Mitigating Safety Risks & Environmental Impacts Questions and Answers

Presentation #1: Development of a Maintenance Prioritization Assessment and Safety Tool

Q1: What type of data or historical information can be used in planning maintenance functions?

A: This was discussed in the presentation consisting of data from past maintenance activities, traffic volume data, past collision and injury data, lane closure data, and experience of the actual crew and others who performed such maintenance activities.

Q2: Does the number of people in a crew affect the level of difficulty to complete a typical maintenance task performed by your group?

A: This is typically the case although there can be exceptions.

Q3: What are the benefits of a systematic method for prioritize maintenance operations?

A: The benefits are to improve safety and efficiency of the operations and the quality of the outcome.

Q4: Was MAZEEP considered as a factor to lower risk?

A: Other studies that we had done in the past have shown the benefits of COZEEP to lower risks and it is recommended for many of the operations.

Q5: What is the matrices used to evaluate the machine learning model (precision and recall, AUC)? What is the criteria for the matrices for the ML model to be considered satisfactory?

A: Model performance comparison for feature selection between logistic regression and extreme gradient boosting (Xgboost) is done based on the AUC score. The same score has been used to tune xgboost's hyperparameters in a cross-validation framework. This is described in detail in our ***publication [1], Section 4.3***.

Generally, an AUC score of more than 0.8 is acceptable. However, for highly imbalanced data sets such as this case, prediction probabilities require calibration before calculating the AUC score which are described in more details in the same ***paper [1], Section 4.4.*** In feature selection, both logistic regression and xgboost achieved an AUC score of 0.9.

Q6: Has the Risk Evaluation model been fully developed and ready to use?

A: Yes, but the data that needs to be input into the parameters needs be extracted in each case necessitating the need for a decision support tool that can assist the user in getting the right values for the needed parameters.

Q7: In the logistics model, did you normalize the input parameters in the feature selection procedure? I noticed the lane closure is a categorical parameter, and others are continuous. How did you do the normalization?

A: Feature values are normalized to a range of [0-1] for continuous features in feature selection procedure. This is described in detail in ***Section 4.2 of [1]***. Continuous features are centered and then scaled to a range between [0-1].

Q8: How many data points are used in developing the logistics model?

A: Both feature selection methods (logistic and xgboost) are trained and tested over similar training and testing sets. There are more than 2 million data points (roadside work zones for different dates and activities) of which only about 2% have a work zone collision within the predefined tolerance distance. The train to test split for this data set has been set at 70% to 30%.

Q9: Besides the logistics model, have you tried other models such as SVM or decision tree? Why use logistic regression?

A: Although we have used logistic regression for feature selection, we did not rely on its result entirely. In fact, the top 4 most important features selected for collision risk index development have been chosen according to the results of the xgboost model because of its better performance. The process is described in detail in ***Chapter 6 of our report [2] and Section 5 of paper [1]***.

Once the top most important features are identified, we used them in a logistic regression function to develop a collision risk index. This is done because of logistic regression's easier interpretation where a feature's coefficient and sign can be translated to how that feature affects the outcome which in this case is the probability of a work zone collision.

Q10: Will the support tool in development consume IMMS data through a pipeline connection, or will this tool need data updates passed through the model (manual)?

A: We have tried to automate the data preprocessing part as automated as possible but there were technical difficulties at the time preventing us from developing a totally automated pipeline. Some of the roadblocks were:

1. We could not identify a portal, gateway, or systematic way to download large chunks of data from the IMMS, PeMS, and SWITRS data sets. In fact, even manual download of data had to be broken down to smaller chunks because of size limitations put in place within those systems.
2. Traffic volume data and Highway element marker data were given to us as manual excel files. We did not have a way of programmatically accessing these data sets as well.

Q11: Can you describe any work done towards keeping workers safe during animal carcass removal? Or is that activity too random and unplanned?

A: Such an activity was not included in the initial list of maintenance functions that we were provided to consider. However, this can be considered in the future and be added with proper data extraction.

Presentation #2: Roadway Crossings for Sensitive Amphibians and Reptiles

Q1: Can you explain more why offsite mitigation can be more expensive than onsite mitigation?
A: Offsite mitigation is primarily done using a mitigation bank in which Caltrans buys credit to offset the impact to habitats and/or water bodies at 3:1 ration. This means, we are required to pay as high as 3 acres of the credit for each acre of impact. This is the main reason that makes offsite mitigation expensive. It does not mean this is always the case as there are cases when onsite mitigation can be more expensive than offsite mitigation especially when Caltrans is required to monitor the rehabilitated habitat once construction is complete.

Q2: How far apart are the turn arounds?
A: It has not yet been defined as this is a new project and the researchers are working on the details of the experimental design.

Q3: Has the design development identified maintenance criteria for crossing, or how well is the crossing able to self-maintain?
A: In any transportation related new innovations, the challenge is to make sure we have a feasible approach for maintenance. We knew it would be a challenge and for this reason we are including Caltrans staff from maintenance to help us develop the best management practice during the research process.

Q4: Would the multiple turn arounds create increased points for the species to be preyed upon?
A: That is generally true that wildlife crossings can expose species to be preyed upon. There is evidence that suggests predators wait on the edge of crossing structures to prey on species gathered to use them. We know that structures with no fences or turnarounds have relatively less usage rate. The risk may be there, but we are trying to alleviate the problem of critters being trampled in thousands by moving vehicles.

Q5: Will the maintenance feasibility of the proposed amphibian structures be considered?
A: As indicated on Q3, a Caltrans maintenance staff is in the research panel to help us advise on this issue.

References:

[1] Nasrollahzadeh, Amir A., Ardalan R. Sofi, and Bahram Ravani. "Identifying factors associated with roadside work zone collisions using machine learning techniques." Accident Analysis & Prevention 158 (2021): 106203.

[2] Nasrollahzadeh, Amir, and Bahram Ravani. "Research to Develop Performance Measures for Maintenance of Roadside Features." California Department of Transportation (2021): CA21-3289

###