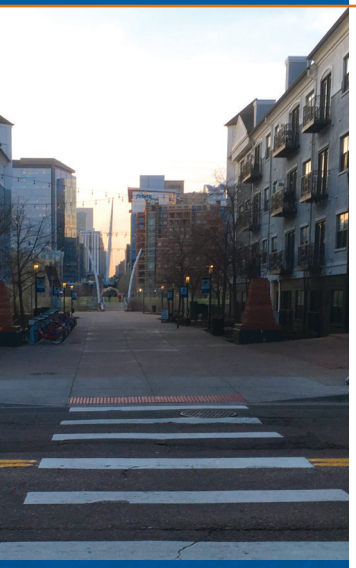
Realistic Artificial Datasets:

Objective Evaluation of Data-Driven Safety Analysis Models

Exploratory Advanced Research . . . Next Generation Transportation Solutions



Crashes occur because of complex interactions between multiple variables, including driver behavior, environmental context, roadway design, and vehicle dynamics. Data-driven safety analysis (DDSA) models help State and local agencies quantify safety data, identify high-risk roadway features, and predict the effects of proposed safety measures. However, even when a model performs well overall, it may not accurately represent the interactions between variables for a specific location or crash because the underlying relationships in the real world are unknown. One proposed solution is to generate realistic artificial datasets (RAD) with predetermined safety relationships built into them. Since these are known, the RAD can serve as a testbed, revealing how well a model reflects those underlying cause and effect relationships.

Two projects supported by the Federal Highway Administration's (FHWA) Exploratory Advanced Research (EAR) Program are generating RADs. The researchers working on "Multidisciplinary Initiative on Methods to Integrate and Create Artificial Realistic Data" are developing open-source datasets for urban interchanges. For "Development and Application of a Disaggregate Artificial Realistic Data Generator for Computationally Testing Safety Analysis Methods," the researchers are building a RAD-generating framework.

RAD Using Simulator Data and Machine Learning

A University of Missouri research team is working with colleagues from Iowa State University and Texas Tech University to apply statistical and machine learning methods to generate open-source RADs capable of revealing how well a given model performs.

"The outcome for our project would be a dataset that other researchers or practitioners who work for State transportation agencies, local municipalities, and so forth, can all use...to quickly find out if a method is the best at teasing out those cause/effect relationships or if it's only good enough for accurately predicting the overall crash rates or crash frequencies," explained Praveen Edara, principal investigator from the University of Missouri.

The researchers selected urban interchanges as a representative facility type for the project because these roadway features lack accurate crash data and are overrepresented in real-world crash data, with more than 50 percent of fatal or injury crashes involving intersections.[1] They are also using de-identified data from the second strategic highway research program naturalistic driving study database, which includes over a million hours of video data, to build a minimum of three simulator testbeds representative of urban interchanges under varying conditions (e.g. daytime, nighttime, angle crash with left-turning vehicle onto a freeway).

The team is documenting its approach to generating the datasets so future researchers can apply those methods to create RADs for other facility types, such as work zones, alternative intersections, or pedestrian facilities.

RAD Tool for Multi-Level Applications

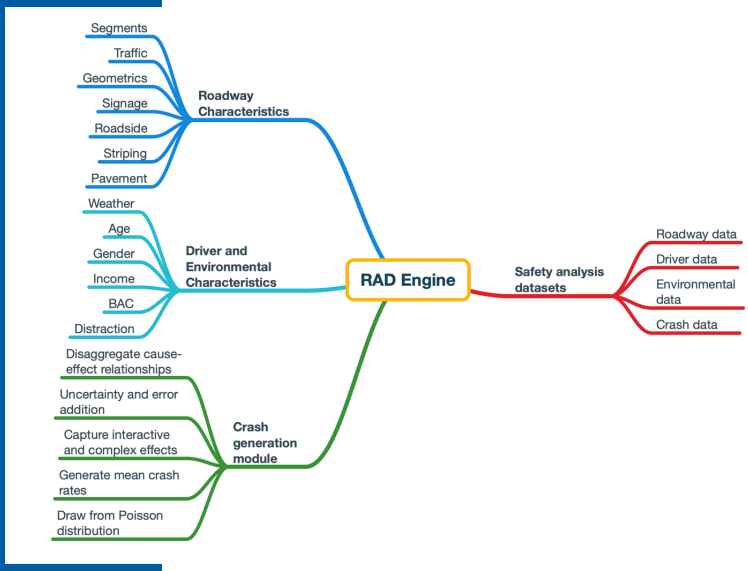
In a collaboration between the University of Connecticut and the University of Central Florida, researchers are building a RAD tool for generating datasets for all facility types using different input combinations. The tool will be customizable and capable of generating datasets at both the macroscopic and microscopic levels.

The macroscopic level is suited for single-step data models, such as those currently used in transportation safety analysis. The microscopic level can compare advanced research models and consider complex factors like human behavior as well as

U.S. Department of Transportation
Federal Highway Administration

Realistic Artificial Datasets:

Objective Evaluation of Data-Driven Safety Analysis Models



© *University of Missouri.* RAD tools are customizable and use variable inputs, such as driver behavior, environmental context, and roadway design to generate safety analysis datasets that help researchers better understand underlying safety relationships.

Photo credit, page 1: *Source: FHWA.* Mid-block crosswalks facilitate non-intersection crossings, providing safety benefits for pedestrians and drivers.

Publication No.
FHWA-HRT-20-047
HRTM-30/08-20(WEB)E

accommodate major changes in transportation, such as the inclusion of autonomous vehicles on roadways.

"I really feel that we are reaching the limits of what we can do with just road characteristics," explained John Ivan, principal investigator from the University of Connecticut. "Because that's not the major variation or causality for crashes occurring. The human factor is much bigger...so what this is going to do is allow approaches that start to get at driver behavior and driver characteristics."

To show proof of concept, the research team will generate two case studies: one focused on vehicle crashes on segments and one focused on vehicle and pedestrian crashes at intersections.

Informing Best Practices for DDSA

The development of RAD tools is key to the FHWA goal of expanding the adoption of DDSA as it will enable users to objectively select the best methods for their data. "In the future, when any transportation researcher comes up with a model, they will not have to scratch their head to determine whether their models are working well or not," said Yusuf Mohamedshah, data analyst with the FHWA Office of Safety Research and Development. "They can run their model across either of these RADs and look at the outputs to figure out how well their models are doing." RAD generation leverages leading-edge machine learning algorithms to develop robust tools suited for current safety analysis approaches as well as emerging and exploratory transportation models.

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 advantages 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 EARProgram, visit www.fhwa.dot.gov/advancedresearch. 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 these projects, contact Yusuf Mohamedshah, FHWA Office of Safety Research and Development at (202) 493-3464 (email: yusuf.mohamedshah@dot.gov).

[1] Federal Highway Administration. (2019). "Intersection Safety." (webpage) Available online: https://highways.dot.gov/research/research-programs/safety/intersection-safety, last accessed April 27, 2020.