Available online at www.sciencedirect.com

ScienceDirect

Transportation Research Procedia 10 (2015) 276 – 285

Transportation Research Procedia

www.elsevier.com/locate/procedia

18th Euro Working Group on Transportation, EWGT 2015, 14-16 July 2015, Delft, The Netherlands

Analyzing traffic patterns on street segments based on GPS data using R

Emilian Necula*

Faculty of Computer Science, University Al. I. Cuza, General Berthelot, 16, Iasi 700483, Romania

Abstract

Nowadays GPS enabled devices are widely spread between drivers making the collection of GPS data more accessible. So there is an opportunity to infer useful patterns and trends. In this research, we plan to apply a statistical approach on 10000 vehicle GPS traces, from around 3600 drivers which are mined to extract the outlier traffic pattern to be used further in an Intelligent Transportation System. We choose to divide the urban area into a grid and organizing the road infrastructure as segments in a graph. Further, at a given time we can make an assumption regarding the congestion level in a specific area taking into account the visits for each vehicle, using the GPS trace data. Over time, the visited segments will settle into a pattern and vary periodically. In this study we will use R software in conjunction with a set of libraries. They provide an environment in which we can perform statistical analysis and produce graphics to annotate different results. Our objective is to identify contiguous set of road segments and time intervals which have the largest statistically significant relevance in forming traffic patterns. Taking into account the number of drivers that submitted their routes in correlation with the entire population on New Haven we can state that a 2-3% penetration rate of smart phones is enough to provide accurate measurements of the traffic flow and identification of traffic patterns.

Peer-review under responsibility of Delft University of Technology

Keywords: GPS patterns; traffic flow; car mobility; R-language; congestion

* Presenting author
E-mail address: emilian.necula@infoiasi.ro

Peer-review under responsibility of Delft University of Technology
doi:10.1016/j.trpro.2015.09.077

1. Introduction

Nowadays the benefits and applications of an accurate measurement and analysis of traffic patterns are broad. It can be used directly in various driver information systems, but it can also be part of larger traffic related systems, such as traffic management systems or personal car navigation. The input for the system that tries to identify movement patterns can also be various. It can consist of a more traditional data originating from stationary in-road sensors (like loop detectors) or it may be the moving car data, generated by the vehicles equipped with a GPS enabled device. The latter type of data is especially interesting for several reasons; first it quickly becomes more widespread and available (even in real-time) with the rising popularity of on-line personal car navigation or vehicles monitoring services. The moving car data also has the potential to provide exhaustive coverage of the whole road network and as a result it can yield better traffic pattern estimation (or generally traffic forecasting) model. Conversely, this kind of data can be very irregularly distributed in time and space (or even missing most of the time in certain areas) so building traffic models can be difficult. More precisely, our goal is to determine groups of street segments which possess a similar traffic load over time. In a second step we interpret these patterns considering their temporal as well as spatial characteristics. We are using available GPS traffic traces and apply statistical data-driven computational methods. An advantage of our solution is the flexibility in incorporating additional explanatory variables. Necula (2014) also used this method in the implementation of a mesoscopic traffic simulator, which is more adequate then the traditional speed–density simulators. While these general methods and tools are pre-existing, their application into the specific problem and their integration into the proposed framework for traffic flow estimation is new. The methodology is applied on a data set from New Haven County, Connecticut, USA. Within the overall analyzing process, data mining is viewed as the sub-process concerned with the discovery of hidden information. We apply a clustering algorithm to identify groups of street segments which possess a similar traffic distribution over the week and weekday to extract knowledge that will facilitate traffic flow analysis. Subsequently, we interpret the resulting clusters according to their temporal pattern as well as their geographic location. To aid in the exploration of traffic data mining we used R software to represent various states of the traffic flow.

The paper is organized as follows. Section 2 discusses related work, while Section 3 starts describing the GPS data sampling and acquisition. Moreover we introduce the analysis process containing data processing step, clustering, as well as the temporal and spatial interpretation of the traffic clusters to make a relevant decision upon congestions occurrence. We discuss our results in Section 4 and conclude the paper with a summary and outlook on future work.

2. Related work

The analysis of GPS trace data is a very active research area and has developed a number of algorithms for the clustering of trajectories presented by Lee et al. (2007), Piciarelli et al. (2008) and Rinzivillo et al. (2008) as well as various definitions of distance functions between traces introduced by Pelekis et al. (2011). Our approach differs from the previous works as we do not concentrate on traces as principal objects of interest. Instead, we also evaluate the frequency and temporal distribution of people passing a specified set of locations of an urban area, in our case street segments. Often such information is collected by traffic management centers directly on the level of street segments. But we do not have access to such databases. Fitschen and Nordmann (2010), using traces collected by the German traffic authority, analyzes and cluster time variation curves using a daily, weekly or yearly granulation of time. A very detailed analysis of traffic distributions has been conducted in Weijermars (2007). In addition to the analysis of weekly patterns also seasonal variations and variations in weather factors were analyzed. However, the basis for this analysis was data from induction loops (continuous observations of vehicular traffic for a selection of major road segments). Extensive analysis of GPS data using methods from visual analytics have been conducted in Andrienko (2008). The author formed spatial aggregation units based on a grid and provided visualization techniques for the analysis of temporal variation in speed and direction. The variations were visualized using geographically aligned mosaic diagrams, which indicate for each day of the week and hour of the day the intensity of the analyzed parameters. However, the exploration of grid cells with similar usage patterns was left to the visual capabilities of the user. In contrast, our data set contains usage patterns for several thousand street segments, which cannot be analyzed by visual inspection only.