Accepted Manuscript

Spatial-temporal travel pattern mining using massive taxi trajectory data

Linjiang Zheng, Dong Xia, Xin Zhao, Longyou Tan, Hang Li, Li Chen, Weining Liu

Please cite this article as: L. Zheng, D. Xia, X. Zhao, L. Tan, H. Li, L. Chen, W. Liu, Spatial-temporal travel pattern mining using massive taxi trajectory data, *Physica A* (2018), https://doi.org/10.1016/j.physa.2018.02.064

Spatial-temporal travel pattern mining using massive taxi trajectory data

Linjiang Zheng [a,b*], Dong Xia [a,b], Xin Zhao[a,b], Longyou Tan [c], Hang Li[c], Li Chen[a,b], Weining Liu[a,b]

a. Key Laboratory of Dependable Service Computing in Cyber Physical Society
of Ministry of Education, Chongqing University, Chongqing, China
b. College of Computer Science, Chongqing University, Chongqing, China
c. Chongqing Integrated Transport Hub Development Investment Co., Ltd, Chongqing, China
Corresponding author:zlj_cqu@cqu.edu.cn

Abstract: Deep understanding of residents' travel patterns would provide helpful insights into the mechanisms of many socioeconomic phenomena. With the rapid development of location-aware computing technologies, researchers have easy access to large quantities of travel data. As an important data source, taxi trajectory data are featured by their high quality, good continuity and wide distribution, making it suitable for travel pattern mining. In this paper, we use taxi trajectory data to study spatial-temporal characterization of urban residents' travel patterns from two aspects: attractive areas and hot paths. Firstly, a framework of trajectory preprocessing, including data cleaning and extracting the taxi passenger pick-up/drop-off points, is presented to reduce the noise and redundancy in raw trajectory data. Then, a grid density based clustering algorithm is proposed to discover travel attractive areas in different periods of a day. On this basis, we put forward a spatial-temporal trajectory clustering method to discover hot paths among travel attractive areas. Compared with previous algorithms, which only consider the spatial constraint between trajectories, temporal constraint is also considered in our method. Through the experiments, we discuss how to determine the optimal parameters of the two clustering algorithms and verify the effectiveness of the algorithms using real data. Furthermore, we analyze spatial-temporal characterization of Chongqing residents' travel pattern.

Keywords: Spatial-temporal, attractive area, grid-based clustering, hot path, trajectory clustering

1. Introduction

A travel pattern, as a collective property of travel, is the common behavior generated by various factors during travel. Understanding the residents' travel pattern is critical to path planning[1,2], location-based social networks[3], intelligent transportation systems[4], and urban computing[5–7]. Because of their wide applications in many fields, travel patterns have been attracting many studies. However, due to the limitations of technology for data acquisition, many of the studies were based on survey data, which have many shortcomings, such as high cost, poor timeliness, and low accuracy. With the rapid development of location-aware computing technologies, residents' travel data acquisition is becoming much easier. Due to privacy concerns, most studies on discovering travel patterns so far are based on public transportation data, such as taxi trajectory data[8–10], smart card transaction data for subway[10,11] and bus[10]. In this paper, our research uses the taxi trajectory data.

Taxi is one of the most important travel methods for urban residents. Different from bus, rail, and other means of transportation with fixed routes and stations, taxi can provide convenient and customized