MPC-543
November 15, 2017

Project Title:
Big Transportation Data Analytics

University:
University of Utah

Principal Investigators:
Xiaoyue Cathy Liu
Assistant Professor
Civil & Environmental Engineering
Phone: (801) 587-8858
Email: cathy.liu@utah.edu
ORCID: 0000-0002-5162-891X

Research Needs:
The world of Big Data is moving fast. There are several private sector Big Data vendors using data from vehicle probes, and from Wi-Fi, cellular, and Bluetooth data traces to synthesize transportation information such as vehicle speeds and origin-destination travel flows. In addition, Utah Department of Transportation (UDOT)'s Traffic Operations Center recently just purchased HERE traffic probe data to assist traffic management and road network performance assessment. UDOT also possesses Big Data in the form of Performance Measurement System (PeMS), that has been archived since 2008. This historic data set should be mined to learn new things about the direct and proximate causes of traffic volume change and travel reliability. There is also an urging need to develop analytical approaches that leverage existing historic data to reveal methods for estimating traffic, with the potential to reduce the burden of the annual traffic count program.

UDOT maintains an annual traffic count program which requires the acquisition of hundreds of short-duration counts each year. This traffic count effort represents a significant cost to UDOT, while also exposing UDOT staff to the dangers inherent to being exposed to traffic. The proposed research will seek to determine whether statistical modeling and/or machine learning methods might be employed to partially or fully supplant the short-duration traffic count program and, in doing so, reduce the effort, cost, and staff exposure of UDOT's traffic count program.

Research Objectives:
1. Determine whether statistical modeling and/or machine learning can be applied to for estimating/predicting traffic conditions
2. Develop analytical methods that can be used by UDOT in the future to estimate traffic volumes and reliability for short-duration traffic count sites

The primary objective of this research is to determine whether statistical modeling and/or machine learning can be applied to for estimating/predicting traffic conditions (e.g. VMT and reliability) when it is conflated with other available data sets, such as demographics, income, highway capacity, and highway improvement projects.

The secondary objective is to develop analytical methods that can be used by UDOT in the future to estimate traffic volumes and reliability for short-duration traffic count sites, based on other available data, and quantitative relationships. A range of analytical methods will be attempted, from well-established methods (e.g. traffic engineering techniques) to more sophisticated methods, such as advanced statistical modeling and machine learning.

Research Methods:
With freeway system carrying the major portion of traffic, UDOT has set up a [freeway performance metrics monitoring system](#) to facilitate improved decision making. The current version includes aggregated reporting and retrieval of travel time, speed, mobility cake, and reliability measures along the major freeway corridors (I-15, I-215, I-80, etc.) The current analyses distinguish the freeway segments into reliably fast, reliably slow, unreliably fast, and unreliably slow based on a robust sensitivity analysis. Such categorization is able to identify the sources that are attributable to the corridor performance levels. These analyses are supported by PeMS, which is a data repository integrating various traffic data sources (e.g. loop detector, radar sensors, incident logs) used by UDOT. UDOT also maintains separate databases for weather and roadway conditions, incidents and special events, and lane closure. Intersections and freeway interchanges are monitored using UDOT signal performance metrics. UDOT provides these metrics that show real-time and a history of performance at signalized intersections. These metrics include approach volumes, approach speeds, split monitor, approach delay, arrivals on red, and travel time. Most notably, UDOT is in the process of purchasing probe data to track the trajectory-level speed and travel time on major arterials. The goal is to offer a network-wide (both freeway and arterials) real-time performance monitoring system to transform traffic management.

This project will begin by collecting and integrating these heterogeneous data sources to enhance scientific traffic prediction, discovery and decision-making. It also offers methods for analyzing the causes and locations of unreliable performance that help identify possible mitigation strategies.

We will employ machine learning techniques to estimate and predict roadway VMT and reliability, potential analytics include decision trees, support vector machines, and deep learning. These methods will be programmed in the R statistical programming language. In other cases it is anticipated that a custom program or a routine from a program library will be utilized to implement these machine learning algorithms.

The research will reveal the value of analytics in predicting traffic volume and reliability changes based on the historic traffic count record, as dependent on several other variables (incident, work zone, adverse weather, etc.) Based on statistical measures of fit, a best analytical approach, or set of approaches, can be incorporated into a DOT's traffic operation program to predict future traffic volumes and reliability for state and federal-aid routes, and for new planned highway

segments. This information could be used to automatically feed the annual Highway Performance Monitoring System (HPMS) reporting UDOT is required to prepare, as well as to provide planning-level traffic forecasts, as an alternative to travel modeling approaches.

Expected Outcomes:
1. Determination of the most effective analytical methods (statistical modeling, machine learning) for predicting traffic volumes and travel reliability on major freeway and arterials within the Salt Lake City metropolitan area
2. Development of methods for incorporating analytics into UDOT Asset Management and Planning functions (e.g. for HPMS reporting, for estimation of future traffic volumes and reliability)
3. Findings that will inform UDOT into methods for incorporating Big Data analytics into their business enterprises, including supporting the development of a Business Plan/Roadmap for evolving the department to leverage Big Data in the future

Relevance to Strategic Goals:
- Economic Competitiveness
- Livable Communities

The expected outcomes will support explicit considerations of mobility and accessibility of the existing roadways to provide users efficient travel experience. The developed machine learning algorithm in this project would inspire a series of research, including incident-induced delay prediction, real-time traveler information system, integrated corridor management, etc. These will provide roadway users, either en route or planning their trip, rich information to make sound travel decisions (e.g. alter departure time, seek for alternative routes and travel modes, etc.) From the agency's perspective, the research outcomes might lead to investment decisions of transportation asset management, such as park-and-ride facilities, dynamic message signs, and transit centers, to increase transportation choices and pursue a well-coordinated transportation system.

Educational Benefits:
One graduate student will be heavily involved in this research. He/she will lead the preparation of journal publications resulting from the work, and in most cases, deliver conference presentations. The project will serve as a basis for his/her dissertation work. The undergraduate course CVEEN 3520 Introduction to Transportation Engineering and graduate course CVEEN 7545 Transportation Network Modeling, are the ideal platform to introduce the concept of machine learning techniques for traffic prediction. The modeling procedure developed in this research will lead to new material included in the course to teach the students practical and interdisciplinary skills on computational techniques for transportation analysis.

Tech Transfer:
The potential audiences for this research are individuals involved in the traffic operations and transportation asset management, including traffic engineers, planners, and senior leaders at FHWA and at individual state DOTs. The following agencies, offices, and committees are those most likely to take a leadership role in implementing the research results:
- Utah Department of Transportation

- FHWA Office of Operations
- TRB Managed Lane Committee
- TRB Highway Capacity and Quality of Service Committee
- TRB Freeway Operations Committee

The proposed principal investigator routinely interacts with UDOT, FHWA, SHRP 2 Reliability Program, and the listed TRB Committees. The 2018 Midyear Meetings of TRB Highway Capacity and Quality of Service Committee, TRB Managed Lane Committee, and TRB Freeway Operations Committee will be an opportunity to share early results and future directions of the research project. The proposed principal investigator will work with the committee chairs to possibly get a presentation on the project added to the agenda. The proposed principal investigator and her graduate students routinely attend TRB's annual meeting as well. At least one TRB paper on this work will be submitted for presentation and publication.

Work Plan:
1. Assemble Steering Committee and set up meeting/communication protocols for the duration of the research project
2. Research synthesis
3. Assemble Data Sets, Identify and Discuss Data Gaps
4. Review Analytical Methods for Application to Time Series Forecasting
5. Apply Analytics
6. Findings and Conclusions
7. Final Report

Task 1 – Assemble Steering Committee (1 month)
Set up meeting/communication protocols for the duration of the research project.

Task 2 – Research synthesis (3 months)
Identify, review, and synthesize relevant published literature regarding econometric and machine learning methods applied to traffic data.

Task 3 – Assemble Data Sets, Identify and Discuss Data Gaps (4 months)
We will assemble data sets that correspond longitudinally to the historic UDOT traffic data. Associated data sets will include land use data, Census data and related demographic data, measures representing the highway network (e.g. density, capacity), and records of UDOT highway improvements. Other data sets will be reviewed for appropriateness

Task 4 – Review Analytical Methods for Application to Time Series Forecasting (2 months)
This task will be conducted in parallel with Task 2 in order to more efficiently determine which analytical methods are best given the available time series data sets. In Task 3, we will review the types of analytical approaches that would be appropriate for predicting traffic volumes (VMT) and reliability given the supporting data assembled in Task 2.

Task 5 – Apply Analytics (4 months)
In this task we will apply three analytical methods for predicting traffic volumes and reliability. In some cases these methods will be programmed in the R statistical programming language. In other

cases it is anticipated that a custom program or a routine from a program library will be utilized to implement machine learning algorithms.

Task 6 – Findings and Conclusions (2 months)
What analytical/machine learning methods are most accurate in predicting future traffic and reliability? What surprise findings are inferred from classifying traffic data and the drivers?

Task 7 – Final Report (1 month)
Describe a future Transportation Data Analytics Framework and the top five priority actions UDOT should take now to advance that Framework. Actions will be described, and include responsible UDOT business unit, approximate costs, and timeline.

Project Cost:
Total Project Costs: $90,000
MPC Funds Requested: $40,000
Matching Funds: $50,000
Source of Matching Funds: Utah Department of Transportation