Journal of Transport Geography 23 (2012) 72–80

Contents lists available at SciVerse ScienceDirect

Journal of Transport Geography

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

Probabilistic potential path trees for visualizing and analyzing vehicle tracking data

Joni A. Downs [a,*], Mark W. Horner [b]

[a] Department of Geography, Environment, and Planning, University of South Florida, 4202 E. Fowler Ave., Tampa, FL 33620, USA
[b] Department of Geography, Florida State University, 323 Bellamy Building, Tallahassee, FL 32306, USA

ARTICLE INFO

Keywords:
Time geography
GPS
Tracking
Density estimation
Network analysis

ABSTRACT

Vehicle tracking data are often used to explore human travel behavior and activity patterns. Time geography is a useful approach for analyzing such datasets, as it provides a means for identifying the set of possible routes and stops for a vehicle between known locations, which is termed a potential path tree. This research extends the utility of the time-geographic approach by developing a method to generate probabilistic potential path trees that represent the space–time potential of a vehicle's movements. First, this research provides the mathematical formulation of the new technique, network-based time-geographic density estimation (TGDE), and demonstrates the computation using a hypothetical tracking dataset and road network. Its formulation operates as a network adaptation of classical TGDE, which has been previously employed to analyze the movements of objects travelling in continuous, Euclidean space. Second, network-based TGDE is applied in the context of analyzing vehicle tracking data collected by GPS and filtered to protect an individual's privacy. The method was used to map and quantify the vehicle's most likely routes, origins, intermediate stops, and final destinations. The results indicate network-based time-geographic density estimation provides a powerful approach for both geovisualizing and analyzing vehicle tracking data.

1. Introduction

Transportation analysts often use vehicle tracking data to gain insights into human travel patterns. In general, tracking data record the known spatial locations of an object periodically over time. Tracking data for vehicles are typically collected using one of two basic protocols. The first and more traditional method involves geocoding a list of origin and destinations that were recorded from a travel survey, global positioning system (GPS) device, or similar instrument (Axhausen, 2008; Collia et al., 2003; Wolf et al., 2003). In these cases, the vehicle's stops are known but travel paths from one location to the next are not specified. The second approach involves using GPS (Beusen et al., 2009; Wu et al., 2006), cell phone (Liu et al., 2008), or satellite tracking (Cleary, 2000) technologies to automatically record locations whenever a vehicle is in motion and/or the device is active. Depending on the temporal sampling interval, the origins, destinations, and exact travel paths of the vehicle may be either known or unknown. The latter situation is common when wider temporal sampling intervals are used to either extend the battery life of the tracking device (Jiang et al., 2008; Oliver et al., 2010) or to protect an individual's privacy (Abul et al., 2010; Falchuk and Loeb, 2010; Gasson et al., 2011; Iqbal and Lim, 2010). Regardless of how the data are collected, if spatial uncertainty about the vehicle's movements exists, analysts may try to reconstruct its possible or most likely movements and stops for use in further analysis (Chung and Shalaby, 2005).

One approach for modelling the possible movements of an object between known locations is time geography. Classical time geography (Hägerstrand, 1970), particularly as presented by Miller (2005a), may be used to compute the set of locations potentially reachable in Euclidean space by an object given its movement capabilities and space–time constraints imposed by the observed locations. For any two consecutive points, this potential path area takes the shape of an ellipse when mapped in two-dimensional continuous space. However, for vehicles travelling on roadways, the classical Euclidean formulations of time geography are not valid (Raubal et al., 2007), and the network-based framework described in Miller (1991) is necessary. In that case, all possible locations reachable between any two successive points on a network form what is termed a potential path tree. Potential path trees and related time-geographic elements have appeared in a wide variety of transport applications (Shaw, 2006).

While time geography has proven useful for analyzing vehicle movements, its main criticism is that potential path trees only delineate locations where an object possibly travelled and do not provide any measure of which routes or stops were the most probable (Pred, 1977). To overcome this limitation, there has been recent interest in GIScience toward developing probabilistic or

* Corresponding author.
E-mail address: downs@usf.edu (J.A. Downs).

0966-6923/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.jtrangeo.2012.03.017

statistical time geography constructs (Downs, 2010; Winter and Yin, 2010b). One advancement in this domain has been the formulation of time-geographic density estimation, which adapts traditional kernel density estimation (Silverman, 1986)—a popular data smoothing technique—for use with point patterns generated by moving objects (Downs, 2010). The technique which uses a distance-weighting function fit to ellipses to generate a two-dimensional probability density surface of an object's movements. The method has been successfully used to analyze movement patterns of objects such as animals that travel in Euclidean space (Downs et al., 2011). However, this technique is not yet compatible with vehicle tracking data or the broader case of discrete network space, as the elliptical distance-weighting function is not valid for networks and a suitable alternative formulation has yet to be developed. Accordingly, this paper presents a network-based time-geographic density estimator and demonstrates how it can be used to analyze vehicle tracking data. Specifically, the technique is used to construct probabilistic potential path trees that can be used to infer the likelihood that a vehicle travelled a particular route and/or stopped at certain locations.

The remainder of the paper is organized as follows. Section 2 reviews the general framework of classical and network-based time geography and also provides some general background context on time-geographic density estimation in two-dimensional space. Section 3 formulates a network-based time-geographic density estimator and introduces the concept of the 'dimension' of a potential path tree which is used in the computation. Section 4 describes an application of the new technique to analyze vehicle tracking data collected by GPS that is filtered to protect the driver's privacy. Finally, Section 5 summarizes this effort and discusses the applicability of network-based time-geographic density estimation to relevant problems in transportation geography and related fields.

2. Background

2.1. Classical time geography

Time geography (Hägerstrand, 1970) is the quintessential approach in GIScience for analyzing the movements and interactions of humans and other mobile agents. Its classical mathematical formulations (Miller, 2005a) rely on observed tracking data to construct the three fundamental time-geographic elements: the space–time path, the space–time prism, and the geo-ellipse (Fig. 1). The space–time path, colored in red, represents the approximate trajectory of the object through space and time. This path through time and space forms the basis of the space–time prism, a three-dimensional delineation of all possible spatial locations for an object during each time instance. The space–time prism is derived from the object's maximum possible velocity and constraints imposed by the spatial locations and time stamps of the control points. When the extent of the space–time prism for any two consecutive control points is mapped in two-dimensional space, it forms what is termed a geo-ellipse, or potential path area, which shows all spatial locations where the object could have been located (Miller, 2005a). Classical time-geographic elements have been utilized in a wide variety of applications in GIScience and related fields. In their simplest form, time-geographic techniques are useful for geovisualizing spatial and temporal patterns in tracking data for any type of mobile object (Gennady et al., 2007; Kwan, 2000). More prominently, however, time geography is used to understand human behavior by analyzing activity patterns (Shaw et al., 2008; Yu, 2007; Yu and Shaw, 2008; Chen et al., 2011) or quantifying potential interactions among numerous individuals (Miller, 2005b; Shaw and Yu, 2009; Yin et al., 2011). Space–time prisms, in particular, are often used to measure people's accessibility to particular opportunities, activities, or facilities (Kwan, 1998; Kwan, 2003; Miller, 2005b). Methods of time geography are also used in such diverse applications as modelling disease transmission (Bian and Liebner, 2007), analyzing crime patterns (Ratcliffe, 2006), or estimating animal activity patterns (Kritzler et al., 2007).

2.2. Network time geography

Many mathematical and statistical techniques have been adapted to a network-based framework. Prominent examples include k-functions (Yamada and Thill, 2004), kernel density estimation (Borruso, 2005; Downs and Horner, 2007; Okabe et al., 2009; Xia and Yan, 2008), cluster analysis (Yamada and Thill, 2007), and other geostatistics (Wintermute et al., 2006). Likewise, the fundamental elements of time geography have been reformulated for use in network spaces (Miller, 1991). Here, the space–time path is calculated as the shortest path along the network that connects consecutive control points. Shortest paths can be determined using either measured distances or travel times. The network space–time prism can then be constructed in one of two ways. The typical way is to assume the object is restricted to the network, and the prism is composed of two-dimensional faces which are affixed to segments of the network, as illustrated in Fig. 2. Alternatively, Neutens et al. (2008b) provide a means for generating three-dimensional space–time prisms on networks, although there the object of interest is permitted to travel off-network to participate

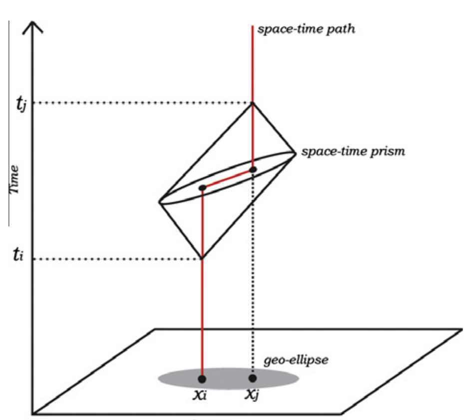

Fig. 1. Classical time-geographic elements for two control points, *i* and *j* (after Miller, 2005a).

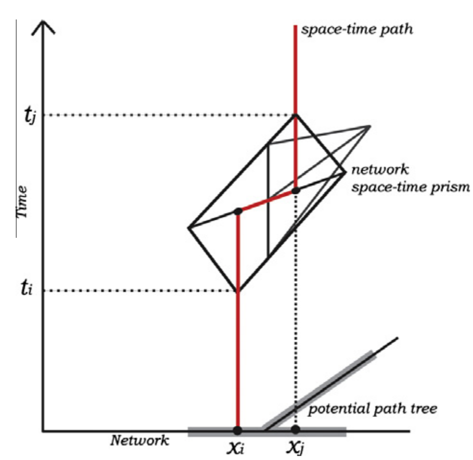

Fig. 2. Network-based time-geographic elements for two control points, *i* and *j*, located on a network.