SJSU SAN JOSÉ STATE UNIVERSITY

MINETA TRANSPORTATION INSTITUTE

Analytical Models for Traffic Congestion and Accident Analysis

Project 2102

November 2021

Hongrui Liu

Rahul Ramachandra Shetty

Traffic congestion and road accidents are major public challenges affecting the daily lives of people in different ways. In this study, the authors analyze U.S. accident data from February 2016 to December 2020 that consists of 4.2 million accident records with 49 variables and use machine learning algorithms to predict traffic accident severity. The goal of this work is to develop more accurate traffic predictive models by utilizing multiple environmental parameters. The proposed models will assist people in making smart real-time transportation decisions to improve mobility and reduce accidents.

Study Methods

The analysis on traffic congestion and road accidents is very complex as they are affected by many factors, such as road user demographics, time of the day/week, weather conditions, and other events that may change in real-time. In this study, we use statistical techniques and machine learning algorithms to process and train models, using the large amount of data offered by advanced information technologies, to obtain robust predictive models that can incorporate environmental parameters in real time. Machine learning algorithms used in the study include logistics regression, decision tree classifier, random forest classifier, and XG-Boost.

Findings

Machine learning algorithms are more flexible in incorporating the impact of multiple environmental parameters and thus can train predictive models with higher level of accuracy. By comparing predictive models trained from four machine learning algorithms that include multivariable logistic regression, decision tree, random forest (RF) and extreme gradient boosting (XG-boost), it was found that all methods have similar accuracy using the original imbalanced data. While the decision tree, RF and XG-boost algorithms exhibited greater robustness and accuracy with balanced data. The RF model handles noise, multi-class variables, and high variance well, while the boosting technique learns from the past errors to improve prediction accuracy. In the study of the trade-off between the computation effort and model accuracy, we simply removed 10 out of the 24 features used in the previous analysis. The accuracy score didn't change significantly for any of the models.

Machine learning models together with the real-time visibility of the environmental conditions offered by advanced information technologies can help build a smart transportation system to guide people making decisions to minimize their time spent on the road.

Policy/Practice Recommendations

The obtained predictive models, together with the real time visibility of the environmental conditions offered by advanced information technologies, can help build a smart transportation system to guide people making decisions to minimize their time spent on the road. More studies need to be conducted to investigate the computational constraints to the implementation of such a system.

About the Authors

Dr. Hongrui Liu is an assistant professor of Industrial and Systems Engineering at San Jose State University.

Rahul Ramachandra Shetty graduated from San Jose State University in Industrial and System Engineering with a specialization in Supply Chain Management. He is currently working as a Supply Chain Business Manager at Lam Research Corporation.

To Learn More

For more details about the study, download the full report at **transweb.sjsu.edu/research/2102**

MTI is a University Transportation Center sponsored by the U.S. Department of Transportation's Office of the Assistant Secretary for Research and Technology and by Caltrans. The Institute is located within San José State University's Lucas Graduate School of Business.