Potential Erroneous Degradation of High Occupancy Vehicle (HOV) Facilities

Determine erroneous degradation magnitude of HOV facilities due to general-purpose mainline detectors misconfigured as HOV detectors.

WHAT IS THE NEED?

According to the Federal Highway Administration (FHWA) and Caltrans guidelines, reports on the degradation status of high-occupancy vehicle (HOV) lanes is required annually. In California, these reports depend on data gathered through the Performance Measurement System (PeMS). According to the April 2018 Report to the Legislature, 65 percent (864 out of 1331 total lane-miles) were degraded in the first half of 2016.

However, recent studies at University of California at Berkeley (UCB) show that general-purpose mainline detectors may be commonly misconfigured as HOV detectors. In the facilities studied, this misconfiguration corresponded to about 8% to 9% of HOV detectors, or about one wrong configuration every 5-6 HOV lane-miles. It is possible that this level of erroneous data may potentially cause HOV facilities to be wrongly evaluated as degraded.

In 2016, one study investigated 32-lane miles of HOV lanes along I-210 through Pasadena. Along that corridor, 5 HOV detectors were found to have been wrongly configured. Two years after the configuration was corrected, there was an Internet Protocol conversion project along the same facilities. Immediately following the project, 6 HOV detectors were found to have been wrongly configured by the contractor. In most cases, HOV detectors are misclassified as general-purpose mainline detectors.

Erroneous degradation of HOV facilities skews performance metrics and investment priorities. In addition, it limits Caltrans’ ability to fashion policy to address multimodal transportation needs. The proposed research will assess the extent of this problem.
**WHAT ARE WE DOING?**

The research team at UCB will perform research that explores the application of data science to improve the accuracy of performance measures, and therefore directly improve the quality of management decisions. The main goal is to develop automated means to identify configuration errors of HOV lanes, leveraging the latest research in data science.

In the process of this research, a survey of data-mining methods will be assembled to cluster similar patterns and for identifying anomalies in HOV data. These methods will be compared for their accuracy in identifying HOV misconfiguration.

The best method will be selected to determine the HOV sensors that are likely misconfigured over a larger region, such as one Caltrans district. These HOV sensors will be compared with their mainline neighbors to see if they are wrongly degraded. Finally, a report will summarize findings and assess the extent and seriousness of the problem.

**WHAT IS OUR GOAL?**

It is currently unknown how widespread or serious this HOV misclassification is. However, there appears to be no protocol for Caltrans maintenance to check whether detectors are connected to the correct physical terminals in the controller. In addition, there are no data analysis tools to check flow patterns and automatically flag this kind of misconfiguration error.

Therefore, the goal of this project is to create a method to detect likely HOV misconfiguration errors and to reveal the magnitude of the problem.

**WHAT IS THE BENEFIT?**

Automated detection of misconfiguration errors will provide a more efficient and timely way of finding problems with Performance Measurement System (PeMS) data and field elements. In addition, it will improve the efficiency and reliability of every future project that uses HOV data from PeMS for any type of analysis, because the initial step of checking that the data are trustworthy will have been completed.

Honest and accurate assessment of HOV-lane facilities performance may show that the reported degradation is greater than actual degradation. This may obviate the need to bring a wrongly degraded facility into compliance under FHWA guidelines thus enabling Caltrans to provide more flexible operating policies for HOV-lane facilities.

Increasing the quality of information available to Caltrans management will directly improve investment decisions related to the operations of HOV-lane facilities. Better decisions will translate to better environmental outcomes and save money by preventing unnecessary projects.

**WHAT IS THE PROGRESS TO DATE?**

Researchers delivered Technical Memo: Potential Erroneous Degradation of High Occupancy Vehicle (HOV) Facilities

Researchers worked on Performance Methods. Thus far, the machine learning training and testing script tested eight different machine learning methods, four supervised classification methods and four unsupervised anomaly detection methods:

- Supervised: K-Nearest Neighbors (KNN), Logistics Regression, Decision Tree, Random Forest;

The contents of this document reflect the views of the authors, who are responsible for the facts and accuracy of the data presented herein. The contents do not necessarily reflect the official views or policies of the California Department of Transportation, the State of California, or the Federal Highway Administration. This document does not constitute a standard, specification, or regulation. No part of this publication should be construed as an endorsement for a commercial product, manufacturer, contractor, or consultant. Any trade names or photos of commercial products appearing in this document are for clarity only.

© Copyright 2021 California Department of Transportation
ALL RIGHTS RESERVED
• Unsupervised: Isolation Forest, Local Outlier Factor (LOF) (Density-based detection), One-class Support Vector Machine (SVM), Robust Covariance (Distance-based detection)

Each method was tested using the I-210 data with simulated ground truth (intentionally misconfigured sensors), evaluating each method's prediction accuracy. This was done using the seven days of daily traffic data in two ways. First as seven separate daily tests and second as a combined dataset of seven consecutive days of traffic data.

Overall, performance results reveal that classification methods yielded the highest accuracy (at or near 100%) but may be prone to overfitting. Considering that accuracy for a single day of data achieved >80% accuracy, the performance difference between a single day of data versus seven consecutive days was modest (5-10%).

The researchers also worked on Magnitude of erroneous degradation. Within the preprocessing, training, and prediction scripts developed in previous task have been extended to now include data for all of Caltrans District 7. The scripts can make predictions for potentially erroneous sensors in District 7.