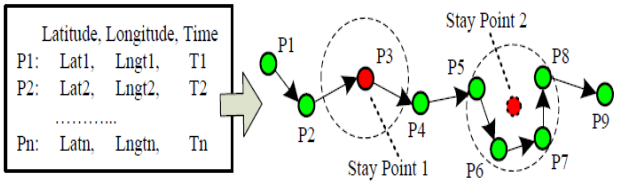


Fig.2 Illustration of stay point extraction from a trajectory

3.2 Significant Place Discovery

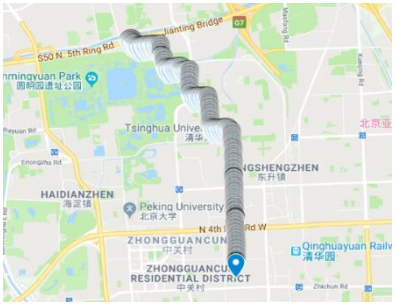
In order to detect a significant place from the stay points, the density-based clustering method (DBSCAN) [4] was applied to retrieve the significant place. The advantage of the density-based clustering method is an efficient clustering algorithm, and it can quickly find the arbitrary shape clusters.

4. Experiment

In this section, we evaluated the experiments for stay point extraction and significant place discovery. We used a real taxi dataset that was collected from the T-Drive project [6].

4.1 Stay Point Extraction

Fig.3 shows the original trajectory point of a vehicle and the result of extracted stay points. In this experiment, the system extracted the stay points by setting, θ_t as 5 minutes, θ_d as 100 meters. The reason was that the sampling frequency of each trajectory in the real taxi data varies from (3~10 minutes).



(a) Original trajectory point

(b) Extracted stay point

Fig.3 Stay points extraction from a vehicle's trajectory

4.2 Significant Place Discovery

For effectiveness, Fig.4 shows a significant place discovery from the stay points that were collected in a single timeframe. The duration of the timeframe was also set as 5 minutes. In this experiment, the parameters for DBSCAN were set as 200 meters for a distance threshold and 5 for a density threshold.

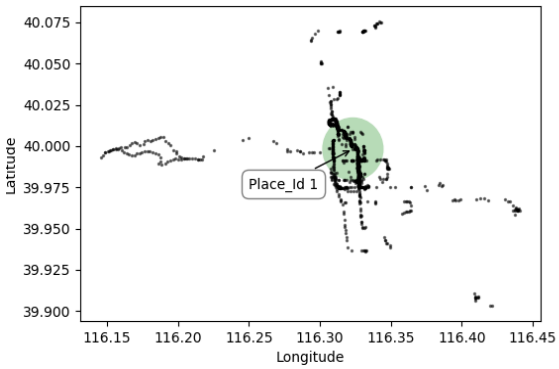


Fig.4 Significant place discovery from vehicles' stay points in a single timeframe

5. Conclusion and Future Work

This paper investigates the significant places (i.e., as congested areas) from the movement of vehicles. But, DBSCAN has still needed to improve the computational time complexity because of the reconstruction of R-tree index in every timestamp. As future work, we will enrich the location with semantic meaning (for example., car parking, shopping complex, road) to improve the accuracy of the system. We will compare the experimental results of other density-based clustering methods (in terms of efficiency and effectiveness).

Reference

[1] Luong, C., Do, S. and Hoang, T., 2015, "A method for detecting significant places from GPS trajectory data". Journal of Advances in Information Technology, 6(1), pp.44-49.

[2] Xiao, H., Wang, W.J. and Zhang, X., 2013, "Identifying the stay point using GPS trajectory of taxis". In Applied Mechanics and Materials (Vol. 353, pp. 3511-3515). Trans Tech Publications.

[3] Zheng, K., Zheng, Y., Yuan, N.J., Shang, S. and Zhou, X., 2013, "Online discovery of gathering patterns over trajectories". IEEE Transactions on Knowledge and Data Engineering, 26(8), pp.1974-1988.

[4] Yu, Q., Luo, Y., Chen, C. and Zheng, X., 2019, "Road Congestion Detection Based on Trajectory Stay-Place Clustering". ISPRS International Journal of Geo-Information, 8(6), p.264.

[5] Li, Q., Zheng, Y., Xie, X., Chen, Y., Liu, W. and Ma, W.Y., 2008, November, "Mining user similarity based on location history". In Proceedings of the 16th ACM SIGSPATIAL international conference on Advances in geographic information systems (p. 34). ACM.

[6] Yuan, J., Zheng, Y., Zhang, C., Xie, W., Xie, X., Sun, G. and Huang, Y., 2010, November, "T-drive: driving directions based on taxi trajectories". In Proceedings of the 18th SIGSPATIAL International conference on advances in geographic information systems (pp. 99-108). ACM.