Systemic Noteworthy Practice

U.S. Department of Transportation
Federal Highway Administration

Ohio's Machine Learning Profile Analysis for Roadway Departure and Pedestrian Crashes

Introduction

A systemic approach to safety considers both risk and crash history when making decisions on where to implement low-cost safety improvements. This approach relies heavily on data to identify locations for safety improvements that may not have been considered using other analysis methods. Analyzing the entire system using a defined set of criteria creates a more comprehensive approach that increases safety and decreases the potential for severe crashes. The systemic approach to safety has increasingly been used by agencies across the United States. This is one of a series of resources highlighting noteworthy applications of the systemic approach to safety.

Ohio's Systemic Approach

In January of 2022, the Ohio Department of Transportation (ODOT) released new funding for safety improvements at sites which meet minimum eligibility requirements for pedestrian crashes and roadway departure crashes. The ODOT used machine learning to develop the eligibility requirements in a time- and cost-effective way that allows for a sustainable funding process. From a systemic perspective, these eligibility requirements defined the focus facility type and risk factors for the focus crash types: pedestrian crashes and roadway departure crashes.

Integrating Machine Learning

Machine learning is an adaptive system that can analyze patterns in data to produce results without needing to be given explicit instructions. This adaptive system is being used in the traffic safety field to analyze large-scale data sets and produce modeled results. Machine learning can produce more accurate crash prediction models than other common models, can define abstract relationships between variables without human interference, can in some cases produce visuals and models that can be easily interpreted, and can be used without the need of specific software licensing. The limitations of machine learning are that little research into machine learning with regard to systemic safety has been done, in some cases there are not enough staff resources available to allow organizations to easily adopt it, and there is limited understanding and marketing of machine learning and its capabilities (Silva et al., 2020). Additionally, there is a high learning curve for both creation and interpretation of results, many methods exist, and the "best" methods are not yet known and are hard to identify for a given situation.

Data Collection and Analysis

ODOT created an integrated data set using the following Transportation Information Mapping System (TIMS) data sources:

- Statewide crash data for 2017 through 2021.
- The latest road inventory data as of August 10, 2022.
- The latest active transportation demand data as of August 10, 2022.
- Adjusted urban areas as of August 10, 2022.

ODOT supplemented these data with Justice40 initiative data from the United States Department of Transportation (USDOT), specifically the DOT Disadvantaged Final Layer per April 2022 and accessed August 26, 2022. The roadway data were filtered to include arterial and collector roadways, excluding access-controlled freeways, access-controlled expressways, and local roads and streets, because ODOT knew those roadways had safety concerns and the highest overrepresentation of severe pedestrian and roadway departure crashes. ODOT addressed missing data using a decision tree classifier algorithm in Python. ODOT then derived several network attributes (speed limit, number of lanes, lane width, and shoulder width) using bins.

The following crash metrics were developed for analysis:

- All Injury (KABC[1]) Crash Density
- All Injury (KABC) Crash Rate

2

Two method types were used to analyze the data: data inputs and tree architecture. Having two metrics and two methods for analyzing the data allowed for multiple levels of comparison which gave ODOT different outcomes to choose from when determining eligible locations for the funding. Finally, ODOT selected a classification and regression tree (CART) model for risk factor identification.

Outcomes

Analysts used machine learning to develop outcomes for many different possibilities based on the data and metrics. This provided different results and findings between variables that were not provided using other methods of selection. ODOT used the results provided from the machine learning to determine eligibility requirements. Table 1 is one example from the machine learning that ODOT used in the eligibility decision. This analysis indicates the critical segments for pedestrian risk screening are muiltilane segments with a pedestrian demand score of 4, which account for 37.9 percent of crashes despite only being 11.5 percent of mileage and 19.1 percent of MVMT.

When analyzing the crash data available from 2017 to 2021, ODOT learned that while the critical and high priority areas make up less than 28 percent of the arterial and collector network mileage, a large majority – 67 percent – of fatal and injury pedestrian crashes occur on these roadway types.

Summary

ODOT was able to show the capabilities machine learning has in determining risk factors, which are then applied for screening and prioritizing system components for systemic safety countermeasures. The results shown by machine learning helped identity key locations that would benefit from funding for safety improvements by looking at crash data and crash risk on the roadways. The results created were easily interpreted and developed in a time- and cost-efficient manner. This will improve ODOT's eligibility requirements and allow for a sustainable process when reviewing funding applications. In the future, ODOT plans to focus their systemic projects around locations that meet "critical" and "high" priorities.

Contact

For more information contact **Jeremy Thompson** (Jeremy.thompson@dot.ohio.gov), the Highway Safety Program Safety Engineer with ODOT.

https://highways.dot.gov/safety/data-analysis-tools/systemic

1 In reference to the KABCO injury severity scale - https://www.nhtsa.gov/mmucc-0.

Table 1 Pedestrian KABC Crashes per Mile per Year

Screening Profiles					Priority*	Avg. Crash Density		Network Milage		Mileage		MVMT Proportion		Crash Proportion	
Rank	Demand Score	Number of Lanes	Lane Width	Intersections per Mile		Profile	Group	Profile	Group	Profile	Group	Profile	Group	Profile	Group
1	4	Multilane	-	15+	Critical	0.53	0.45	650	1,620	4.6%	11.5%	6.1%	19.1%	20.1%	37.9%
2	4	Multilane	-	<15		0.40		970		6.9%		13.0%		17.8%	
3	4	Two-lane	-	15+	High	0.25	0.20	630	2,366	4.5%	16.9%	3.5%	23.7%	14.1%	28.9%
4	3	Multilane	-	-		0.19		1,125		8.0%		15.7%		9.8%	
5	4	Two-lane	Wide	<15		0.17		611		4.4%		4.5%		5.0%	
6	3	Two-lane	-	15+	Medium	0.09	0.06	469	3,951	3.3%	28.1%	2.5%	29.3%	5.1%	19.0%
7	4	Two-lane	Narrow	<15		0.09		412		2.9%		1.1%		4.6%	
8	1-2	Multilane	-	-		0.05		1,022		7.3%		13.8%		2.2%	
9	3	Two-lane	-	<15		0.05		2,047		14.6%		11.9%		7.1%	
10	1-2	Two-lane	-	-	Low	0.02	0.02	6,078	6,078	43.4%	43.4%	27.9%	27.9%	14/1%	14.1%

References

Jacobs. (2023, January 19). *Machine Learning Systemic Profile Machine Learning Systemic Profile Systemic Profile Analysis for Ohio Rodway Departure and Pedestrian Analysis for Ohio Rodway Departure and Pedestrian Crashes* (Project No. C6A15200).

Preston, H, Storm, R, Bennett, J. D., & Wemple, B. (2013, July). *Systemic Safety Project Selection Tool* (Report No. FHWA-SA-13-019). Department of Transportation Federal Highway Administration Office of Safety. https://safety.fhwa.dot.gov/systemic/fhwasa13019/

Silva, P. B., Andrade, M., & Ferreira, S. (2020, October 31). *Machine learning applied to Road Safety Modeling: A systematic literature review.* Journal of Traffic and Transportation Engineering (English Edition). Retrieved August 22, 2022, from https://www.sciencedirect.com/science/article/pii/S2095756420301410