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

October 31, 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: April 1, 2018 – September 30, 2018

Report Frequency: Semiannual

Signature:

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.

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 ran test cases, and we are refining our approach and data sets.

Planned Activities:

The research team is currently at the final editing stage and will submit the following working papers in the next few weeks:

- Wu L, JE Kang, Y Chung, and A Nikolaev, Monitoring Multimodal Travel Environment Using Automated Fare Collection Data: Data Processing and Reliability Analysis.

- Wu L, JE Kang, Y Chung, and A Nikolaev, Inferring Origin-Destination Demand and User Preferences in a Multi-modal Travel Environment Using Automated Fare Collection Data.

b). 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:

- Data preprocessed for deep learning model;
- Developed super-resolution convolutional NNs to predict pick-up events in Shenzhen
- Developed a dynamic programming algorithm to maximize the average profit of an eTaxi in long term 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.
- 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:
- Developed a time-dependent graph representing road networks which includes road segments and charging stations and designed a time-dependent recursion strategy to generate optimal driving route with consideration of both long-term profits from recharging actions and short-term profits from carrying passengers.

Planned Activities:

Prepare to submit one or two papers on relevant conferences.

c). 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 estimated based on possible influencing factors, including travel speed, points of interests and weather conditions. 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. 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 on correctly inferring city-level traffic volumes.

- 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. We plan to evaluate the designed statistics on the traffic data collected in Guiyang, China, and visualize the results on the map.

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

Summary:

The principal objective of this project is to develop a predictive statistical framework to efficiently estimate the ability of a bike-sharing system to serve incoming bike requests. By mining user data collected from the system's smartphone app, an operator can utilize the proposed models to predict the likelihood that any potential user who desires to use the system decides to do so under the given the system's conditions the user encounters. These include the location from where a bike request originates and its proximity to the nearest available bike, the weather conditions, the time of the day, and the customer profile, among others. The proposed statistical models are coupled with an operational bike redistribution model to analyze the cost-effectiveness of triggering quick bike redistribution tours to raise the service quality of the system to desired levels.

Project Status: Completed

Conclusions:

The statistical framework is comprised of five comprehensive components: (1) a clustering component which groups users based on their demographics and usage behavior; (2) a demand forecasting component that estimates the OD demand matrix using the trip information collected by the system for each of the clusters generated before; (3) a rejection predictor component that acts as a classifier in order to identify the probability that a person from a given cluster decides to use the system; (4) a bike redistribution component modeled as a mixed-integer optimization problem that identifies fast redistribution routes to test whether performing a redistribution trip is cost effective; and (5) a data preparation component that prepares the raw data and extracts the relevant information that is later fed into the statistical framework.

Planned Activities:

A portion of the technical implementation of the proposed framework, particularly the training component is still an ongoing endeavor that requires further considerations. We are currently refining the data sets to better train the rejection predictor.

e). 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. The findings can be leveraged for the development of educational tools, policy strategies for risk-taking drivers and road safety programs.

Project Status: On going

Milestone Accomplishments and Dates:

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

- Statistical modeling of perceived and observed aggressive driving behavior is conducted. A bivariate grouped random parameters probit model is estimated for the entire dataset. On the basis of the results of the full-dataset model, likelihood ratio tests are also conducted in order to identify if the estimation of separate models is warranted for male and female participants, and for the investigation of aggressive driving behavior due to driver's fatigue, and under distracted driving conditions.
- The preliminary results of the likelihood ratio tests warrant estimation of separate models for perceived and observed aggressive driving behavior for non-fatigued and fatigued participants, distracted and non-distracted participants, as well as for male and female participants.
- The research team has estimated preliminary models of perceived and observed (based on the driving simulation experiments) aggressive driving behavior arising from driver's fatigue, and distracted driving conditions.

Planned Activities:

- Completion of statistical analysis and finalization of the model estimation results.
- Estimation of (pseudo-) elasticities of the explanatory variables affecting observed and perceived aggressive driving behavior.
- Interpretation of the estimation results and discussion on the variations of factors affecting aggressive driving behavior across different groups of drivers.
- Preparation of 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.

f). 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. It includes the following tasks: 1) model input parameters for characterizing workload; 2) quantify changes in driving performance based on mirror checks and system alerts; and 3) investigate data-driven modeling approaches for performance prediction.

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 and system alerts and evaluate these changes with respect to gold standard guidelines
  - Investigate data-driven modeling approaches for DSP prediction, including structural analysis and machine learning approaches
- Significant Results:
  - From April-September 2018, we were able to extract the main input parameters for workload, including trip duration, time of day, median speed, percent of time looking down at the phone screen, and percent of events when the driver was on the phone.
  - The response variable of average mirror check rate was used for modeling changes in driving performance. Independent of trip duration, the average mirror check rate was just under 6 per minute (see figure). The standard recommendation for commercial drivers is to check their mirrors every 8-10 seconds (6-7.5 per minute). The mirror check rate was considered both as a continuous and dichotomous variable. For the dichotomous form, we considered a gold

standard target rate of 6 mirror checks per minute. Using a classification tree, with 5-fold cross validation, we achieved a classification accuracy of 63% with an AUC of 0.66.

Planned Activities:

We are working to improve the models for predicting changes in driver safety performance. Further input, explanatory variables are being investigated to better explain changes in mirror check rates. We will investigate the influence of the predictors on the change over time. A manuscript submission is planned following completion of data analysis and modeling.

g). Border Crossing Delay Prediction

Summary: This research developed several models for predicting short-term border crossing traffic, along with queueing models for predicting the anticipated delay. Based on these models, an Android smartphone application 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: Nearing completion

Milestone Accomplishments and Dates:

- Our most recent paper on this work has been published in the *Transportation Research - Part C.* The paper entitled, *A Hybrid Machine Learning Model for Interval Prediction of Short-term Traffic Volume and its Application to Optimal Staffing Level Plan Development. Transportation Research,* describes the development of 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. The predicted intervals are then utilized to develop optimal staffing plans for a border crossing agency.

Planned Activities: Prepare Final report.

h). 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: Nearing completion

Planned Activities: Prepare the Final Report.

j). Testing and Evaluation of Connected and Autonomous Vehicles

Summary: TransInfo researchers at the University at Buffalo are involved in an-going project, primarily funded by New York State Energy and Research Development Authority (NYSERDA) and New York State Department of Transportation (NYSDOT) aimed at evaluating the safety and reliability of the self-driving shuttle, Olli. The shuttle, which is mostly 3D printed, is capable of seating eight passengers and reaching 25 mph. It is also a fully electric vehicle. Olli is a level 4/5 Autonomous Shuttle, capable of operating to the afore mentioned levels in a fixed route, fixed stop manner in a controlled environment.

Project Status: The testing and evaluation of Olli has begun. A series of tests were designed. The study intended to evaluate performance under normal and inclement weather conditions.

Planned Activities: Continue the testing and evaluation.

1.2.1.2 Collaborative Research between UB & George Mason University

a) Using Social Media Data to Enhance Travel Behavior Analysis

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. The second step removes the sampling bias of social media from user's accurate demographics. The third step estimates the origin-destination demand.

Project Status: On going

Milestone Accomplishments and Dates:

This study first linked travelers' Twitter account with their Facebook account, and verified their demographics from Facebook data. We then conducted manual labeling of more than 1500 Facebook users and their associated tweet account with two undergraduate students and one graduate student. Next, we obtained around 1000 valid users with correct demographics labels. Several models were then proposed for predicting social-demographics. Afterward, the study resampled social media data and compared it to the 2009 California Household Travel Survey data. The resampled data showed comparable characteristics to the survey data. Two papers were submitted for presentation and/or publication. The first was submitted to the 2019 TRB Annual Meeting, and the second was submitted to Transportation Research Part A.

Planned Activities:

- Write final report; work on the Buffalo GPS & social media data for activity modeling.

1.2.1.3 Collaborative Research between UB & UPRM

a). 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, and (iii) to evaluate the value of the classification models' predictions in planning and operations.

Project Status: On going.

Milestone Accomplishments and Dates:

- Machine learning tasks finalized by UPRM team. Initial results generated on trip booking and Routing models
- Meeting with DARTS executive director in Hamilton, Canada, where the researchers presented their work and agreed with the director to collaborate on a future project.
- Presentation at 2018 4th Annual Symposium on Transportation Informatics, Buffalo, NY

Planned Activities:

- Finalize research findings and finish writing the final report.

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

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. Accomplished on 01/30/2017
- The research team has designed a data collection program based on the smartphone app and tested the deployment through a pilot program. 100% on 01/30/2017.
- The research team is conducting a smartphone-based survey to collect travel behavior data during the Washington Metro Safetrack project. 100% on 06/30/2018
- The research team is developing algorithm to analyze the trajectory data collected through the smartphone app to infer important travel behavior information including the mode, departure time, and routes. 100% on 09/30/2018
- The research team is compiling the final report to summarize all research findings. 50% on 09/30/2018

Planned Activities:

- One presentation, "Inferring Multimodal Day-to-day Travel Dynamics using Smartphone App Data", will be given at INFORMS Annual Meeting, Phoenix, AZ, November 4-7, 2018

- An invited seminar at NYU, Assessing Travel Behavior Responses to Washington Metro SafeTrack Project Based on an Integrated Data Platform, October 29, 2018.
- A second paper, "Modeling Revealed Mode Choices using Emerging Data Sources: A Case Study on the Washington Metro SafeTrack Project", is under review at Transportation.
- The research team is compiling the final report to summarize all research findings

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. The project will pool data collected from sources such as TIFIA, InfraAmerica, OECD, and others to build an initial database. The strength of this database will be demonstrated through case studies and pilot transportation planning and policy analysis.

Project Status: Completed

Milestone Accomplishments and Dates:

- The research team reviewed literature and had a teleconference with researchers from USDOT Bureau of Transportation Statistics and International Transport Forum, Organization for Economic Co-operation and Development (OECD) to discuss the data need to support decision making process of Public-Private Partnership in transportation.
- The research team investigated the existing database on P3 transportation projects, including database compiled by TIFIA, InfraAmerica, OECD, and European banks to identify the gap in existing data sources.
- The research team worked with CDM Smith Inc. on a database of toll roads in the U.S. The research team is developing a model for toll elasticity of demand, a key issue for the user-pay model in P3 projects and conventional design build projects, using empirical data.
- The research team developed a model that can inform the decision making process of P3 project using data collected through the case studies.

Planned Activities:

- Due to the success of this project, the research received a gift from Cintra Inc. to study the performance of managed toll road network in the Dallas Fort-Worth area.
- Two papers are currently under review for journal publication.

1.2.1.5 Collaborative Research between RPI & 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: Completed

Milestone Accomplishments and Dates since start of the project:

- The research team reviewed existing literature on using smartphone apps for transportation data collection. The team then developed an app entitled "GMU Parking Helper" in both iOS and Droid environment.
- Worked with Mason Parking Department to collect historical parking lot usage data.
- The RPI team completed the traffic state estimation from reduced GPS data task, by specially designed sampling techniques based on traffic states.
- Developed the travel trajectory recording and sharing module which l enhanced the "GMU Parking Helper" app, and made it an effective travel behavior data collection platform.
- 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.
- Deployed the apps to collect travel trajectory data to track longitudinal travel behavior, and developed an algorithm to analyze the travel trajectory data.

Planned Activities:
Due to the success of this project, the research is awarded a new project, "Improving Inventory and Investment of Bicycle and Pedestrian Facility through Targeted Public Outreach", by the Virginia DOT to investigate innovative ways based on smartphone apps and crowd-sourcing approach to collect inventory and feedback data from users on biking and pedestrian facilities.

1.2.1.6 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 devices installed in public transit vehicles, and merging that information with transportation demand related data available from other sources. This type of synthesized real time information has the capability to improve decision making at the operational and planning levels. This project is taking place using 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 prepared the first draft of the final report that is currently under review by the PI's. The team summarized our work on three main areas:

- 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.
- Merging the demand related data and the data generated by GPS devices installed in public transit vehicles, to propose a new integrated approach based on the merged data.
- The statistical analysis and models necessary to generate new performance measures.
- Our research team used the software to test the performance measures developed and prepared the reporting of our project with recommendations.

Planned activities:
The next reporting period will be focused in preparing the final report.

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 mobile application CARS has been finalized, including the web services application. The research team is focusing their efforts in writing the final report and coordinate meetings with pertinent agencies in order to present the results of the research.
Planned Activities:
The next reporting period will be focused in preparing the final report and reaching out to government agencies to show the final product.

1.2.1.7 Research at CUBRIC / University at Buffalo Research Center

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

Project Status: Completed
Summary: In an ever growing internet of things (IoT) environment the amount of data produced and available for analysis is growing exponentially. In order to access and process these data sources new

methods and tools are needed that can sufficiently process big data as well as ensure the quality and completeness of the data. With that in mind, this research proposes a prototype system that would enable researchers to efficiently work with 'Big Data' sources.

The proposed prototype system employs an informatics approach which considers how transportation safety datasets are best accessed and ingested, and how they can be rapidly processed using a distributed parallel architecture. This design will produce an analytic framework that looks to the future and provides system extensibility to facilitate the incorporation of new sources of data as they become available, while also considering future data mining opportunities.

The proposed prototype architecture has unique characteristics that address the formidable challenges of multisource data alignment and analytics. These characteristics include the alignment of multiple, disparate data sources via a common data model; the rapid integration of new data sources; mature entity and event co-reference resolution capabilities; and advanced analytics such multi-source social network analysis and location normalization. A key feature to this prototype system that there is no need for a data warehouse, which might otherwise act as a common data model. Instead, a non-data-warehousing solution is built on cloud-enabled technologies. On demand, raw data is translated into a singular graph structure. The system in essence operates as a virtual fusion center, and can be installed on any cloud system. Cloud technologies such as Hadoop and Storm allow the system to take full advantage of the parallelism and scalability that is native to the cloud. The overall system architecture has been carefully designed so that it is compatible with the technologies currently employed in the state-of-the-art cloud environments.

Identify 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 the 2nd 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.
- March 31st, 2018 Completed technical work on project, documentation and final report to follow.
- August 10th, 2018 presented summary of project at the 4th TransInfo Symposium.
- September 30th, 2018 Completed final report.

Planned Activities:

Journal submissions are currently being prepared.

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 (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"** will be presenting at the upcoming 2018 INFORMS Annual Meeting.
- 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.
- Nine PhD students are currently being supported by TransInfo funded projects at UB.
- UPRM Undergraduate students who participated in the research project "**Predictive Analytics for No-Shows and Cancellations for Paratransit Operations"** during the last two quarters were trained on the use of software to fit machine learning models.
- Two UB graduate students who participated in the project "**Predictive Analytics for No-Shows and Cancellations for Paratransit Operations"** were trained on developing routing models. This training will be continued throughout the Fall Semester 2018. These two graduate students will travel to INFORMS annual meeting.

1.2.3 Outreach and technology transfer activities

1.2.3.1. TransInfo's Fourth Annual Symposium on Transportation Informatics
Transportation and big data professionals from academia, industry and government gathered for the Fourth Annual Symposium on Transportation Informatics, an event hosted by Transportation Informatics Tier I University Transportation Center (TransInfo) at its lead institution, the University at Buffalo on August 9th and 10th, 2018. Both days of the symposium were teeming with informative lectures, discussions, and enlightening activities. The Symposium showcased not only the experience and innovative thinking of our speakers and guests, but TransInfo's ability to present Transportation Informatics in a way that was both stimulating and relevant. The focus for this year was on the challenges and opportunities of Connected/Automated Vehicles (CAVs) applications, testing and evaluation with presentations, panel discussions and keynote addresses from Alain L. Kornhauser, PhD, Professor of Operations Research & Financial Engineering at Princeton and Raj Rajkumar, PhD, George Westinghouse Professor at CMU. More details are available at: (https://www.buffalo.edu/transinfo/news-and-events/events/annual-symposium/2018.html).

1.2.3.2. Self-Driving Vehicles for Smart & Sustainable Mobility - Evaluation and Feasibility Study for Educational and Medical Campuses
As alluded to above, TransINFO researchers at UB's School of Engineering & Applied Sciences (SEAS) received funding from the New York State Energy Research and Development Authority (NYSERDA) to purchase and test an Olli bus. 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. In addition, 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.

Regarding the project's potential for the state's future, Alicia Barton, President and CEO of NYSERDA, said: "The arrival of the Olli shuttle represents a significant milestone in advancing Governor Cuomo's commitment to providing cleaner technologies to reduce harmful emissions from the state's transportation sector. The Olli project serves as another example of the private public collaboration that is driving the expansion of our clean energy economy and innovative technologies." The Olli bus was unveiled at the Fourth Annual Symposium on Transportation Informatics at UB's North Campus.

1.2.3.4. Miscellaneous Activities

- TransInfo Researchers at UB's School of Engineering & Applied Sciences received $1.0 M 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**".
- Due to the success of the project **"Developing a Smartphone App Platform to Decipher Travel Behavior**", the research team is awarded a new project, "Improving Inventory and Investment of Bicycle and Pedestrian Facility through Targeted Public Outreach", by VDOT to investigate innovative ways based on smartphone apps and crowd-sourcing approach to collect inventory and feedback data from users on biking and pedestrian facilities.
- Due to the success of the project **"Developing a P3 Projects Database to Support Transportation Planning and Policy Analysis",** the research team received a gift from Cintra Inc. to study the performance of managed toll road network in the Dallas Fort-Worth area. Cintra Inc. is a leading private-sector transportation infrastructure company in the U.S. transportation P3 market and will lead the express lanes on I-66 outside the Beltway project.
- The PIs for the project "**Predictive Analytics for No-Shows and Cancellations for Paratransit Operations**" have visited the Disabled and Aged Regional Transportation System (DARTS @ https://www.dartstransit.com/) site in Hamilton, ON. The PIs are collaborating with DARTS on a new research proposal in paratransit studies to be submitted to the TRB TRANSIT IDEA based on the results of this research.
- A seminar on Paratransit Operations was organized for a group of employees of the Metropolitan Bus Authority of Puerto Rico in March 13, 2018. Concepts developed in the project "**Predictive Analytics for No-Shows and Cancellations for Paratransit Operations**" along with a performance evaluation of the paratransit system operated by the Metropolitan Bus Authority were discussed in the seminar**.** This presentation helped the participants understand, relative to other systems in the US, the scope of the problems of the system they are managing.
- An event is being organized at the University of Puerto Rico in Mayaguez (UPRM) which will discuss, among other things, the use of predictive analytics to solve transportation problems. The event will be advertised as a school-wide dialogue on transportation problems on campus.
- Presentations about the research project "**Development of a Mobile Computer Application for the Process of Data Collection and Data Sharing for Vehicle Accidents**" have been given to undergraduate students of the Civil Engineering Department at the University of Puerto Rico, Mayaguez Campus with the purpose of promoting graduate studies related to Transportation Engineering.

- **The PIs of the "Development of Transit Performance Measures using Big Data" are in discussion with** the UPRM Administration in order to apply the results of this project to the trolley system that is available to the academic community in Mayaguez.
- The research work from the project **"Monitoring Behavior Reactions to Washington Metro SafeTrack Project Using Advanced Travel Data Collection Techniques"** has been featured by the Washington Post 6 times.
- TransInfo published and delivered **three issues 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.
- An invitation seminar is given at NYU, Assessing Travel Behavior Responses to Washington Metro SafeTrack Project Based on an Integrated Data Platform, October 29, 2018.
- The research Monitoring Behavior Reactions to Washington Metro SafeTrack Project Using Advanced Travel Data Collection Techniques has attracted board interest from the research community. Part of the research was subsequently funded by the National Science Foundation through a project entitled "RAPID: Transit Network Disruption, Service Reliability, and Travel Behavior"

2. Products for this Reporting Period

2.1 Publications, conference papers, and presentations

Publications

- 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. Transportation Research Part C 92 (2018) 323–348.
- Jeong Yun Kweun and Shanjiang Zhu, Modeling Revealed Mode Choices using Emerging Data Sources: A Case Study on the Washington Metro SafeTrack Project", currently under review at Transportation Research Part C: Emerging Technology.
- Jeong Yun Kweun and Shanjiang Zhu, Measuring Demand Elasticity for US Toll Roads: An Aggregate Analysis of Panel Toll Data, currently under review at Transportation Research Part A
- Jeong Yun Kweun and Shanjiang Zhu, Toll Price Elasticity of Travel Demand: A Meta Study of Current Evidence currently under review at Transportation Research Part A.
- Cui, Y., Q. He, "Inferring Twitters' Demographics to Correct Sampling Bias of Social Media Data in Travel Behavior Analysis", submitted to Transportation Research Part A.

Presentations:

Fourth Annual Symposium on Transportation Informatics – August 9-10, 2018:

- Cavuoto, Lora, "Commercial Driver Workload and Risky Behavior Evaluation using Naturalistic Driving Data"
- Cui, Y., "Sampling Bias Correction in Using Social Media for Travel Behavior Analysis by Predicting the Demographics of Social Media Users"
- Majka, Kevin, "An Evaluation of Knowledge Discovery & Dissemination Techniques for Big Data Transportation Research"
- Rodriguez-Roman, Daniel, "Predicting Trip Cancellations and No-Shows in Paratransit Operations"
- Wu, Laiyun, "Inferring Origin-Destination and User Preference in Multi-modal Travel Environment by Using Automated Fare Collection data"

- Zhu, Xiaohang, "Paratransit Routing and Overbooking: Accounting for No-Shows and cancellations"
- Cruzado, Ivette, "CARS Mobile Application, Its Side Projects, and Future Work"
- Wang, Enshu, "A Cost-effective Routing and Recharging Recommendation System for Electric Taxi Drivers"
- Zhu, Shanjiang, "Assessing Travel Behavior Responses to Washington Metro SafeTrack Project Based on an Integrated Data Platform"
- Bartlett, Andrew, "Predicting Border Crossing Delay through Newly Available Real-Time Data and Deep Learning Methods"

He, Qing, "Inferring Twitters' Demographics to Correct Sampling Bias of Social Media Data in Travel Behavior Analysis", NYSDOT, September 2018

Poster presentation at the University of Cincinnati Education and Research Center Pilot Research Symposium on Occupational Safety and Health titled: "Predicting changes in driving safety performance on an individualized level under naturalistic driving conditions"

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 iTune 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 at GMU 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 at GMU 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).
- The research team at UPRM has developed the mobile application CARS – Car Accident Reporting System.
- The research team at UPRM has developed the mobile application for traffic citations, Electronic Traffic Infraction and Citation System (E-TICS). 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).
- The research team at UPRM has developed software and installed new GPS systems in the AMA buses to replace the old ones.
- 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"
- 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	
DARTS	Hamilton, Ontario, Canada	
District Department of Transportation (DDOT)	55 M Street, SE, Suite 400, 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 count movements 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 for selected locations throughout Erie County, NY
New York State Department of Transportation (NYSDOT)	Albany, NY	In kind labor hours to generate extract 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, 17th Floor, Ave.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 the third Symposium on Connected and Automated Vehicles and on submitting a proposal for Connected Vehicle Deployment in NY State.
TrafficCast, Inc.	Madison, Wisconsin	Proposal collaborations
University at Buffalo	Amherst, NY	Developed advanced analytics for the prediction of motor vehicle crashes.
University of Dayton	Dayton, OH	
University of Puerto Rico, Medical Science Campus	Gobernador Piñero, San Juan, 00921, Puerto Rico	
University Transportation Research Center (UTRC)	New York, NY	Research Funding and Collaboration on organizing the 3rd &4th Symposium on Connected and Automated Vehicles, the First Annual Symposium on Transportation Informatics, and on submitting a proposal for Connected Vehicle Deployment in NY State.
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 M5P-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 research team for the project: **"Using Social Media Data to Enhance Up-to-date Origin-Destination Demand Monitoring"** found a sampling bias of social media to be a critical issue for OD estimation during their investigation. A demographics-estimation study was conducted to correct the sampling bias from social media data. After these two studies, not have enough time for OD estimation with social media data was left. Hence, the title of the project was changed to "**Using Social Media Data to Enhance Travel Behavior Analysis**".