Analyzing Transportation Big Data with GIS: Detecting Over-speeding Vehicles from Traffic GPS Data

Authors: Yulu Chen, Doctoral Student, ychen2793@sdsu.edu, Department of Geography.
Ming-Hsiang Tsou, Professor of Geography and Director of the Center for Human Dynamics in the Mobile Age, mtsou@sdsu.edu
Atsushi Nara, Assistant Professor of Geography and Associate Director of the Center for Human Dynamics in the Mobile Age, anara@sdsu.edu
Department of Geography, San Diego State University

With the popularity of automobiles and increasing volume of vehicles, the traffic accident rates in U.S. cities are increasing year by year. The aggressive driving behaviors can be considered as one of the main reasons to lead to serious traffic crashes according to AAA (American Automobile Association) Foundation for Traffic Safety. Over-speeding is one major problem in the category of aggressive driving behaviors. The Center for Human Dynamics in the Mobile Age (HDMA) at San Diego State University is cooperating with Virginia Tech Transportation Institute to study "*Big Data Visualization and Spatiotemporal Modeling of Aggressive Driving"* under the Safety Through Disruption (SAFE-D) Project (SAFE-D, 2018) funded by US Department of Transportation. This project aims to identify aggressive driving behaviors and visualize them by using various big data analytics and spatial analysis technologies. We utilized the experimental data for Safety Pilot Model Deployment (SPMD) program, which was a real-world data collection program with vehicle-to-vehicle (V2V) and vehicle-to-infrastructure (V2I) communication devices in October 2012 and April 2013 in Washtenaw County, Michigan (Gold, 2012).

Big Data are commonly defined as the size or the complexity of data is too big to be processed effectively by traditional software. For example, the one-month GPS tracks of SPMD data contains more than 1.5 billion GPS points (205 GB in a comma-separate value file format), which cannot be handled by traditional GIS software easily. The purpose of this study is to utilize various SQL and No-SQL databases to process big data and to detect over-speeding vehicles or aggressive driving behaviors using environmental GIS layers with GPS trajectory data. It will help transportation researchers to develop a real-time visualization application in the future. We wrote Python programs to parse the original large data into smaller datasets. The smaller datasets were converted into both PostgreSQL and MongoDB databases and link them with ArcGIS Pro. This study used two types of data, vehicle points and roads. The points dataset is a series of GPS points extracted from the randomly selected vehicles from the SPMD database in April 2013. Those points contain basic safety messages (BSM) for a vehicle such as speed, location, direction, yaw rate, and heading collected at the rate of approximately 10 Hz. A road line dataset contains major road condition and speed limit, which comes from the Southeast Michigan Council of Governments (SEMCOG) Annual Average Daily Traffic (AADT) program.

To find the vehicle points with over-speeding cases, we used sampled SPMD data (0.1%) with ESRI ArcGIS Pro and buffered the roads with 5 meters wide since the major roads should have at least 5 meters wide. Then we used a spatial join to aggregate vehicle points data into the road buffers. We compared speed recorded at each vehicle point with the road speed limit and extracted four categories of over-speeding clusters: class 1: Over speed 1-5 MPH; class 2: Over speed 5-10 MPH; class 3: Over speed 10-20 MPH; class 4: Over speed more than 20 MPH (Figure 1). We visualized these data on GIS maps to identify their spatial patterns and point density. Figure 2 illustrated the severity of overs-speeding vehicles and their cluster patterns. The visualization result showed that points with more than 10 MPH over-speeding mostly occurred in the intersections of major highway segments and ramps. In addition, it also appeared when the speed limit changes between two segments of local roads. Finally, most over-speeding less than 10 MPH occurred in the Ann Arbor downtown. The spatial pattern of over-speed

activity can be further evaluated in the future using advanced machine learning methods.

The automatic detection of aggressive driving behaviors can facilitate the investigation of behavioral and environmental risks in space and time from a large collection of sensor-based datasets. Identifying and addressing those risks is one of the central factors in building a safe transportation system, especially with the potential of autonomous vehicles and self-driving cars in the near future. This pilot study helps us to understand the potential and challenges of using GIS to analyze transportation Big Data. We continue to strive for finding better approaches to challenges in big data analytics, such as data partitioning, data sampling from the original datasets, big data visualization, and developing parallel computation algorithms.

Reference

- SAFE-D (2018). Big Data Visualization and Spatiotemporal Modeling of Aggressive Driving: URL: https://www.vtti.vt.edu/utc/safe-d/index.php/projects/big-data-visualization-and-spatiotemporal-modeling-of-aggressive-driving/ (Last accessed on December 03, 2018).
- Ariel Gold (2012). SAFETY PILOT MODEL DEPLOYMENT DATA, sponsored by the U.S. Department of Transportation (USDOT). URL: https://www.its.dot.gov/factsheets/pdf/SafetyPilotModelDeployment.pdf (Last accessed on December 03, 2018).

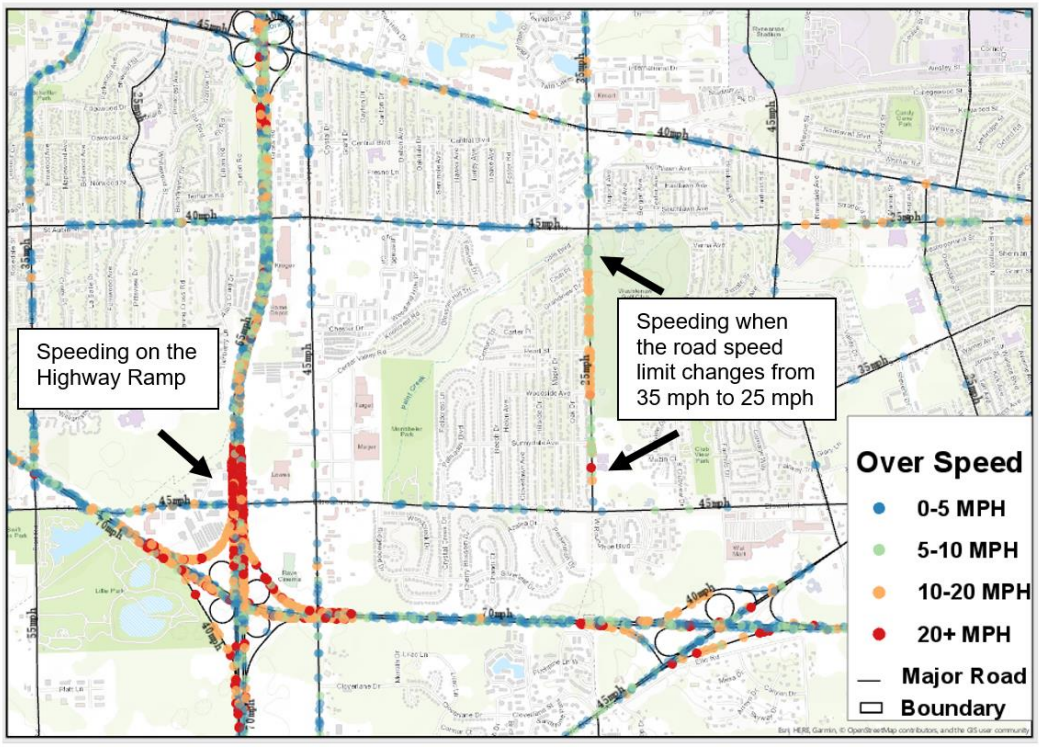

Figure 1. Identifying over-speeding locations in major roads using GPS trajectory datasets.