Transportation Research Part B 101 (2017) 1–27

Contents lists available at ScienceDirect

Transportation Research Part B

journal homepage: www.elsevier.com/locate/trb

Scalable space-time trajectory cube for path-finding: A study using big taxi trajectory data

Lin Yang[a], Mei-Po Kwan[b,*], Xiaofang Pan[a,c], Bo Wan[a], Shunping Zhou[a]

[a] Faculty of Information Engineering, China University of Geosciences, 388 Lumo Road, Wuhan 430074, China
[b] Department of Geography and Geographic Information Science, University of Illinois at Urbana-Champaign, 255 Computing Applications Building, MC-150, 605 E Springfield Ave., Champaign, IL 61820, USA and Department of Human Geography and Spatial Planning, Utrecht University, 3508 TC Utrecht, Netherlands
[c] School of Geographic Sciences, Xin Yang Normal University, 237 Nanhu Road, Xin Yang, 464000, China

ARTICLE INFO

Article history:
Received 2 August 2016
Revised 4 January 2017
Accepted 18 March 2017
Available online 29 March 2017

JEL classification:
C63
R41
R42

Keywords:
Path-finding
Road network
Taxi trajectory
Space-time constraint
Driver's experience
Navigation system

ABSTRACT

Route planning is an important daily activity and has been intensively studied owing to their broad applications. Extracting the driving experience of taxi drivers to learn about the best routes and to support dynamic route planning can greatly help both end users and governments to ease traffic problems. Travel frequency representing the popularity of different road segments plays an important role in experience-based path-finding models and route computation. However, global frequency used in previous studies does not take into account the dynamic space-time characteristics of origins and destinations and the detailed travel frequency in different directions on the same road segment. This paper presents the space-time trajectory cube as a framework for dividing and organizing the trajectory space in terms of three dimensions (origin, destination, and time). After that, space-time trajectory cube computation and origin-destination constrained experience extraction methods are proposed to extract the fine-grained experience of taxi drivers based on a dataset of real taxi trajectories. Finally, space-time constrained graph was generated by merging drivers' experience with the road network to compute optimal routes. The framework and methods were implemented using a taxi trajectory dataset from Shenzhen, China. The results show that the proposed methods effectively extracted the driving experience of the taxi drivers and the entailed trade-off between route length and travel time for routes with high trajectory coverage. They also indicate that road segment global frequency is not appropriate for representing driving experience in route planning models. These results are important for future research on route planning or path finding methods and their applications in navigation systems.

1. Introduction

Route planning is an important daily activity. It has been intensively studied due to its broad relevance to many areas of social and scientific concerns (Huang et al., 2007; Zeng and Church, 2009). Substantial attention has been given to the development of efficient path-finding algorithms for navigation or route guidance systems due to the complexity of transportation systems, especially in the fields of transportation and geographic information science

* Corresponding author.
E-mail addresses: yanglin_2002_wh@163.com (L. Yang), mpk654@gmail.com (M.-P. Kwan), xfpanem@163.com (X. Pan), magicwan1105@163.com (B. Wan), zhousunping@mapgis.com (S. Zhou).

http://dx.doi.org/10.1016/j.trb.2017.03.010
0191-2615/

(Chen et al., 2012; Shirabe, 2014; Chen and Nie, 2013; Srinivasan et al., 2014; Zhang et al., 2016). An efficient route not only saves travel or driving time but also reduces energy consumption. This can greatly help both end users and governments to ease traffic problems and protect the environment (Yuan et al., 2011a; Ma et al., 2016; Liu and Qiang, 2016). A good routing or path-finding service should take real-time traffic conditions and human travel behavior into account. However, these attributes are difficult to extract and incorporate into existing routing services (Tang et al., 2010).

Taxi drivers are experienced professional drivers who can usually make near-optimal route choice based on their extensive local knowledge. The routes they use tend to take less travel time and incur lower cost than the routes computed by navigation systems (Ottomanelli and Wong, 2011; Bates et al., 2001). This experience-based and highly efficient driving, which integrates real-time traffic conditions and human travel behavior, is entailed within a huge number of taxi trajectories, which in turn provide researchers with excellent opportunities to learn about the best routes and to develop better methods for supporting dynamic route planning (Li et al., 2011; Yuan et al., 2011a). Indeed, many recent studies recommend and construct travel routes using GPS trajectories, including taxi trajectories, social media, and location-based service (LBS) check-in data (Arase et al., 2010; Wei et al., 2012; Lu et al., 2010; Kurashima et al., 2010).

Drivers' experience is essential for finding satisfactory routes for users of navigation systems, which need to take both real-time traffic conditions (e.g., congestion) and the driving preference of users into account. For the former, research has been conducted to improve the reliability of path finding algorithms through taking into account the uncertainty, stochasticity and dynamics in travel time (Khani and Boyles, 2015; Yang and Zhou, 2014). Probe vehicle data are used to estimate and forecast travel time (Jenelius and Koutsopoulos, 2013). Crowd-sourced data from a fleet of cooperative vehicles and online social media contents are also utilized to detect real-time traffic conditions (Liu and Qiang, 2016). The latter includes preferred road category, intersection avoidance, traffic light avoidance, left turn avoidance, toll fee avoidance, fuel consumption and crime risk. All of these factors are implicated in observed experienced trajectory routes. There are two main types of experience extraction in terms of different utilizations of trajectory. The first one is the factor-based perspective, which aims to predict traffic conditions and model drivers' preference directly (Yuan et al., 2011a; Balteanu et al., 2013; Jenelius and Koutsopoulos, 2013). The other one is the trajectory-based perspective, which identifies experience-based routes (preferred routes) as the routes taxi drivers frequently use to travel from a particular origin to a particular destination (Zeng, 2010; Gong et al., 2007).

From the trajectory-based perspective, there are three main approaches to route planning based on big taxi trajectory data. The first approach is the sub-trajectory growth method. This method detects shared structure among a collection of trajectories by dividing trajectories into smaller segments (decomposition) and then aggregating these segments into clusters to compute routes (composition) (Gudmundsson et al., 2012; Zhou and Huang., 2008; Chen et al., 2013). The paths computed based on the augmented graph naturally incorporate high-order mobility patterns. This method reveals the travel choice of a taxi driver with the finest granularity. However, the path segments identified with this method only reflect the instant movements of taxis and their drivers. They may not reflect the actual origins and destinations of passengers, since the origin and destination of a particular taxi trajectory does not necessarily correspond to the actual origin and destination of the passenger since this depends on where the person got on and got off the taxi.

The second approach extracts experience-based graph from historical trajectory data to compute routes using the shortest path algorithm (Wei et al., 2012; Yuan et al., 2010). Routable graph is constructed according to a specified location sequence and time span. From the node perspective, trajectory points with high space-time correlation are clustered together or frequently used top-k road segments are represented as landmarks to form the graph's nodes, which have higher potential to be included as part of the resulting route. From the edge perspective, edge weight is quantified mainly by the global visit frequency (i.e., the total number of distinct trajectories that traversed it) between two adjacent nodes. Edges with higher visit frequencies will have higher route scores and greater probability to be included as a portion of the planned route. Based on the experience-based graph, an experiential hierarchical graph is also developed to compute routes (Zeng., 2010; Tang et al., 2010; Li et al., 2011; Hu et al., 2013). Travel frequency, travel speed, travel time and centricity have been used to generate connected hierarchical road networks. Statistically significant frequency distribution of taxis over different road segments within different time periods is derived to group road segments into different experience classes.

The above discussion indicates that travel frequency reflects consideration of both real-time traffic conditions and the driving preference based on drivers' experience and plays an important role in experience-based knowledge models in the context of route planning and path finding. Travel frequency is computed by the number of distinct trajectories traversing a road segment and reflects the segment's traffic flow. The interesting phenomenon here is that higher frequency means higher potential to be included as part of the resulting route for traveling from a particular origin to a particular destination. However, whether a road segment is chosen or not should not be decided mainly by global trajectory frequency (i.e., the total number of trajectories traversing a particular road segment).

First, the global trajectory frequency of a particular road segment is not relevant when different origin-destination pairs are considered. Current approaches notice that the value is not a static number. It not only varies between different times but also varies as the origin and destination change. In other words, the trajectory frequency of a particular road segment will be different for trajectories with different origin-destination pairs. However, the global frequency on a road segment does not reflect or utilize the semantic information of the origins and destinations of the actual trips involved. For instance, high trajectory frequency on street A can only reflect the high traffic flow between the two end nodes of this road segment. However, it does not necessarily mean high frequency on street A for trips with a particular pair of origin and destination. For example, Fig. 1(a) shows the global trajectory frequency of each road segment using the width of a green line. Road