

Program Progress Performance Report

Submitted to

United States Department of Transportation (USDOT), Office of the Assistant Secretary for Research and Technology (OST-R)

Federal Grant number: DTRT13-G-UTC48

Project Title: Transportation Informatics Tier I University Transportation Center (TransInfo UTC): Harnessing the Power of Big Data in Support of USDOT Strategic Goals

Adel W. Sadek, Ph.D.

Professor and Director of TransInfo UTC

Department of Civil, Structural and Environmental Engineering

University at Buffalo, the State University of New York

Buffalo, NY 14260

E-mail: asadek@buffalo.edu

Phone: (716) 645-4367

April 30, 2018

DUNS#: 038633251

EIN#: 14-1368361

Recipient Organization: The Research Foundation for SUNY, 402 Crofts Hall, University at Buffalo, Buffalo, NY 14260

Recipient No.: 66473

Project/Grant Period: October 2013 – September 2018

Reporting Period: October 1, 2017 – March 31, 2018

Report Frequency: Semiannual

Signature:

A handwritten signature in black ink, appearing to read "Adel W. Sadek", with a long horizontal flourish extending to the right.

1. Accomplishments

1.1 Major Goals of the Program

The volume, variety, quality and resolution of transportation-related “Big Data” currently present the transportation community with an unprecedented opportunity for improving system performance. Specifically, the wealth of data can be *studied, analyzed, and mined for insights and applications that can improve the efficiency, safety, sustainability, resiliency and reliability of the transportation system, and can inform and guide transportation policy*. It is to this goal that TransInfo’s activities are dedicated. TransInfo’s mission is to undertake research, education, training, and technology transfer activities aimed at realizing the full potential of “*Big Data*” and *Transportation Informatics* in: (1) improving transportation system performance; and (2) guiding investments and policies. The activities performed under the umbrella of the TransInfo Center has helped advance the state of knowledge in the emerging field of transportation informatics, while preparing and educating both the existing transportation workforce, as well as the next generation of transportation professionals, in how to harness the power of “Big Data” to address transportation challenges.

Because improved utilization of transportation data helps improve system performance, and because transportation serves as the very foundation of our nation’s economy, the Center’s activities directly address the US Department of Transportation (US DOT) Strategic Goal of “Economic competitiveness”. However, “Economic competitiveness” is not the only goal that TransInfo UTC has been addressing. The common thread behind all of TransInfo’s research and educational activities is to compile, fuse, and mine various data streams to support a wide range of transportation applications in traffic operations, safety, emergency operations, travel behavior modeling, and performance measurement. As such, TransInfo also touches upon the goals of “Safety”, “Environmental Sustainability”, “Livable Communities”, and the “State of Good Repair”.

1.2 Progress and Accomplishments

TransInfo’s progress and accomplishments, during the reporting period covered by this report, are organized under the following three headings: (1) research activities at the different institutions of TransInfo; (2) graduate student activities; and (3) outreach and technology transfer activities. An additional section will summarize some of the recent accomplishments of TransInfo researchers which were facilitated by the support provided by TransInfo.

1.2.1 Research Activities

1.2.1.1 Research at the University at Buffalo

a) Inferring Origin-Destination Demand and Utility-Based Travel Preferences in Multi-Modal Travel Environment Using Automatic Fare Collection Data

Summary: The objective of this project is to develop scalable inference methods for understanding and expressing public transit system utilization based on fundamental travel behavior. The researchers have proposed a methodology that identifies both the preference vector and the true OD-pairs by collecting and analyzing Automatic Fare Collection (AFC) system-type data (stop-level ODs), as travelers make their multi-modal route choice decisions within the stochastic travel environment. The proposed methodology captures system-wide demand changes with respect to changes in stochastic travel environment.

Project Status: On going

Milestone accomplishments and dates:

- The research team has developed a methodology to reconstruct the travel environment, which is both dynamic and stochastic. Major environment variables include travel time, level-of-service, transfer time, and cost.
- The research team has estimated travel disutility function based on the observed route choices in multi-modal travel environment (through AFC records).

- The researchers have created synthetic experimental data to be used in estimating the OD demand matrix. The research team then developed and tested methodologies with the synthetic data set generated.
- The research team is currently working on further testing the methodologies for their scalability on real multi-scale cases. We have run test cases, and we are refining our approach and data sets.

Planned Activities:

The research team will focus on studying the scalability of the developed methodologies and applying them to real multi-scale cases.

b). Predictive Analytics for No-Shows and Cancellations for Paratransit Operations

Summary:

The objective of this research project is threefold: (i) to develop a classification methodology to predict no-shows and cancellations of trips in paratransit systems, (ii) to incorporate the classification model predictions into trip booking and routing models for paratransit operations (including novel overbooking problems in the context of paratransit), and (iii) to evaluate the value of the classification models' predictions in paratransit planning and operations.

Project Status: On going.

Milestone Accomplishments and Dates:

- Major Activities: literature review, model and solution development
- Specific Activities: we are reading literature in Dial-A-Ride Problem and its variations, Monte-Carlo scenario generation. Implementing column generation and dynamic programming algorithms.
- Significant Results: testing of the model
- Key outcomes: small toy case study results

Planned Activities:

Finalizing model and solution approach. Large-scale case study.

c). Using Social Media Data to Enhance Up-to-date Origin-Destination Demand Monitoring

Summary:

This project aims to enhance the estimation of travel demand by social media data. This project involves three steps. The first step involves the estimation of social media users' demographics (gender, race, education, age, etc.). The second step removes the sampling bias of social media from user's accurate demographics. The third step finally estimates the origin-destination demand from corrected social media samples and other data sources.

Project Status: On going

Milestone Accomplishments and Dates:

We conducted manual labeling of more than 1500 Facebook users and their associated tweet account with two undergraduate students and one graduate students. Finally, we have obtained around 1000 valid users with true label of demographics. We have applied initial classification analysis. We found that classifying gender and race is relatively easy. However, classifying age is very difficult. We are exploring more features and developing deep learning related models for further tests.

Planned Activities:

Finish existing work in demographics estimation and continue working on sampling bias correction and OD demand estimation.

d). R2Deep: Recharging Recommendation System for Electric Taxis based on Deep Learning

Summary:

We developed a recharging recommendation system for electric taxi (eTaxi) drivers to maximize their profits by giving recommendations on when/where/how long to recharge and where to find passengers. Under the constraint of state of charge and dynamics of battery consumption, the problem is formulated as

a dynamic programming problem. We developed a time-dependent graph representation of road networks by mining the historical taxi GPS traces and provided a time-dependent recursion strategy to generate optimal route.

Project Status: On going

Milestone Accomplishments and Dates:

10/1/17-11/1/17: Data preprocessed for deep learning model;

11/1/17-1/1/18: Developed super-resolution convolutional neural network to predict pick-up events in Shenzhen and the predicted result as below:

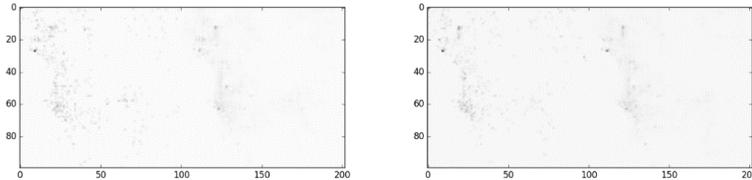


Fig.1 Predicted Result (Left) and Ground Truth (Right)

1/1/18-2/1/18: Developed a dynamic programming algorithm to maximize the average profit of an eTaxi under the constraint of state of charge and dynamics of battery consumption. The computational complexity is $O(nm)$, where n is the total number of battery level intervals, and m is the total number of time slots.

2/1/18-3/1/18: Applied map-match algorithm and defined the potential profit of road segment v at time t as $g(v,t)$, computed by mining the historical eTaxi and traditional taxi GPS traces:

$$g(v, t) = \frac{\sum_{i=1}^{N_r} \frac{Fee(i; v, t)}{t_i}}{N_r} P(v) - (1 - P(v))L(r)p_e$$

3/1/18-4/1/18: Developed a time-dependent graph representing road and designed a time-dependent recursion strategy to generate optimal driving route.

| | Net Profit | Drivers' profit |
|--------------------|------------|-----------------|
| Mean | 0.471587 | 0.351793 |
| Standard deviation | 0.017932 | 0.019766 |

Table1. Net profit per Unit Time

Planned Activities:

Prepare to submit one or two papers on relevant conferences.

e). City-wide Transportation Information Estimation with Heterogeneous Urban Data

Summary:

We propose to develop novel data mining methodologies that integrate heterogeneous urban data for the estimation of city-wide transportation information, including the inference of traffic speed, volume and emission. We propose to intelligently utilize all the available sources to achieve an accurate estimation. We also propose to detect traffic anomalies and derive a confidence measurement together with each estimate.

Project Status: On going

Milestone Accomplishments and Dates:

- Major activities: We investigated heterogeneous urban data sources, including loop detector data, taxi trajectories, points of interest, road networks and weather data to learn their effect on traffic

volume. We developed a novel method that integrates these heterogeneous sources for traffic volume estimation, and tested the approach on real urban data collected in a city in China.

- **Specific objectives:** To identify speed and volume patterns, find the influencing factors of traffic volume, and design an effective approach for traffic volume estimation.
- **Significant results:** We found the periodic patterns in speed, which correlate with traffic volume. The other influencing factors include the spatial closeness in roads, the weather, the similarity in the point of interests. Based on this observation, we constructed a graph in which each node represents a road segment, and the edge indicates the similarity in traffic volume. The travel speeds are estimated from taxi trajectories. Based on the similarity graph, we proposed a semi-supervised learning approach to infer traffic volumes of all road segments. On some road segments, we have the traffic volume information obtained from loop detectors. The proposed semi-supervised approach can infer the traffic volume information by propagating volume information via the similarity graph. Real-world data was collected from 155 loop detectors and 6,918 taxis over a period of 17 days in Guiyang China. The experiments performed on this large urban data set demonstrate the advantages of the proposed method.
- **Key outcomes:** Research results were summarized in a paper published in SIGSPATIAL'17 conference (peer-reviewed, acceptance rate 18%).

Planned Activities:

We plan to conduct the task of traffic anomaly detection, which is to identify road segments that deviate from the expected behavior significantly at a certain set of time slots. The idea is to design some test statistics to test whether the volume of a road exceeds its expected value or not. We plan to evaluate the designed statistics on the traffic data collected in Guiyang, China, and visualize the results on the map.

f). Improving the Service Quality of Bike Sharing Systems via the Analysis of Real-Time User Data

Summary:

This project develops a statistical framework to estimate the ability of bike-sharing systems to serve incoming requests. By mining user data from the system's smartphone app, an operator can utilize these models to predict the likelihood that any potential user decides to use the system under the given the system's conditions the user encounters. The models are coupled with a bike-redistribution model to analyze the effectiveness of quick redistribution tours.

Project Status: On going

Milestone Accomplishments and Dates:

The principal activities conducted so far involved the development of the statistical framework and the selection of the proper techniques based on the available data.

- We have spent some time mining and cleaning publicly-available data from several bike-sharing systems around the world. A major difficulty that we encountered is that most publicly available data is either incomplete or do not fully include information regarding demand rejections from a system. Similarly, most data sets are typically not linked to user profiles and the ones that can be found with such features, lack other component required for the proposed models. Some of these needed features thus required to be compiled statistically over time.
- We have conducted a thorough literature review and have observed that, despite data analytics models being ubiquitous in bike-sharing applications, no current project is framing the redistribution problem from a similar perspective, which is quite encouraging.
- We spent time processing data that could be used not only for this project, but for subsequent efforts on similar topics. We plan to make all the data and models that we have publicly available.
- We have a clear selection of the models to be used in the different stages of the framework.
- We are currently working on possessing the data we have found to start testing the models.

Planned Activities:

Our principal objective at the moment is to process the available data to test the proposed models, which we hope we can finish by the end of May. We also have plans for a few meetings with our industry partner

to validate the proposed models and expectations. After the data is preprocessed, we plan to conduct preliminary runs of the learning models and updated our framework based on the preliminary results.

g). Factors affecting perceived and observed aggressive driving behavior: An empirical analysis of driver fatigue, and distracted driving

Summary:

This project investigates the influential factors of observed and perceived aggressive driving behavior under driver's fatigue, as well as under normal and distracted driving conditions. Using driving simulation data and advanced statistical modeling techniques, significant driver-, environment- or trip-specific factors that have the potential to affect aggressive and distracted driving behavior, will be identified.

Project Status: On going

Milestone Accomplishments and Dates:

Following the original plan, the research team has so far completed the following tasks:

- The literature on various types of aggressive driving behavior, including driving under fatigue and distracted driving behavior, has been extensively reviewed.
- The data from the driving simulation experiments and the survey responses of the participants have been linked and compiled in one integrated dataset. The dataset includes socio-demographic and behavioral information and driving simulation raw data for 42 participants.
- Upon the compilation of the comprehensive dataset, incidents of driver's fatigue either before or after any phase of the driving simulation experiment, were identified. In addition, incidents of driving under various types of distraction were also identified; such distractions were inducted to the driving simulation experiments through the consideration of different driving scenarios.
- Statistical modeling of perceived and observed aggressive driving behavior is currently in progress. A bivariate grouped random parameters probit model has been estimated for the entire dataset. On the basis of the results, likelihood ratio tests are also conducted to identify if the estimation of separate models is warranted for the investigation of aggressive driving behavior under driver's fatigue and under distracted driving conditions.
- The research team is in process of estimating models of perceived and observed aggressive driving behavior under driver's fatigue and distracted driving conditions.

Planned Activities:

- Estimate and finalize the remaining statistical models for perceived and observed aggressive driving behavior.
- Calculate and verify (pseudo-)elasticities for the explanatory variables of the statistical models.
- Implement the estimated models for examining their forecasting accuracy.
- Prepare a final, comprehensive report summarizing the literature review, data collection and collation process, data analysis, statistical modeling, interpretation of statistical findings, and implications of findings.

h). Predicting changes in driving safety performance on an individualized level under naturalistic driving conditions

Summary:

The goal of this project is to examine how driver safety performance varies by location, time of day, hours on duty, and driver workload and to model changes in performance. \

Project Status: On going

Milestone Accomplishments and Dates:

- Major Activities:
 - Model input parameters for characterizing workload: tasks performed, cognitive load, miles driven, road location, hours on duty, time of day, driving characteristics
 - Quantify changes in driving performance based on characteristics of mirror checks

- **Specific Objectives:**
 - Two datasets were received from Maven Machines in October 2017. The data were collected from a headset worn by truck drivers. Each row of the data is an “event”, considered when a driver looks either at the left or right mirror or down. Other variables in the dataset include date and time, employee id, GPS location (latitude and longitude), event type, maximum angle of head turn, duration of event, and speed of driving at the time of event. The first step was to merge the datasets and clean any bad data from the files. The files were edited to create consistent variable naming and formatting. Bad data was identified based on the following criteria: status indicators (based on binary bit data) from events when the system was not being worn or the truck was not being driven, data points when the driving speed was > 90 mph (deemed to be infeasible for real data), events where the sensor position indicated an angle of > 180 degrees (infeasible). The total dataset consists of 253 drivers, with 12 million events (rows). The data cleaning took until January 2018.
 - Following initial data cleaning, exploratory data analysis was conducted to isolate the relevant input parameters. Common data visualization approaches were employed. Variables identified for analysis include: distribution of events that are left, right, or down (mirror checks vs. potential distraction to look at other in-cab items); number of event statuses that indicate the driver was on a phone call or their phone screen was on (a sign of distraction); duration of trips; time of day when the trip occurs; average speed while driving with different events; duration of glances away from center; miles driven; continuous miles driven between breaks; and road types (highway vs. local road). Statistical analysis was conducted to capture central tendency, variability, and distribution of the parameters. This process is ongoing, and is expected to be completed by the end of April 2018.
 - The output variable is the time between mirror checks, here defined as the time between consecutive events. Only left-left, left-right, right-left events are considered, since a down event would not indicate a mirror check. This output variable will be analyzed over time and with respect to the recommended gold standard of a mirror check every 8 seconds and a balance of left and right mirror checks. This task is ongoing, and is expected to be completed by the end of April 2018.
- **Significant Results:**

At this point, the data has been prepared for the driver safety performance analysis and prediction modeling. The time between mirror checks, and changes to this parameter with respect to the input parameters, will be used as the outcome variable.

Planned Activities:

Following completion of the tasks described, the next step will be to establish trip and route level outcomes, and to construct driver-level baselines. Any metrics which exhibit low variability across driver, trip, or route will be excluded from further analysis as inappropriate for modeling. Different fusion techniques will be investigated to combine the remaining metrics into an overall driver safety performance score. We will then investigate data-driven modeling approaches for driver safety performance prediction, including structural analysis and machine learning approaches. We will begin with clustering approaches to see whether a model can detect an increase in the likelihood of an unsafe behavior. The goal is to understand the pattern in the structure of the input variables, such as to arrive at a classification of how they influence changes in a driver’s state over time. A submission is planned for the August 1 deadline for the Transportation Research Board Annual Meeting.

I) Border Crossing Delay Prediction

Summary: This research developed several models for predicting short-term border crossing traffic, along with queuing models for predicting the anticipated delay. Based on these models, an Android smartphone application called the Toronto Buffalo Border Waiting (TBBW), was designed to collect, share and predict waiting time at the three Niagara Frontier border crossings (i.e., the Lewiston-Queenston Bridge, the Rainbow Bridge, and the Peace Bridge).

Project Status: On-going (nearing completion).

Milestone Accomplishments and Dates:

- The paper summarizing the most recent work involving developing a hybrid machine-learning model for predicting a prediction interval within which future traffic volumes are expected to lie with a certain degree of confidence, has been revised and resubmitted. It is currently undergoing its second round of reviews by the Transportation Research – Part C journal is currently being revised. The paper is entitled, “Hybrid Machine Learning Model for Interval Prediction of Short-term Traffic Volume and its Application to Optimal Staffing Level Plan Development”.

Planned Activities: We are in the process of securing additional direct delay data collected from the blue-tooth based systems deployed at the Niagara Frontier Border Crossings. With the additional data, we plan to use deep learning methods to directly predict the delay and compare this approach with our previous models. We are planning to submit a paper to our fourth Annual Symposium on Transportation Informatics to be held in August 2018, summarizing this new work, and then to prepare Final report.

j) Novel Machine Learning Methods for Accident Data Analysis

Summary: With the recent advances in data collection, storage and archival methods, the size of accident datasets has grown significantly. This in turn has motivated research on applying data mining and complex network analysis algorithms, which are specifically designed to handle datasets with large dimensions, to traffic accident analysis. This project explored the potential for using a number of machine learning and data mining methods to accident data analysis.

Project Status: Complete

Planned Activities: We are currently preparing the Final Report.

k) Buffalo-Niagara Transportation Data-warehouse Prototype and Real-time Incident Detection

This project is complete and the final report has been submitted and posted online.

l) Towards Quality-Aware Big Data Integration for Crowdsourced Road Sensing System

Summary:

This project proposes a Quality of Information (QoI) aware crowdsourced road sensing system that can monitor a wide spectrum of road events, such as congestion, accidents, potholes, bumps, and road signs. Our crowdsourced road sensing system gathers information from a variety of sources, including not only connected vehicular smartphones, but also social media as well as internet users. To integrate the information from multiple sources, we develop a novel QoI aware information integration framework that can automatically estimate the QoI of each source and incorporate the estimated QoI into the information integration process.

Project Status: Completed

Milestone Accomplishments:

In this project, the researchers presented the design, implementation, and evaluation of VehSense, a vehicular sensing platform that uses smartphones and OBD-II adaptors for slippery road detection. The system utilizes standardized OBD-II comments and Android sensor readings, which makes it compatible with most of the passenger vehicle. It does not require user calibration and can be readily integrated into navigation applications. Our pilot study at Buffalo, NY, shows that the system is an effective skidding detector during snow days.

Planned activities:

The researchers want to add more capabilities such as integrating information from heterogeneous sources and supporting environmental sensors. More rigorous and larger scale experiments are also necessary. The objective of the VehSense system are: 1) develop a low-cost crowdsourcing data collection system; 2) explore how to integrate this data with other contextual information and use it to detect slippery road, pavement roughness and other potential uses; and 3) evaluate how to inform driver about safety information without creating distractions.

1.2.1.2 Research at Rensselaer Polytechnic Institute (RPI)

Urban System Modeling and Performance Measurement Using Multiple Data Sources

Summary:

As technologies advance, emerging urban data are increasingly available for wide urban areas, which are referred to as *Urban Hybrid Traffic Data (U-HTD)* in this report. U-HTD provides great opportunities for urban transportation/traffic system performance evaluation, modeling, and management. It also poses great challenges in data collection, processing, storage, and use, including data privacy and security issues. This research investigates the issues of applying U-HTD for urban traffic modeling, including (i) fusion methods of mobile and fixed data, (ii) freight transportation modeling using multiple sources of data, and (iii) new econometric method for analyzing interaction behavior data.

Project Status: Completed and final report submitted.

Milestone Accomplishments:

For the first application, a data fusion and information integration approach was proposed to interpret traffic flow and matching of data between different locations or sources along arterial corridor. This was done by making connections between the upstream and downstream traffic data at various sensor locations. Historical traffic information such as travel time, lane choice, and traffic merging were integrated into the matching algorithms. By validating the fixed-location data against mobile sensing data, some upstream and downstream fixed-location records can be directly matched, and the matching results can therefore be significantly improved. In addition to the ideal scenario which assumed the fixed-location data and mobile sensing data are accurate and perfectly matched, some practical issues were also studied related to the use of heterogeneous traffic data, such as data errors, and detection failures.

For the second application, the goal was to study time dependent effects in freight trip attraction models. The research showed that freight trip generation models are not temporal stable but rather, that time has a significant effect on the model parameters. This implies that models implemented without taking into account the time effects, serious problems of under or overestimations in the number of freight trips would occur. The results also reveal that freight trip generation patterns are highly responsive to market disruptions such as financial and economic crises. Freight trip generation models also tend to be linear rather than exponential, with the inclusion of time-dependent variables and the revenue. Finally, another important implication of this study is that freight trip generation models are very responsive to market disruptions taking place over the time which is captured by including in the models for freight trip generation time-related variables and variables accounting for changes in the revenue of establishments.

The third application developed an innovative econometric model to analyze joint responses with the consideration of partner selection and joint decision making. The proposed model consists of two parts: the first part uses a limited dependent variable equation to explain the matching process in a one-to-many matching structure; the second uses a multinomial probit model to characterize joint responses in the form of categorical values. The two parts are connected with each other by a variance-covariance matrix, which is determined by empirical data.

1.2.1.3 Collaborative Research between Rensselaer Polytechnic Institute & George Mason University

a). Developing a Smartphone App Platform to Decipher Travel Behavior

Summary: This project is focused on the development of a smartphone app for parking management based on cloud sourcing information. Specifically, the app will be designed to collect information about parking usage.

Project Status: On-going but nearing completion

Milestone Accomplishments:

- The research team reviewed existing literature on using smartphone apps for transportation data collection.
- The research team developed an app entitled “GMU Parking Helper” in both iOS and Droid environment. The prototype app has been published in both iTunes Store and Google Market.

- The RPI team has completed the traffic state estimation from reduced GPS data by specially designed sampling techniques based on traffic states.
- The research team developed the travel trajectory recording and sharing module for the “GMU Parking Helper” app, to allow for its use for travel behavior data collection platform.
- The research team has developed an algorithm to estimate real-time parking lot usage information through cloud-sourcing. The algorithm will integrate parking lot usage information submitted by users, historical parking lot usage data, and real-time data reported by Mason Parking Department Officials, and the information will be provided to app users in real time. The research team is deploying the apps to collect travel trajectory data to track longitudinal travel behavior.
- The research team developed an algorithm to analyze the travel trajectory data collected using the smartphone app and to test its potential for use as a travel behavior monitoring tool.

Planned Activities: (next quarter):

The research team is finalizing the final project report summarizing all the findings.

1.2.1.4 Research at George Mason University

a) Monitoring Behavior Reactions to Washington Metro SafeTrack Project Using Advanced Travel Data Collection Techniques

Project Status: On-going but nearing completion

Milestone Accomplishments and Dates:

- The research team has designed a smartphone app based longitudinal travel behavior data collection scheme to track travel behavior changes during the Washington Metro SafeTrack project, a series of transit network disruptions that involve either continuous single track or complete shutdown of a metro segment.
- The research team has designed a data collection program based on the smartphone app and tested the deployment through a pilot program.
- The research team is conducting a smartphone-based survey to collect travel behavior data during the Washington Metro Safetrack project.
- The research team is developing algorithm to analyze the trajectory data collected through the smartphone app to infer travel mode, departure time, and routes.

b) Developing a P3 Projects Database to Support Transportation Planning and Policy Analysis

Summary: This project proposes to develop a P3 project database to support transportation planning practice and policy analysis. This project will review existing data in infrastructure finance and develop a data structure that provides a platform for projects of different sizes, financial sources, ownership, delivery methods, and age to be compared and analyzed. It will pool data collected from sources such as TIFIA, InfraAmerica, OECD, and others to build an initial database. Efforts will be dedicated to explore how qualitative features associated with a project could be analyzed using latest data mining techniques.

Project Status: Ongoing

Milestone Accomplishments: The research team:

- investigated the existing database on P3 transportation projects, including the database compiled by TIFIA, InfraAmerica, OECD, and European banks to identify the gap in existing data sources.
- is working with CDM Smith Inc. on a database of toll roads in the U.S., and modelling toll elasticity of demand, a key issue for the user-pay model in both P3 projects and conventional design build projects.
- is developing a model that can inform the decision making process of P3 projects.
- is collecting data on Virginia P3 projects and is identifying the gaps between existing P3 data and data needs for effective evaluation of project delivery method.

Planned Activities:

- The research team will continue to develop the P3 model using existing data and new data to estimate the costs and benefits associated with P3 project delivery mechanism.

- The research will continue with the data collection efforts to identify new experience on transportation P3s in the U.S as new P3 projects are planned or closed.

1.2.1.5 Research at the University of Puerto Rico at Mayaguez

a) Development of Transit Performance Measures using Big Data

Summary: The purpose of this project is to develop real time performance measures using Big Data generated by GPS or AVL (Automatic Vehicle Locator) devices installed in public transit vehicles, and merging that information with transportation demand related data available from other sources, including Census data. This project is working with the Metropolitan Bus Authority system (AMA, for its acronyms in Spanish), which is the main public transit operator in the San Juan Metropolitan Area of Puerto Rico.

Project Status: On going

Milestone accomplishments and dates:

During this reporting performance period, the research team has continued working on three main areas. First, using the software KNIME Analytics Platform (a typical big data software) to continue reproducing the analysis of big data available from the automatic positioning system of the San Juan Metropolitan Area transit system. Second, incorporating the demand related data, obtained from other data sources and the data generated by GPS devices installed in public transit vehicles, to make a new approach based on the merged data. Finally, regarding the third area, our research team has been working in statistical analysis and models to generate new performance measures and has been writing the final report.

Planned activities:

In the next period, our research team plans to use the software to test the performance measures developed and prepare the reporting of our project with recommendations. Workshops and other training activities are also planned as part of the technology transfer activities in this project.

b). Development of a Mobile Computer Application for the Process of Data Collection and Data Sharing for Vehicle Accidents

Summary:

The research project focuses in developing two mobile applications to be used by police officers in Puerto Rico: one for vehicle crash reports (CARS – Car Accident Reporting System) and one for traffic citations (E-TICS – Electronic Traffic Infraction and Citation System). The data collected by these applications could be used for easy sharing between the pertinent agencies as well as investigations focused in reducing the number of vehicle crashes.

Project Status: On going

Milestone Accomplishments and Dates:

The application was optimized in GitHub (a software development platform) and changes were made to the file structure and its organization in order to improve asset usage. We also worked on debugging the application and fixing errors in the WebServices' part of the application and fixing syntax errors due to framework update. An undergraduate student was also hired to work on the application's dashboard. In a very short amount of time (one month) this student got acquainted with the code base of the project and the technologies being used. The student has added a list of people involved in a vehicle crash, designed a customized website, and added user-interaction improvements.

Planned Activities:

The next reporting period will be focused in preparing the final report, polish the mobile application and its website, and reach out to government agencies to show the final product.

1.2.1.6 Research at Calspan / University at Buffalo Research Center

a) An Evaluation of Knowledge Discovery Techniques for Big Transportation Data

Summary: As an extension of the work that was completed for “Developing Highway Safety Performance Metrics in an Advanced Connected Vehicle Environment Utilizing Near-Crash Events from the SHRP 2 Naturalistic Driving Study” this project is investigating the application of knowledge discovery (KD)

techniques to analyze the same data. The primary output of this research will be the identification of methodologies used in other areas of data science that are applicable to transportation safety research.

Project Status: On going

Milestone accomplishments and dates:

- November 20th, 2015: Finalized the compilation of research database s. The datasets being utilized for this project include the SHRP2 Roadway Inventory Data (RID), Archived Clarus weather data and the traffic safety database compiled in CUBRC's previous UTC project.
- January 8th, 2016: Finished ingesting and aligning databases in the Hadoop Distribution environment to enable initial data search, query and mining techniques.
- February 12th, 2016: Compiled a summary list of the advanced analytics and KD techniques which would be run against the ingested data.
- August 3rd, 2016: Presented summary of project at TransInfo Symposium.
- September 23rd, 2016: Completed summary document on the state of the art and best practices for research in big data transportation research.
- January 27th, 2017: Finished performing benchmark testing on the operational environment.
- April 28th, 2017: Completed model evaluation and validation
- September 29th, 2017: Adjusted optimization strategies to improve benchmark measures
- Completed technical work on project, documentation and final report to follow.

Planned activities:

Document components, best practices, state-of-the-art technologies, and lesson learned; summarize overall findings and produce final report. The findings of this project will be provided in written form as well a final report and a journal submission. Parts of this research will also be presented at the Annual TransInfo Symposium in August of 2018.

1.2.2. TransInfo Educational Activities

1.2.2.1 New Interdisciplinary Degree in Sustainable Transportation and Logistics

TransInfo researchers at the University at Buffalo have developed a new interdisciplinary Masters' degree in Sustainable Transportation and Logistics. The new program is a 30-credit, full-time, 3-semester program, jointly administered by UB's School of Engineering and the School of Management. It can also be pursued on a part-time basis to enable those working in industry to benefit from the program. The curriculum consists of five courses that are designed to provide a common platform of relevant engineering principles & practices, coupled with managerial concepts & practices. Upon entering the program, the candidates will complete these five core courses as a cohort, followed by five courses in one of the following two tracks: (1) Sustainable Transportation; and (2) Logistics. The program is designed to train globally competitive graduates who are well rounded technically and managerially, and who intend to assume leadership positions in transportation and logistics, which have emerged as major sectors of the economy. The program welcomed its first students in Fall 2017. Seven more students are expecting to join the program for Fall 2018 (<https://www.buffalo.edu/istl/Education.html>).

1.2.2.2. TransInfo graduate student activities

The TransINFO consortium continues to engage dozens of undergraduate or graduate students in its research activities with each project facilitating the involvement of at least one student. Some highlights of TransInfo engaged students during this reporting period include:

- The TransINFO sponsored project: **Development of a Mobile Computer Application for the Process of Data Collection and Data Sharing for Vehicle Accidents at UPRM** has two students working on topics related to mobile app development and improvements.
- Laiyun Wu, Ph.D. student at UB working on the project: **Inferring Origin-Destination Demand and Utility-Based Travel Preferences in Multi-Modal Travel Environment Using Automatic Fare Collection Data** has presented initial findings at INFORMS Annual Meetings in 2016, 2017 and TRB 2018. He will also present in this upcoming 2018 INFORMS Annual Meeting.

- FHWA T3 webinar: Dr. Qing He was invited by FHWA to conduct a T3 webinar in social media https://www.pcb.its.dot.gov/t3/s160928_Next_Generation_Traveler_Information_Agency_Social_Media.asp
- UB Civil PhD student Yu Cui will develop her dissertation based on the TransInfo funded project “**Using Social Media Data to Enhance Up-to-date Origin-Destination Demand Monitoring**”. She is expected to graduate in the beginning of 2019.
- 8 PhD students are currently being supported by TransInfo funded projects at the University at Buffalo.
- University at Buffalo’s PhD student in Industrial and Systems Engineering supported by the TransInfo funded project “**Predictive Analytics for No-Shows and Cancellations for Paratransit Operations**” has participated in INFORMS 2017.
- The PIs of the **Inferring Origin-Destination Demand and Utility-Based Travel Preferences in Multi-Modal Travel Environment Using Automatic Fare Collection Data** project are working with two MS students in a similar topic area developed from this research project.

1.2.3 Outreach and technology transfer activities

1.2.3.1 TransInfo to Host its Fourth Annual Symposium on Transportation Informatics

The Symposium attracts distinguished practitioners, academicians, policy makers, and industry. The two-day event will feature distinguished speakers including keynote addresses, presentations, panels and guided discussions covering a broad range of topics related to transportation. The focus for this year will be on challenges and opportunities of Connected and Automated Vehicles (CAVs) applications, testing and evaluation (<https://www.buffalo.edu/transinfo/news-and-events/events/annual-symposium/2018.html>).

1.2.3.2. 2018 Transportation Research Board Annual Meeting

The Transportation Research Board (TRB) 97th Annual Meeting took place January 7-11, 2018 at the Walter E. Washington Convention Center, in Washington, D.C. The TransInfo University Transportation Center alongside the University at Buffalo (UB) hosted a reception on Monday evening. In addition to hosting the reception, a large contingent of TransINFO faculty and students participated in the conference with presentations, papers and posters. Shweta Dixit, a PhD graduate in the Transportation Engineering Program of George Mason University, was selected as the 2017 Student of the Year for the Transportation Informatics Tier I University Transportation Center. She was honored by USDOT at a special ceremony held during the meeting.

1.2.3.3. TransINFO Open House @ University of Puerto Rico at Mayaguez (UPRM)

The TransInfo Team at University of Puerto Rico at Mayaguez hosted its annual Open House on January 18, 2018. The event showcased the Car Accident Reporting System (CARS), the mobile computer application developed by TransINFO researchers for documenting and sharing data regarding vehicular accidents. Attendees had the opportunity to test the app through several fictitious crash scenarios and make recommendations for improvement. The developed application will benefit the police workforce, the Puerto Rico Department of Transportation and higher education institutions by providing the means to collect accident data accurately and making it available for further research.

1.2.3.4. Community Outreach @ University of Puerto Rico at Mayaguez (UPRM)

On Thursday December 14th, 2017, the TransInfo team at University of Puerto Rico at Mayaguez (UPRM) joined the Institute of Transportation Engineers (Puerto Rico Section & UPRM Student Chapter) in assisting the Puerto Rico Department of Transportation with a Community Outreach Event. The long-term plan for Puerto Rico's transportation infrastructure and the changes in Puerto Rico's travel patterns after hurricane Maria were discussed during the event. The Car Accident Reporting System (CARS) mobile application developed by TransINFO researchers was also presented to the attendees. On March 20-22, 2018, the application was again presented to the public during the traditional Transportation Week of the ITE-RUM, which took place at UPRM.

1.2.3.5. Self-Driving Vehicles for Smart & Sustainable Mobility - Evaluation and Feasibility Study for Educational and Medical Campuses

Researchers at UB's School of Engineering & Applied Sciences received funding from the New York State Energy Research and Development Authority (NYSERDA) to purchase and test an Olli bus. Olli is a self-driving, electric vehicle that can accommodate up to 12 passengers and uses sensors, including radar and cameras, to monitor the driving environment. UB plans to test the bus later this year. The project will help answer questions that relate to the technical feasibility, safety and reliability of using Autonomous Vehicle (AV) technology and the public policy changes needed to allow for AVs to be driven on New York State public roads. Furthermore, the project will conduct a detailed evaluation of the costs and benefits of using AV technology on a realistic case study involving the Buffalo Niagara Medical Campus (BNMC) in downtown Buffalo.

1.2.3.6. IRG Track 2: Towards Quality Aware Crowdsourced Road Sensing for Smart Cities, NSF Award 1737590

TransInfo Researchers at UB's School of Engineering & Applied Sciences received \$1.0 Million in funding from NSF in support of their proposal: **Towards Quality Aware Crowdsourced Road Sensing for Smart Cities**. The goal of the project is to make the acquisition and dissemination of road/traffic condition information (traffic congestion, road surface defects, malfunctioning traffic regulation infrastructures) accurate, efficient, and timely. *The large-scale proposal was based on the TransINFO funded project "Towards Quality-Aware Big Data Integration for Crowdsourced Road Sensing System".*

1.2.3.7. Miscellaneous Activities

- **Dr. Daniel Hess named Chair of the Department of Urban and Regional Planning at UB:** TransINFO's Daniel B. Hess, professor of Urban Planning who studies the socio-economic dynamics of housing, transportation and land use, has been appointed Chair of the Department of Urban and Regional Planning at UB, effective January 1, 2018.
- **Dr. Qing He receives the 2017 SEAS Early Career Researcher of the Year Award:** Dr. Qing He, Stephen Still Assistant Professor in Transportation Engineering and Logistics (joint appointments in Civil, Structural and Environmental Engineering and Industrial and Systems Engineering) received the 2017 School of Engineering & Applied Sciences Early Career Researcher of the Year Award for the development of innovative models and algorithms for multi-modal transportation. Specific problems tackled include problems arising from highway transportation, rail transportation, and supply chain management and logistics. Domains include connected vehicles, smartphones and social media.
- Portions of the project improving the Service Quality of Bike Sharing Systems via the Analysis of Real-Time User Data were discussed with Michael Galligano, the Executive Director of Shared Mobility Inc., who is the operator of the Reddy Bike share system of the city of Buffalo for future collaboration.
- TransInfo published and delivered **one issue of its newsletter** during the reporting period in an effort to continue to provide information on the Center and updates on transportation informatics more generally. The newsletter mailing continues to increase, with over 525 qualified recipients.
- PIs for the project titled **"Inferring Origin-Destination Demand and Utility-Based Travel Preferences in Multi-Modal Travel Environment Using Automatic Fare Collection Data"** have obtained an access to UB Parking & Transportation database where individual students' card swipes for UB bus system are recorded.
- NSF ERC Proposal: TransInfo researchers, Prof. Qiao and Prof. Sadek, are currently working with Prof. Krishna Rajan of the Department of Materials Design and Innovation (MDI) on an NSF ERC proposal in the area of Active Cyber Transportation Systems (ACTS).
- The PIs for the project "Predictive Analytics for No-Shows and Cancellations for Paratransit Operations" have reached out to a paratransit agency, DARTS (<https://www.dartstransit.com/>), and initiated a collaboration in order to submit a new research proposal in paratransit studies to TRB TRANSIT IDEA.

2. Products

2.1 Publications, conference papers, and presentations

Publications during the reporting period

- Wu L, JE Kang, Y Chung, and A Nikolaev (2017), Monitoring Multimodal Travel Environment Using Automated Fare Collection Data, To be submitted to *Transportation Research Part C: Emerging Technologies*.
- Laiyun Wu, Jee Eun Kang, Younshik Chung, and Alexander Nikolaev “Monitoring Multimodal Travel Environment Using Automated Fare Collection Data: *Data Processing and Reliability Analysis*” in submission.
- A paper, “Modeling Revealed Mode Choices using Emerging Data Sources: A Case Study on the Washington Metro SafeTrack Project”, is under review at *Transportation Research Part C: Emerging Technology* (Project: Monitoring Behavior Reactions to Washington Metro SafeTrack Project Using Advanced Travel Data Collection Techniques).
- Ebadi Ny, JE Kang, and S Hasan (2017), "Constructing Activity-Mobility Trajectory Patterns of College Students based on Smart Card Transaction Data", *International Journal of Transportation Science and Technology*, 6-4:314-329
- Zhang Ay, JE Kang, and C Kwon (2017), "Incorporating Demand Dynamics in Multi-Period Capacitated Recharging Location Planning for Electric Vehicles", *Transportation Research Part B: Methodological*, 103:5-29
- Caceres, H., R. Batta, and Q. He, "School Bus Routing with Stochastic Demand and Duration Constraints", *Transportation Science*, 51(4), 2017, 1349-1364.
- Zhang, Z., Q. He, J. Gao and M. Ni, "A Deep Learning Approach for Detecting Traffic Accidents from Social Media Data" *Transportation Research Part C*, Volume 86, January 2018, pp 580-596.
- Zhang, Z., Q. He, and S. Zhu, "Potentials of Using Social Media to Infer the Longitudinal Travel Behavior: A Sequential Model-based Clustering Method", *Transportation Research Part C: Emerging Technologies*, Volume 85, December 2017, pp 396-414.
- Devari. A, A. Nikolae, Q. He, "Crowdsourcing the Last Mile Delivery of Online Orders by Exploiting the Social Networks of Retail Store Customers", *Transportation Research Part E: Logistics and Transportation Review*, Volume 105, September 2017, pp 105-122.
- L. Lin, J.C. Handley, X. Wen, and A.W. Sadek. A Hybrid Machine Learning Model for Interval Prediction of Short-term Traffic Volume and its Application to Optimal Staffing Level Plan Development. Submitted to *Transportation Research – Part C*.

Presentations:

- Pipicano, W., Valdes, D., Ramos, I., Figueroa C., “*A Metodology to Restructure a Transit System: The Case of San Juan, Puerto Rico*” (Title in Spanish: *Metodología para Reestructurar un Sistema de Transporte Colectivo: Caso de San Juan, Puerto Rico*) LACCEI International Multi-Conference 2017: Global Partnership for Development and Engineering Education,” Boca Raton, FL, July 2017
- A poster “An Evaluation of Knowledge Discovery Techniques for Big Transportation Data” poster describing the research activities to date was produced and displayed at the TransInfo reception at TRB 2018 in Washington, D.C.
- Chuishi Meng, Xiuwen Yi, Lu Su, Jing Gao, Yu Zheng, “*Citywide Urban Traffic Volume Inference with Loop Sensor Data and Taxi Trajectories*”. Proceedings of the 25th ACM International Conference on Advances in Geographical Information Systems, Redondo Beach, California, USA.
- “*Bike Sharing using Toggle Locations and Portable Stations*” was presented at the INFORMS Annual Meeting 2017 held at Houston.

2.2 Website(s) or other Internet site(s)

Several Websites were developed during this period to showcase the eight new TransINFO's research projects (<https://www.buffalo.edu/transinfo.html>). Also a new website is being developed for the Fourth Annual TransINFO Symposium (<https://www.buffalo.edu/transinfo/news-and-events/events/annual-symposium/2018.html>)

2.3 Technologies or techniques

- Experiences of using smartphone app as a travel behavior data collection tool and technologies developed in the project “Developing a Smartphone App Platform to Decipher Travel Behavior” project have been used in the study to assess the impact of the Washington Metro SafeTrack projects.
- The smartphone app, entitled “Mason Parking Helper” is available in both iTunes Store and Google Market. The app is designed to provide real-time parking lot usage information at George Mason University based on data from different sources. The updated version has the capacity of disseminate on campus special event information and sends app users notifications if they choose to receive them.
- The above smartphone app may become a transportation data collection platform that can benefit other studies or data collection efforts.
- The GMU research team is developing the travel trajectory recording and sharing module that will enhance the “GMU Parking Helper” app, and make it an effective travel behavior data collection platform.
- An algorithm was developed to estimate real-time parking lot usage information through cloud-sourcing. The algorithm will integrate parking lot usage information submitted by users, historical parking lot usage data, and real-time data reported by Mason Parking Department Officials, and the information will be provided to app users in real time.
- The research team has designed a smartphone app based longitudinal travel behavior data collection scheme to track travel behavior changes during the Washington Metro SafeTrack project, a series of transit network disruptions that involve either continuous single track or complete shutdown of a metro segment (Monitoring Behavior Reactions to Washington Metro SafeTrack Project Using Advanced Travel Data Collection Techniques).
- The research team has designed a data collection program based on the smartphone app and tested the deployment through a pilot program (Monitoring Behavior Reactions to Washington Metro SafeTrack Project Using Advanced Travel Data Collection Techniques).
- Regarding the mobile application for traffic citations, Electronic Traffic Infraction and Citation System (E-TICS), this application is completely finished in terms of performance. There are some minor changes planned for the next period that focused on the application to be more "user-friendly". The application includes the option of determining the exact location of the traffic violation. This information could be used in future research projects that wish to determine locations in which driver behavior can be considered risky (Development of a Prediction Model for Crash Occurrence by Analyzing Traffic Crash and Citation Data).
- Developed software and installed new GPS systems in the AMA buses to replace the old ones that had problems providing the information that the project was using. The new GPS and the software developed by the research team will be installed in several buses the next month therefore obtaining real time data once again (New Performance Metrics and Operational Strategies based on Bus Location and Passenger Count Data).
- Functional prototype to obtain the data required to continue and finish project. It is expected that the installation of equipment and testing the new software will be performed in May (Application Development for Mobile Data Collection and Sharing for Vehicle Accidents).
- Mobile app on collecting mobile trace data at varying locations and time intervals, based on traffic states learned from the data (Urban System Modeling and Performance Measurement Using Multiple Data)

- Researchers at UPRM have been in conversations also with the UPRM Administration to apply the results of their project: “**Development of Transit Performance Measures using Big Data**” to the trolley system that is available for our academic community in Mayaguez.
- As part of another project that UB at CUBRC collaborate on, "The development of a Transportation Research Informatics Platform (TRIP)" funded by FHWA, UB has developed advanced analytics for the prediction of motor vehicle crashes. In order to apply their methods on a massive scale techniques were evaluated and selected under this program.

2.4 Inventions, patent applications, and/or licenses

- Multi-agent Simulation of Partially Observed Complex Systems, under submission (Variational Inference for Agent-Based Models with Applications to Achieve Fuel Economy)
- Event-Based Social Network Construction Using Distributed Wi-Fi Access Points, under submission (Variational Inference for Agent-Based Models with Applications to Achieve Fuel Economy)

2.5 Other products

- With support from their collaborators at Maven Machines, UB researchers have created a dataset that includes naturalistic truck driving data collected from a set of 253 drivers. The dataset includes over 12 million mirror check events from a period of 15 months. All driver identifying data has been removed from the dataset. This dataset, along with the codes and models developed for the project Predicting changes in driving safety performance on an individualized level under naturalistic driving conditions, will be made available through a GitHub site that has been developed.
- Final version of the research databases needed to support “An Evaluation of Knowledge Discovery (KDD) Techniques for ‘Big’ Transportation Data” have been compiled
- Revised traffic safety database as an update to last reporting period’s database to include the revised elements of the SHRP2 NDS data available through the InSight website (Developing Highway Safety Performance Metrics in an Advanced Connected Vehicle Environment Utilizing Near-Crash Events from the SHRP 2 Naturalistic Driving Study)
- A survey was developed in order to gather data regarding traffic violations and crash history. The survey is available on both paper and electronic forms (Development of a Prediction Model for Crash Occurrence by Analyzing Traffic Crash and Citation Data)
- A day of population and travels on UB’s North Campus, synthesized data set (Variational Inference for Agent-Based Models with Applications to Achieve Fuel Economy)

3. Participants and Collaborating Organizations

Work on TransInfo projects have involved close collaborations across traditional disciplinary lines. For several of our research projects and initiatives, transportation researchers (from the Departments of Civil Engineering) are working very closely with their counterparts in the Departments of Computer Science and Engineering. In addition TransInfo researchers are collaborating with government, private industry, and other academic partners on several projects including (1) Connected Vehicle Pilot Deployment proposal, (2) the USDOT Smart Cities Challenge, (3) MRI Proposal for Connected and Autonomous Vehicle Evaluation and Experimentation, and (4) BNMC Green Commons & Living Transportation Lab.

3.1 Organizations which have been involved as partners

| Organization Name | Location | Contribution to the Project |
|--|--|------------------------------------|
| Committee for Traffic Safety, Puerto Rico | Minillas Government Center, South Tower, Suite 5, Santurce | |
| District Department of Transportation (DDOT) | 55 M Street, SE, Washington, DC 20003 | |

| | | |
|---|---|---|
| Erie County Department of Public Works | Buffalo, NY | |
| George Mason University Parking Department | Fairfax, Virginia | In-kind support |
| Governors Traffic Safety Council (GTSC) | Albany, NY | Provided project guidance in selecting research area |
| CDM Smith Inc. | Fairfax, Virginia | Collaboration on database development |
| Greater Buffalo Niagara Regional Transportation Council (GBNRTC) | Buffalo, NY | Provided database of traffic volumes and turning counts for Erie County, NY |
| INRIX | | Data for research |
| Korea Transport Institute (KOTI) | 370 Sicheong-daero, Sejong-si, 339-007, Republic of Korea | |
| “Llame y Viaje” (Call and Travel; LV) | Puerto Rico | |
| Maven Machines | Pittsburgh, PA | |
| Metropolitan Bus Authority (AMA, by its acronym in Spanish) | Puerto Rico | GPS and APC data sharing, in-kind support (data) |
| Microsoft Research Asia | China | |
| Miami University | Oxford, Ohio | |
| National Science Foundation | Maryland | |
| National Transportation Center at the University of Maryland | Maryland | |
| New York State Department of Transportation – Region 5 (NYSDOT) | Buffalo, NY | In kind labor hours to compile crash statistics |
| New York State Department of Transportation (NYSDOT) | Albany, NY | In kind labor hours to generate extracts from Safety Management System Crash Database |
| New York State Energy Research & Development (NYSERDA) | Albany, NY | |
| New York State Thruway Authority | Albany, NY | In-kind support (data) |
| Niagara Falls Bridge Commission | Niagara Falls, NY | In-Kind support (data) |
| Niagara Frontier Transportation Authority (NFTA) | Buffalo, NY | |
| Niagara International Transportation Technology Coalition (NITTEC) | Buffalo, NY | In-kind support and personnel exchanges |
| Peace Bridge Authority, Niagara Falls Bridge Commission | Buffalo, NY | In-kind support (data) |
| Police Workforce (Hormigueros Municipality) | Puerto Rico | In-kind support |
| Police Workforce (San German Municipality) | Puerto Rico | In-kind support |
| Puerto Rico Department of Transportation and Public Works (DTOP, by its acronym in Spanish) | Puerto Rico | In-kind support (data) |
| Puerto Rico Integrated Transportation Authority | Minillas Government Center, South Tower, De Diego, Santurce | |
| SAMMAT Engineering Services, LLC | P.O. Box 780 Mount Airy, MD 21771 | |
| ShenZhen Urban Transport Planning Center | China | |
| SUNY-Albany | Albany, NY | Collaboration on organizing 3 rd Symposium on Connected and Automated Vehicles |

| | | |
|---|--|---|
| TrafficCast, Inc. | Madison, Wisconsin | Proposal collaborations |
| University at Buffalo | Amherst, NY | Developed advanced analytics for the prediction of motor vehicle crashes. |
| University of Puerto Rico, Medical Science Campus | Gobernador Piñero, San Juan, 00921, PR | |
| University Transportation Research Center (UTRC) | New York, NY | Research Funding and Collaboration on organizing the 3 rd & 4 th Symposium on Connected and Automated Vehicles, the First Annual Symposium on Transportation Informatics. |
| Urban Transportation Associates (UTA) | Cincinnati, OH | APC Installation and file sharing |
| Virginia Department of Transportation | Richmond, VA | In-kind support |
| Virginia Office of Public-Private Partnerships | Richmond, Virginia | |

4. Impact

4.1 What is the impact on the development of the principal discipline(s) of the program?

TransInfo activities are envisioned to help advance the state-of-the-art in the application of advanced data mining, Artificial Intelligence (AI), Social Network analysis and Advanced Statistical and Econometric models to transportation Big Data. Applying such methods is envisioned to result into invaluable insight into how to improve transportation system efficiency, safety, sustainability, resiliency and reliability. It is also envisioned to help support sound transportation decision making through the development and application of appropriate performance metrics. Our research to-date has already resulted in the development of new methods for data analysis. Examples include the recently developed combined MSP-HBDM model for incident duration prediction, and the new methods developed at RPI for probe vehicle data fusion and analysis.

4.2 What is the impact on other disciplines?

TransInfo activities are likely to have an impact on the field of Big Data Analytics, as it pertains specifically to transportation data. The transportation Big Data context has several unique features which distinguish it from other application domains of Big Data. TransInfo research and educational initiatives are thus likely to have an impact on the emerging field of Transportation Informatics and Analytics.

4.3 What is the impact on the development of human resources?

Several graduate students are supported by TransInfo either through fellowships or graduate research assistantships. TransInfo also has held a number of outreach initiatives aimed at encouraging high school students to consider careers in transportation, including the National Summer Transportation Institute at UB. Moreover, TransInfo's project focused on the development of a mobile computer application for vehicle accidents also presents an opportunity in training police officers in using the mobile application for accident reports. We also hope to contribute toward improving the diversity of the workforce.

4.4 What is the impact on physical, institutional, and information resources at the university or other partner institutions?

TransInfo work is helping build several important data repositories which can help support future research in the area of Big Data Analytics. These repositories could be made available to other researchers via our website to promote research and technology transfer in this field of inquiry.

4.5 What is the impact on technology transfer?

The vision is for many of the transportation research projects undertaken by TransInfo to result in either products which can be implemented to address transportation challenges or improve the traveler experience (e.g., the smart phone app which was already developed for predicting border crossing delay), or in

strategies and/or policies to improve transportation system performance and to support sound decisions regarding transportation investments. We hope to work with our stakeholders and partners in the future to make sure that TransInfo research and educational initiatives are having a positive impact on technology transfer.

4.6 What is the impact on society beyond science and technology?

The benefits to individual drivers and society in general with respect to the Android smartphone application, the Toronto Buffalo Border Waiting (TBBW) can be tremendous. For individual drivers, the app can help them choose the right crossing and the arrival time which would minimize their wait time, thereby saving them time, gas, and money. For society, the app can help reduce the cost of border crossing delay on a region's economy as well as on the environment in the area, which is estimated to be in the order of billions of dollars annually in lost business productivity, wasted fuel, traffic-related pollutants and associated health hazards. The same can also be used about other TransInfo initiatives and projects such as the work aimed at improving incident management in Northern Virginia, and at building a P3 (Public-Private-Partnership) projects database to support transportation planning and policy analysis. Such projects have the potential to benefit society at large and to save tax-payers millions of dollars.

5. Changes/Problems

- Regarding the project “Predictive Analytics for No-Shows and Cancellations for Paratransit Operations”, there was a small delay in receiving the final dataset from our partners at Maven Machines. The full dataset was received in October 2017, at which point data preparation and integration began. This has led to a shift in the timeline for the other tasks of the project. The proposed scope of the project remains as planned.
- **The project: “Using Social Media Data to Enhance Up-to-date Origin-Destination Demand Monitoring”** found sampling bias of social media is a critical issue for OD estimation during our investigation. Therefore, we are conducting a demographics-estimation study to correct the sampling bias.
- The project “City-wide Transportation Information Estimation with Heterogeneous Urban Data” requested an extension of the project to further improve the traffic volume estimation techniques and build a prototype to illustrate the estimation results.