SMA's Crystal Ball

Ryan Rivas
Department of Computer Science and Engineering
University of California, Riverside

Kamili Shaw, PhD
Safety and Mission Assurance Directorate
NASA Stennis Space Center

National Aeronautics and Space Administration

Safety incidents are costly for both NASA and contractors. Information on these incidents, as well as other safety data such as close call reports and safety inspection findings, is collected and stored within NASA data systems. The leading safety indicators present in these datasets provide the potential to predict safety incidents before they happen, allowing NASA personnel to direct safety resources to high-risk areas and ultimately avoid the costs incurred by these incidents. To demonstrate the feasibility of predicting future incidents at NASA centers, we combined data from some of these systems. We then used the combined dataset to train several machine learning algorithms and evaluate their performance.

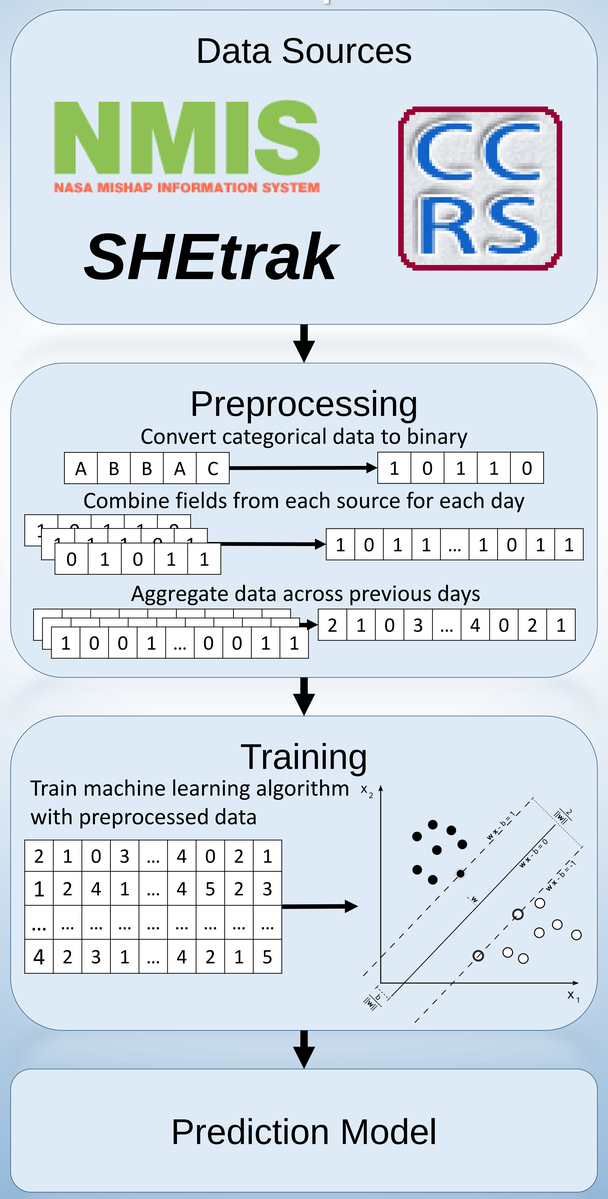
Introduction

- Part of the mission of the Safety and Mission Assurance (SMA) Directorate is to prevent safety incidents, including injuries, property damage and mission failures
- NASA SSC data systems contain large amounts of valuable information on safety and quality, e.g.:
  - Close Call Reporting System (CCRS)
  - NASA Mishap Information System (NMIS)
  - Safety, Health and Environmental Tracking System (SHEtrak)
- The goal of this project is to use this data to predict safety incidents by training a machine learning algorithm to recognize data that represents a high probability of an incident occurring:

Objectives

- Combine data from NASA SSC safety and quality data systems (CCRS, NMIS and SHEtrak)
  - Consists largely of categorical data that must be represented using "one-hot encoding" (one binary field for each possible value)
  - Other data types: dates, Boolean (i.e. "yes/no"), numeric, text (text data is not yet implemented in this phase of the project)
- Identify candidate algorithms to provide the basis for a predictive capability for safety incidents, e.g. decision trees, support vector machines, neural networks, etc.
- Conduct experimental evaluation of machine learning methods using safety and quality data
- Explore means to improve accuracy of promising safety incident prediction models

Data Pipeline



Expected Outcomes

- Implementation of the project's data model using three sources (CCRS, NMIS and SHEtrak)
- Demonstration of the feasibility of building a prediction model from the data (see Data Pipeline, left)
- Performance metrics of machine learning algorithms under cross-validation with the dataset, e.g.:
  - Accuracy – Percentage of correctly classified days
  - Precision – Ratio of true positives vs. all positives
  - Recall – Ratio of true positives vs. all days with an incident
  - F_1 score – $2\frac{precision \times recall}{precision + recall}$
  - ROC – A curve representing the true positive rate vs. false positive rate under various thresholds
- Proposal of directions for further improvement of the prediction model

Summary

- NASA holds a wealth of safety data: incident investigations, close call reports, inspections, etc.
- This data can be used to train a machine learning algorithm
- A prediction model built from this data may be able to predict safety incidents at NASA centers
- Future work:
  - Text data – Text can also be represented in a prediction model in many ways, e.g. bag-of-words, Doc2Vec, text CNN
  - Other NASA data sources will contribute to the prediction model:
    - Audit Tracking and Information System (ATIS)
    - Design and Data Management System (DDMS)
    - Integrated Risk Management Tool (IRMA)
    - Maximo (asset management)

Acknowledgements: NASA John C. Stennis Space Center Safety and Mission Assurance Directorate.
References: Shaw K. SMA's Crystal Ball: Project Management Plan. 2017 Dec 22; SPLN-1710-0001. National Aeronautics and Space Administration.