The Exploratory Advanced Research Program Fact Sheet

Knowledge Discovery in Massive Transportation Datasets

Merging Information from Disparate Sources to Enhance Traffic Safety

Exploratory Advanced Research . . . Next Generation Transportation Solutions

Broad adoption of engineering and policy advances, including air bags, highway safety barriers, and distracted driving laws, contributes to increased vehicle safety on our Nation's roadways. Although the number of deaths and injuries from crashes decreased slightly in 2017, the number of 2017 deaths is still at a level not seen since 2007 (National Safety Council 2017). One promising avenue for reducing crashes lies in extracting and analyzing safety-related information from vast and expanding datasets related to driver behavior, vehicle performance, traffic patterns, weather, and infrastructure characteristics. Identifying and making sense of this information will require new techniques. The Federal Highway Administration (FHWA) Exploratory Advanced Research (EAR) Program is supporting research projects that can process massive amounts of transportation-related data from structured, semistructured, and unstructured datasets using open-source tools and technology. The Palo Alto Research Center, Inc. (PARC) is developing automated methods to integrate information from large unrelated datasets. CUBRC, a Buffalo, New York-based systems integration research organization, is developing a layered infrastructure to ingest, store, analyze, and display information.

Acquiring and Compiling Big Data for Traffic Safety

For decades, traffic safety researchers have developed and expanded datasets that describe human behavior, vehicle conditions, and other contextual information related to highway crashes. FHWA's Highway Safety Information System (HSIS) compiles quality data on accident, roadway, and traffic variables collected by States for managing highway systems and studying safety. The second Strategic Highway Research Program (SHRP2) includes a naturalistic driving study (NDS), a resource that includes trip summary records describing more than 3,400 drivers and vehicles involved in roughly 36,000 baseline driving events, including crashes and near-crashes. A related SHRP2 Roadway Information Database (RID) contains detailed information about NDS trips on the most frequently traveled roadway sections, including roadway curvature, number and type of lanes, intersections, guardrails and barriers, and lighting.

Other sources of information, such as Clarus roadway-weather data and video logs that capture roadway features and characteristics, can provide important data related to traffic safety. "Merging traffic-related information from disparate sources makes it possible to detect safety issues that might not be identified by looking at traditional datasets only," says Ana Maria Eigen of FHWA's Office of Safety Research and Development. EAR Program-supported researchers at PARC are developing automated machine learning methods that will replace slower manual methods to extract, clean, and restructure data. PARC is using video, radar, and still photography information gathered at Chicago intersections. Tools developed through this project will be refined for use with similar data-rich traffic information resources.

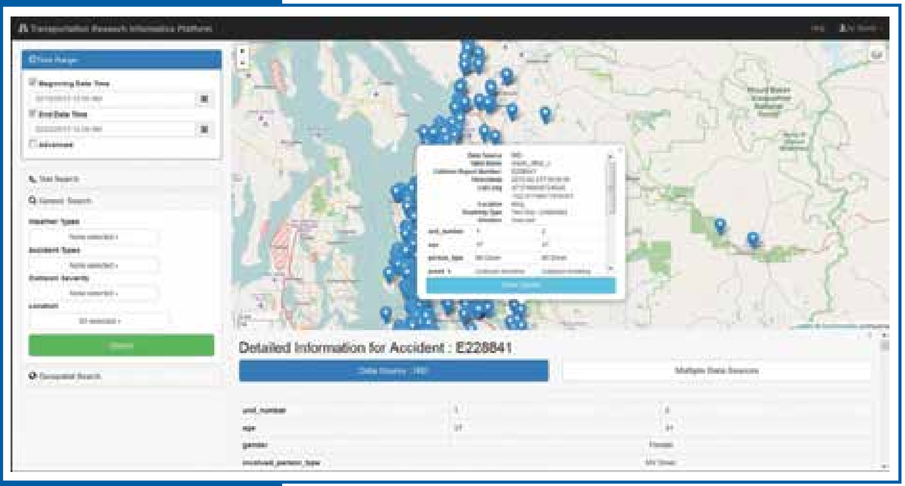
A Platform for Integrating and Analyzing Big Data

The Transportation Research Informatics Platform (TRIP) is designed to make massive amounts of transportation-related data accessible for knowledge discovery and analyze the data to reveal patterns related to traffic safety. At CUBRC, researchers developed TRIP in two layers. The first layer involves the platform itself on a LINUX operating system foundation. A second layer integrates open-source tools to

U.S. Department
of Transportation
Federal Highway
Administration

Knowledge Discovery in Massive Transportation Datasets

Merging Information from Disparate Sources to Enhance Traffic Safety



© *2017 CUBRC.* The TRIP User Interface allows users to set up search parameters, such as weather, accident type, and collision severity, and look for driver patterns.

Photo page 1: © *Gwoeii/Shutterstock.com*

ingest, transform, align, and store structured and unstructured data. TRIP's processing, warehousing, and query functions rely on programs such as Apache and SQL. To provide common and well-supported maximum flexibility for analysts and researchers, TRIP incorporates linkages to many popular analytics packages and visualization tools.

TRIP's dashboard can give users a rapid and versatile method for visualizing streaming data, as well as historical information such as HSIS information on State road crashes, traffic volumes, and roadway characteristics (e.g., curve and grade); RID information on roadway geometrics; supplemental data on historical crashes, volumes, weather, traffic laws, safety campaigns, and work zones; data from the Clarus Initiative, which provides information on atmospheric weather and roadway surface conditions; and imagery data of Nexrad weather information from the Iowa Environmental Mesonet. Nexrad and mesonets are both networks of weather stations that collect data on local weather conditions such as precipitation, temperature, and wind speed. CUBRC is testing TRIP with sample datasets from the Seattle, Washington, region.

"The TRIP project will allow researchers to leverage the decades of legacy data and legacy systems that exist, work with data sources not typically considered in the safety domain, and extend our understanding beyond the conventional wisdom on safety countermeasures," says James Pol of FHWA's Office of Safety Research and Development.

EXPLORATORY ADVANCED **RESEARCH**

What Is the Exploratory Advanced Research Program?

The EAR Program addresses the need for longer term, higher risk research with the potential for transformative improvements to transportation systems. The EAR Program seeks to leverage advances in science and engineering that could lead to breakthroughs for critical, current, and emerging issues in highway transportation by experts from different disciplines who have the talent and interest in researching solutions and might not do so without EAR Program funding.

To learn more about the EAR Program, visit http://highways.dot.gov/research/exploratory-advanced-research. The website features information on research solicitations, updates on ongoing research, links to published materials, summaries of past EAR Program events, and details on upcoming events.

Learn More

For more information about the PARC high-performance data fusion project, contact Ana Eigen at 202-493-3260 (email: ana.eigen@dot.gov). For more information about the development of TRIP, contact James Pol at 202-493-3371 (email: james.pol@dot.gov).

National Safety Council, Statistics Department. (2017). *NSC Motor Vehicle Fatality Estimates*, Itasca, IL. Available online: https://www.nsc.org/Portals/0/Documents/NewsDocuments/2018/December_2017.pdf, last accessed September 14, 2018.

Publication No.
FHWA-HRT 17-117
HRTM-30/02 18(WEB)E