Reliability Engineering of Autonomous Systems incorporating Machine Learning

Christian Ellis, Lance Fiondella
University of Massachusetts, Dartmouth

SYSTEMS ENGINEERING RESEARCH CENTER

Introduction

Artificial intelligence and machine learning have attracted significant interest as enablers of autonomous systems.

However, autonomous systems are susceptible to a variety of failures as well as adversarial attacks, suggesting the need for more formal reliability and resilience engineering methods.

In the absence of such methods, it will be difficult to ensure these systems will be accepted as trusted partners by humans.

This research explicitly documents how machine learning test and evaluation methods can be incorporated into existing processes by mapping traditional concepts from reliability engineering in the following:

(i) reliability growth modeling and reliability engineering
(ii) fault tolerance
(iii) software testing
(iv) failure modes and effects criticality analysis (FMECA)

Objectives

This research seeks to bridge the gap between traditional and emerging methods to support the engineering of autonomous systems incorporating machine learning.

The proposed approach will provide organizations with additional structure to comprehend and allocate their risk mitigation efforts.

Furthermore, we seek to:

(i) Provide the Test and Evaluation Community with a familiar framework in which to assess autonomous systems
(ii) Facilitate effective communication between diverse stakeholders such as system engineers, advanced algorithm designers, testers, and leadership

Methodology

Machine learning is an enabler of autonomy that resides in software. Therefore, test and evaluation may be first be regarded as an extension of the software reliability field requiring the integration of learning enabled components into software and system architectures.

Performance of a machine learning component depends on how well the training data represents its operational environment.

We trained a neural network to represent a classifier, which can be used in autonomous systems to inform decision making.

Data & Analysis

Reliability Growth Modeling

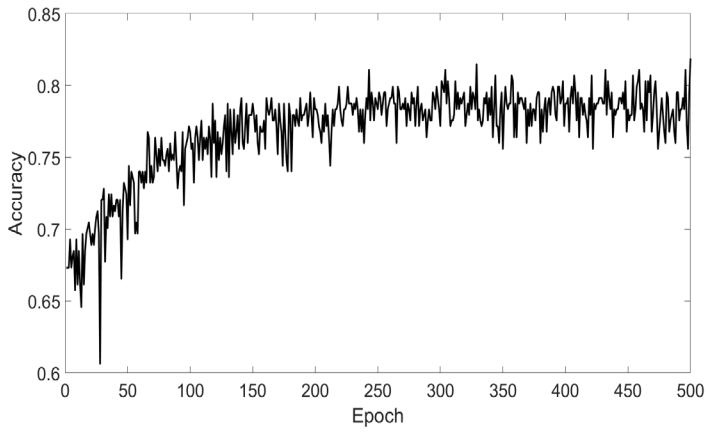

Figure 1: Accuracy as a function of iterations (epochs)

Since the accuracy of a machine learning algorithm tends to improve with more training data, time and reliability are replaced with training iterations (epochs) and accuracy on the x- and y-axes respectively.

Data & Analysis (continued)

Reliability Engineering

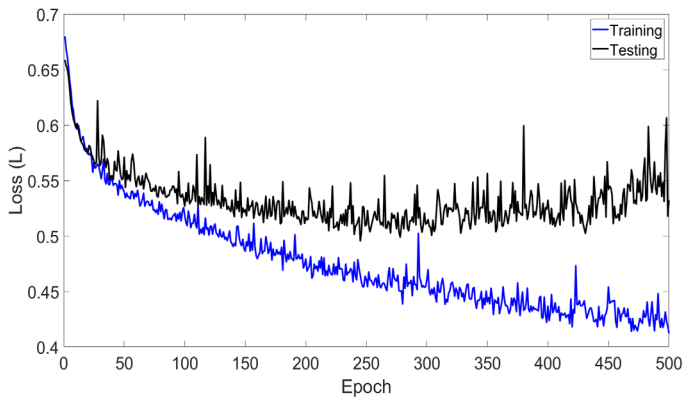

Figure 2: Loss as a function of iterations (epochs)

On average, the loss decreases on the training data, but decreases and then increases on the testing data, indicating that model overfitting has occurred because the algorithm performs well on the training data, but worse on the testing data. Regularization can be applied to deter overfitting.

Software Testing

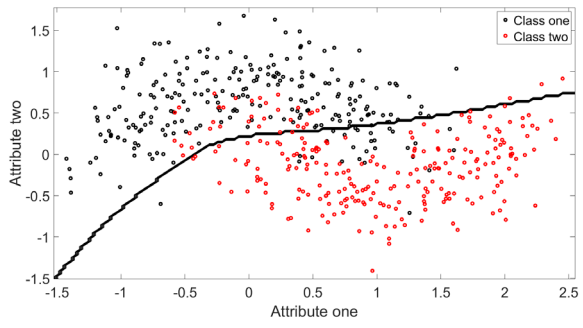

Figure 3: Input domain view of testing machine learning algorithm

Black dots above the black line and red dots below the line represent correct classifications. The two input attributes are continuous valued, making the input space infinite. Several misclassifications of class one and two are visible.

Failure Modes and Effects Criticality Analysis (FMECA)

		Actual Values	
		Pedestrian	No Pedestrian
Predicted Values	Pedestrian	TP	FP (Minor)
	No Pedestrian	FN (Catastrophic)	TN

Table 1: Confusion matrix for pedestrian identification

In a cost matrix C based on Table 1, a false negative (c_{21}) is significantly more costly than a false positive (c_{12}). Thus, the cost of misclassification need not be symmetric and traditional methods from binary systems reliability theory are inadequate.

To address failure modes and their severity, cost sensitive learning trains a classifier to minimize loss. For example,

$$C(Pedestrian) = c_{21} \times \Pr\{FN\} + c_{12} \times \Pr\{FP\}$$

Future Research

Future research will further elaborate the connections between reliability engineering and machine learning methods.

The relationship between adversarial machine learning and failure modes, effects and criticality analysis is an emerging area, where more thorough understanding will assist in the engineering of resilient autonomous systems.

Application of techniques from machine learning to support resilience systems engineering will also be explored.

Contacts

Lance Fiondella Ph.D.
lfiondella@umassd.edu

Christian Ellis
cellis3@umassd.edu

Acknowledgement

Christian Ellis is supported by an Army Research Laboratory Research Associateship Program Fellowship.