A novel tool to understand travel behavior using Floating Car Data

Abdallah Abu-Aisha, Ralph Harfouche[1], Christos Katrakazas[1*], Constantinos Antoniou[1]

[1]Chair of Transportation Systems Engineering, Department of Civil, Geo and Environmental Engineering, Technical University of Munich, Arcisstrasse 21, Munich 80333, Germany

[*] Corresponding author

Extended Abstract

Understanding mobility patterns and human travel behavior plays a key role in urban planning, traffic forecasting, public transport management and location-based mobile applications, among others. This requires reliable and high-quality traffic information. With the advent of ubiquitous sensing technology, massive amounts of rich mobility data (e.g., human daily activities and vehicle trajectories) have become increasingly available from various sources such as cellular network data, geo-tagged social media data, and GPS floating car data (FCD). The growing availability of such data opens up new opportunities for performing more sophisticated and comprehensive spatial and temporal analyses for planning and management of transportation systems. These new data sources have the ability to improve data quality and accuracy and the potential to complement data collected using conventional methods.

This research examines the use of Big Data to infer mobility patterns and investigate users' travel behavior. More specifically, in this research, a novel tool is introduced to better understand the travel behavior and route choice decisions. One of the most popular Big Data techniques used in analyzing mobility data is spatial clustering, which identifies distinct groups of trips or trajectories based on their geographical characteristics, such that there is greater similarity within each group than between groups.

For decades, in order to provide long-term guidance and short-term strategies for transportation planning and urban development, several studies have been conducted to identify, understand, and predict human mobility patterns and travel behavior. For instance, some studies tried to investigate drivers' route choice and the factors influencing their decisions. However, there were no studies found in the literature that investigated the area-based deviation between drivers' actual chosen route and the least-cost route between two zones. Therefore, this research tries to find a way to fill this gap by presenting a novel tool that computes the deviation area between any two trajectories sharing the same origin and destination zone. This tool can help leveraging trajectory data collected by passive collection methods, such as FCD and cell phone GPS data to detect mobility patterns, and understand drivers' travel behavior and route choice decision.

In this research, a six-day dataset of floating car data from Munich city is clustered to extract meaningful urban mobility patterns. The trajectory dataset contains information like waypoints' GPS coordinates and speed for approximately 100,000 trips. K-means, hierarchical, and DBSCAN are the clustering methods that have been tested on a 10% sample of the dataset's trips. In addition, a theoretical comparison among them is conducted. The comparison and test results showed that hierarchical clustering method performs the best out of the three methods in this context. Therefore, hierarchical clustering analysis, based on Euclidean distance and Ward's linkage method, is used to spatially cluster the trips in each day according to the coordinates of their origin and destination points, such that trips that have similar origin and destination points are put together in one cluster. The statistical methods failed to provide meaningful number of clusters in the context of this research, where it is desired to have as many OD pairs as possible (while satisfying the minimum cluster size constraints) in the study area to analyze the travel behavior and mobility patterns within these pairs. Therefore, special criteria are developed to determine the optimal number of clusters for each day, taking into consideration clusters' dimensions, number of trips per cluster, and the percentage of clusters with no overlap between their origin and destination zones. The output of this clustering process is clusters of trips, where each cluster represents an OD pair.

To explore mobility patterns and drivers' travel behavior within the resulting clusters, a new tool is presented in this research. Relative Deviation Area (RDA) is a tool that aims to find the deviation between two trajectories that share the same origin and destination zone. Ideally, one of the trajectories presents the least-cost route (referential path) between the two zones. In this research, the trajectory that has the highest average speed (V) while traveling from one zone to another is considered as the least-cost route between the two zones. RDA computes the relative area by which a vehicle traveling from one zone to another is deviating from the least-cost route, while relative area means that the resulting deviation area is divided by the length of the vehicle's trajectory section that extends between the nearest pair of OD points of the two trajectories. In this research, the trips that have the highest average speed in each cluster are defined as the referential paths for that cluster. Next, RDA is computed for each trip in each cluster with all referential paths in that cluster, such that the minimum RDA value is considered as the final RDA value for that trip.

Subsequently, all resulting RDA and V values from all clusters are aggregated together in each day, as it was not possible to separately investigate each cluster or each zone in the context of this research. The relationships between V and RDA and between ΔV and RDA are investigated for each day attempting to understand drivers' travel behavior, where ΔV is the average speed difference between a given path and the referential path that gives the minimum RDA value. Before applying nonparametric regression to understand these relationships, outliers are detected and excluded using bagplot method. Afterward, Kernel regression method is applied to investigate the aforementioned relationships for each day and for all days together as well. It was found that the resulting regression curves in both relationships are almost consistent throughout all weekdays and weekend day as well. In the relation between V and RDA, it was found that V, generally, decreases as RDA increases. In other words, it was noticed that the higher the deviation area between the route that a vehicle takes and the respective fastest route between two zones, the lower the trip average speed the vehicle can achieve. As for the other relation between ΔV and RDA, it was found that ΔV increases as RDA increases. This means that the higher the deviation area between a trip's path and the respective fastest path, the higher the average speed difference between them, and thus the lower the average speed of that trip. The resulting regression curves in both cases are found to be sensible and consistent; therefore, this might indicate a potential association between deviation area and trip average speed. However, these relationships should be deeper investigated and validated by applying the presented methodology, for example, on different datasets in different locations.

In addition, the relationship between V and RDA is temporally investigated at peak and off-peak periods of one weekday and one weekend day. The regression curves suggest no big difference between weekday and weekend day, where RDA values are only a bit higher on weekday than on weekend day. With regard to peak and off-peak periods, V

values at peak periods are found to be lower than those at off-peak periods for the same value of RDA on both days. Another case is tested where only private cars are considered on one weekday, excluding all other vehicle types like taxicabs and trucks. The results showed that RDA values in private cars case are much higher and decrease more steeply than for those in all vehicle types case.

The output of this research illustrates the potential of using Big Data to infer mobility patterns and travel behavior. The developed RDA tool is expected to have several applications in different fields such as urban and transportation planning, transportation demand management, and traffic monitoring.

Keywords: Big Data, Spatial Clustering, Mobility Patterns, Relative Deviation Area (RDA), Trajectory Analysis.