Transportation Research Part C 59 (2015) 375–390

Contents lists available at ScienceDirect

Transportation Research Part C

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

21st International Symposium on Transportation and Traffic Theory

Spatial and temporal characterization of travel patterns in a traffic network using vehicle trajectories

Jiwon Kim [a,*], Hani S. Mahmassani [b]

[a] The University of Queensland, Brisbane St Lucia, QLD 4072, Australia
[b] Transportation Center Northwestern University, 600 Foster Street, Evanston, IL 60208, USA

ARTICLE INFO

Article history:
Received 12 April 2015
Received in revised form 14 July 2015
Accepted 14 July 2015
Available online 26 July 2015

Keywords:
Trajectory clustering
Spatial and temporal traffic patterns
Traffic stream clusters
Trajectory similarity
Longest Common Subsequence (LCS)
DBSCAN

ABSTRACT

This paper presents a trajectory clustering method to discover spatial and temporal travel patterns in a traffic network. The study focuses on identifying spatially distinct traffic flow groups using trajectory clustering and investigating temporal traffic patterns of each spatial group. The main contribution of this paper is the development of a systematic framework for clustering and classifying vehicle trajectory data, which does not require a pre-processing step known as *map-matching* and directly applies to trajectory data without requiring the information on the underlying road network. The framework consists of four steps: similarity measurement, trajectory clustering, generation of cluster representative subsequences, and trajectory classification. First, we propose the use of the Longest Common Subsequence (LCS) between two vehicle trajectories as their similarity measure, assuming that the extent to which vehicles' routes overlap indicates the level of closeness and relatedness as well as potential interactions between these vehicles. We then extend a density-based clustering algorithm, DBSCAN, to incorporate the LCS-based distance in our trajectory clustering problem. The output of the proposed clustering approach is a few spatially distinct traffic stream clusters, which together provide an informative and succinct representation of major network traffic streams. Next, we introduce the notion of Cluster Representative Subsequence (CRS), which reflects dense road segments shared by trajectories belonging to a given traffic stream cluster, and present the procedure of generating a set of CRSs by merging the pairwise LCSs via hierarchical agglomerative clustering. The CRSs are then used in the trajectory classification step to measure the similarity between a new trajectory and a cluster. The proposed framework is demonstrated using actual vehicle trajectory data collected from New York City, USA. A simple experiment was performed to illustrate the use of the proposed spatial traffic stream clustering in application areas such as network-level traffic flow pattern analysis and travel time reliability analysis.

1. Introduction

Vehicle trajectories are traffic data obtained from tracking individual vehicles' movements over time, where each trajectory is a sequence of consecutive geo-referenced coordinates and the corresponding timestamps, typically recorded every few seconds. Compared to traditional traffic data obtained from fixed loop detectors, vehicle trajectories provide much richer

* Corresponding author. Tel.: +61 7 3346 3008; fax: +61 7 3365 4599.
E-mail address: jiwon.kim@uq.edu.au (J. Kim).

http://dx.doi.org/10.1016/j.trc.2015.07.010

information—not only information on user-centric travel experiences (e.g., speed profile of each vehicle along its journey and travel times experienced by individual vehicles) but also system-wide spatial and temporal trip and traffic patterns (e.g., origin–destination pairs, routes, and bottlenecks). These are considered as by far the most comprehensive traffic data available in transportation systems analysis. With the ubiquitous use of GPS devices (e.g., smartphones and in-vehicle navigation systems) and RFID tags (e.g., electronic fare cards and electronic toll collection systems), trajectory data of moving objects in traffic networks are becoming increasingly available and this opens up new opportunities for performing more sophisticated and comprehensive analyses in several application areas, including individual mobility pattern analysis (Calabrese et al., 2013; Chen et al., 2014), real-time traffic monitoring (Feng et al., 2014), travel time reliability analysis considering both vehicle-to-vehicle and day-to-day travel time variability (Kim and Mahmassani, 2015), data fusion combining probe and fixed sensor data (Mehran et al., 2012).

Great challenges remain, however, due to the massive size of such data and the complexity of dealing with both the spatial and temporal dimensions. Data mining has thus become an important part of transportation research and practice, calling for the in-depth investigation and development of domain-specific analysis methods and tools. This paper addresses these needs by presenting a systematic framework and detailed methods for processing, clustering, and classifying vehicle trajectories, with a primary application in characterizing spatial and temporal travel patterns that can be used in network-level traffic flow analysis and travel time reliability analysis. In particular, this study considers the problem of clustering trajectories as a means to discover major traffic flow groups across the network, which may then be used in analyzing the temporal evolution of traffic states for each region more effectively. As the main contribution of this paper, we present the step-by-step procedure for clustering vehicle trajectories based on their spatial characteristics (i.e., route similarity) and show how network space can be partitioned into and represented by a few major traffic stream clusters. The proposed trajectory clustering method is demonstrated using actual vehicle trajectory data collected from New York City, USA.

1.1. Related work

Trajectory clustering has been gaining increasing interest in recent years and several clustering methods tailored specifically for trajectories have been proposed. Clustering analysis in general requires two essential components, namely, similarity measure and clustering algorithm. Trajectory clustering approaches proposed in the literature also largely depend on the choice of the combination of these two. Nanni and Pedreschi (2006) describes an approach to clustering trajectories using the average distance between objects' positions and density-based clustering algorithm OPTICS (Ankerst et al., 1999). They further extend their OPTICS-based trajectory clustering approach to identify a specific time interval where trajectories are clustered in the most meaningful way and focus on those interesting time intervals to improve the quality of clustering results. Rinzivillo et al. (2008) proposed progressive clustering that allows successive application of several different distance functions and gradual refinement of clustering results obtained from the previous steps. They suggest a number of intuitive distance functions such as 'common source', 'common destination', 'route similarity', and 'time steps'. Lee et al. (2007) proposed a partition-and-group framework for clustering trajectories (TRACLUS), which partitions a trajectory into a set of line segments and discovers common sub-trajectories by grouping similar line segments. Abraham and Sojan Lal (2012) focus on the trajectory similarity problem, where they present a spatio-temporal trajectory similarity search process to identify similar trajectories given Points of Interest (POI) and Times of Interest (TOI).

Earlier studies in trajectory clustering mainly considered trajectories of general moving objects, which can move freely in Euclidean space. Recognizing that in the case of vehicles the movements are actually constrained by the underlying road networks, another line of studies have emerged focusing on the clustering of network-constrained trajectory data. Kharrat et al. (2008) proposed a two-step clustering algorithm (NETSCAN), where the first step clusters road segments into a set of dense paths and the second step groups trajectories according to their similarity to each dense path. Roh and Hwang (2010) proposed a distance measure that reflects road-network proximity, computed using the shortest path calculation, and applied in their trajectory clustering algorithm (NNCluster). Han et al. (2012) proposed a road network aware approach for clustering trajectories (NEAT) and Mahrsi and Rossi (2013) proposed a graph-based approach to clustering network-constrained trajectory data.

While it is attractive to consider the network constraint and take advantage of the underlying network representation, it requires a pre-processing step for mapping each point in a trajectory to a road network, known as *map-matching*, and that process itself requires non-trivial effort. Above all, it is highly desirable to work purely on the available vehicle trajectory data, without having to prepare the associated network topology. In this paper, we focus on the clustering approach that does not require the information on the underlying road network. Although we do not attempt to map trajectories to known road segments, we take into account the fact that vehicle movements are constrained by the road network in defining a similarity measure. We consider the fact that trajectories on the same route are completely overlapping (at a macroscopic scale) due to the network constraint and this allows us to use the algorithm for finding the Longest Common Subsequence (LCS) between two sequences. We measure how much two trajectories overlap each other to define their level of closeness, relatedness, and connectivity. Conceptually, this LCS-based similarity measure lies between the Euclidean distance that is used for free moving objects and the graph-based distance (e.g., shortest path) that is used for fully network-mapped objects. Below, we describe a number of characteristics for vehicle trajectory data that are assumed in this study.