Transportation Research Part C 66 (2016) 176–194

Contents lists available at ScienceDirect

Transportation Research Part C

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

Estimating the most likely space–time paths, dwell times and path uncertainties from vehicle trajectory data: A time geographic method

Jinjin Tang [a], Ying Song [d], Harvey J. Miller [b], Xuesong Zhou [c,*]

[a] School of Traffic and Transportation, Beijing Jiaotong University, Beijing 100044, China
[b] Department of Geography, The Ohio State University, Columbus, OH 43210, USA
[c] School of Sustainable Engineering and the Built Environment, Arizona State University, Tempe, AZ 85287, USA
[d] Department of Geography, Environment and Society, University of Minnesota, MN 55455, USA

ARTICLE INFO

Article history:
Received 27 January 2015
Received in revised form 25 July 2015
Accepted 25 August 2015
Available online 9 September 2015

Keywords:
GPS map matching
Traffic state estimation
Dynamic shortest path
Uncertainty estimation

ABSTRACT

Global Positioning System and other location-based services record vehicles' spatial locations at discrete time stamps. Considering these recorded locations in space with given specific time stamps, this paper proposes a novel time-dependent graph model to estimate their likely space–time paths and their uncertainties within a transportation network. The proposed model adopts theories in time geography and produces the feasible network–time paths, the expected link travel times and dwell times at possible intermediate stops. A dynamic programming algorithm implements the model for both offline and real-time applications. To estimate the uncertainty, this paper also develops a method based on the potential path area for all feasible network–time paths. This paper uses a set of real-world trajectory data to illustrate the proposed model, prove the accuracy of estimated results and demonstrate the computational efficiency of the estimation algorithm.

1. Introduction

Emerging mobile computing and sensor techniques have improved capabilities to collect and process real-time traffic data for traffic state monitoring and management applications. For example, Global Positioning System (GPS)-based in car navigation has reached a significant level of penetration rate, and most smart phones are equipped with GPS receivers with high-speed data communication links. Nevertheless, current vehicle location data are still associated with location errors, typically within a wide range of 5–300 m given the ground-truth vehicle trajectory. A critical data processing component in emerging Big Data applications is how to systematically use latitude, longitude, and time stamps of a single probe vehicle or a set of probe trajectories to estimate traffic states at different scales.

In this paper, we present a time-geography based approach (Hägerstrand, 1970) to consider not only the geometry and topology of the road network, but also the time attributes in available GPS samples. We also incorporate a space–time network-based representation adapted from the Time-Dependent Shortest Path problem (TDSP). Essentially, a sequence of GPS traces with both location and timestamp information can be mapped as a space–time trajectory or path. Within a space–time network, there are a large number of possible paths with different degrees of spatial and temporal distance to the vehicle trajectory records. Our approach aims to find the most likely network–time path that minimizes the total

* Corresponding author. Tel.: +1 480 9655827.
E-mail addresses: jinjintang@hotmail.com (J. Tang), yingsong@umn.edu (Y. Song), miller.81@osu.edu (H.J. Miller), xzhou74@asu.edu (X. Zhou).

http://dx.doi.org/10.1016/j.trc.2015.08.014
0968-090X/

map-matching distance among all possible alternatives. Our method also allows estimation of dwell and detour times at intermediate nodes and the uncertainties associated with the likely paths. The proposed mathematical programming model can integrate with various sensor data sources and finds the optimal solution that takes into account the distance measure at different time stamps of driving traces.

This paper is organized as follows. Section 2 provides the background to the trajectory map-matching problem, alternative solutions and the features of our time geographic approach. Section 3 illustrates the proposed space time network representation for finding the most likely network–time paths, followed by an introduction on the time dependent shortest path model in Section 4. Section 5 uses illustrative example to further explain the fundamental of the model. Section 6 presents dynamic programming algorithms to find the time dependent shortest path with generalized cost functions specific to the network–time path estimation problem. After a discussion on uncertainty quantification for potential accessible space–time nodes within the estimated paths in Section 6.3, numerical experiments on a simple network and a real-world network are presented in Section 7. Section 8 concludes the paper with summary comments and future research steps.

2. Background

Many traffic management and planning agencies have started using the vehicle trajectory data for estimating travel time, route choice behavior and activity location patterns. The traffic states of interest include the most likely paths and related traffic speed and density on the transportation network. In this research, we are interested in providing a time-expanded graph modeling framework for fully utilizing location-based data and sensor measurements from different sources, such as automated vehicle counts, virtual detection lines as well as fixed detectors. Within the last 30 years, a large number of algorithms have been proposed to the general problem of GPS map matching, which aims to find the closest or most likely matched link sequence and travel speed. These methods can be broadly categorized in the following.

(i) *Geometric map-matching algorithm*. Geometric map-matching algorithms compare raw GPS points with the geometries of the underlying road network, in order to obtain a sequence of likely links (Greenfeld, 2002). A simple approach along this line is to match each point with the nearest road node (Bernstein and Kornhauser, 1998), while other sophisticated methods involve point-to-curve or curve-to-curve geometric distances (White et al., 2000). As this approach generally does not consider connectivity, it is possible that the matched links are disconnected from one other. To address this issue, Fu et al. (2004) proposed a hybrid map-matching algorithm by examining the geometry of the road network and fuzzy comprehensive judgment. Kong et al. (2013) recently integrated curve-fitting-based method and a vehicle-tracking-based method to estimate traffic states from GPS probe data along a path without detour.
(ii) *Topological map-matching algorithm*. An approach proposed by Greenfeld (2002) aims to find a topologically feasible (but time-invariant) path through the road network, with the arc weights in the related topological path-search algorithm without considering any heading or speed information from GPS data. Meng (2006) further considered other topological features such as road intersections, road curvature and road connections. Some other topological map-matching algorithms (Yin and Wolfson, 2004; Yang et al., 2003) utilize the connectivity and contiguity information of road networks to improve link identification rates in GPS map matching applications.
(iii) *Statistical algorithms*. Honey et al. (1989) first introduced a probability-based algorithm to clearly definite an elliptical or rectangular confidence region around a position. Zhao (1997) suggested that the error region can be derived from the error variances associated with GPS positions. To further quantify and determine map-matching probabilities given noisy data, this type of algorithms have integrated various statistic methods, to name a few, Kalman Filters and Extended Kalman Filters (e.g. Kim et al., 2000; Krakiwsky et al., 1988; Obradovic et al., 2006;Jo et al., 2012), fuzzy logic (e.g. Quddus, 2006; Zhao, 1997; Syed and Cannon, 2004), Bayesian inference (e.g. Pyo et al., 2001), and Particle Filter (Peker et al., 2011). Within an optimal filtering framework, the above algorithms recursively estimate the likely path and error covariance matrix associated with the estimated states under different measurement error assumptions.

Beside these time-invariant algorithms, studies have started to account for travel time beside distance of vehicle trajectories. Aiming to assist informed traffic management decisions, traffic state estimation techniques are often used to estimate end-to-end trip travel time and congestion levels of the traffic system using heterogeneous data sources. Early studies such as Szeto and Gazis (1972), Cremer and Papageorgiou (1981), and recent studies such as Wang and Papageorgiou (2005), Muñoz et al. (2003), Sun et al. (2003) and Work et al. (2010) focus on how to use detector data and various traffic state filtering methods to estimate traffic flow, density and queue lengths on each link segment of the freeway corridor. Herrera and Bayen (2010) also proposed a novel method of virtual trip lines to estimate traffic states based on trajectory data from both arterial and freeway corridors. A recent study by Deng et al. (2013) extended Newell's three-detector method as a stochastic measurement equation to quantify the value of Automated Vehicle Identification (AVI), Automated Vehicle Location (AVL) and point sensor measurements.

While significant progress has been made in recent years, there are still a number of challenging questions to be addressed to systematically estimate vehicle paths within a transportation network based on trajectory data. First, many GPS map-matching algorithms mainly focus on spatial attributes of GPS records within a short time window, in conjunction