QUANTITATIVE RELIABILITY AND RESILIENCE ASSESSMENT OF A MACHINE LEARNING ALGORITHM

Karen da Mata¹, Zakaria Faddi¹, Priscila Silva¹, Vidhyashree Nagaraju², Susmita Ghosh³, Gokhan Kul¹, and Lance Fiondella¹

¹University of Massachusetts Dartmouth, MA, USA, ²Stonehill College, MA, USA, ³Jadavpur University, Kolkata, IN

April 2024
MOTIVATION

• Machine learning (ML) applications face dynamically changing and actively hostile conditions that lead to system failures and degraded performance
• Systems incorporating ML must be reliable and resilient, especially in safety-critical domains
• Many studies propose techniques to improve the robustness of ML algorithms, but few consider quantitative techniques to assess the reliability and resilience of ML models under likely stresses and attack scenarios.

CONTRIBUTION

• Demonstrate how to collect relevant data during training and testing of machine learning models suitable to apply
  • Software reliability models without covariates
  • Software reliability models with covariates
  • Resilience models
• Enable quantitative assessment of ML reliability and resilience during testing and field operations monitoring to ensure systems operate dependably under critical conditions.
MACHINE LEARNING TECHNIQUES

Convolutional Neural Networks
- Image classification
- Object recognition

Generative Adversarial Attacks
- Fast Gradient Sign Method (FGSM)
- Projected Gradient Descent (PGD)

Defense Measure
- Adaptive Adversarial Training

\[
\min_{\theta} \left[ \max_{\delta \in \Delta} [\mathcal{L}(x + \delta, y, \theta)] \right]
\]
RELIABILITY AND RESILIENCE MODELING

RELIABILITY

Non-homogeneous Poisson Process Software Reliability Growth Models (NHPP SRGM) estimate the number of defects remaining at any given time and the rate at which defects will be detected and removed.

- Models without covariates
  - Goel-Okumoto (GO)
  - Weibull (Wei)
  - Delayed S-shaped (DSS)
- Models with covariates
  - Geometric (GM)
  - Discrete Weibull of order two (DW2)
  - Type III discrete Weibull (DW3)
  - “S” distribution (S)
  - Truncated logistic (TL)
  - Increasing Failure Rate Salvia and Bollinger (IFR SB)
  - Increasing Failure Rate Generalized Salvia and Bollinger (IFRGSB)

RESILIENCE

A discrete resilience curve incorporating covariates is described as

\[ P(i) = P(i-1) + \Delta P(i) \]

- Regression Models
  - Multiple Linear Regression (MLR)
  - Multiple Linear Regression with Interaction (MLRI)
  - Polynomial Regression (PR)
DATA COLLECTION
Factors collected before and after Step 5.

<table>
<thead>
<tr>
<th>Before</th>
<th>After</th>
</tr>
</thead>
<tbody>
<tr>
<td>Factor Name</td>
<td>Acr.</td>
</tr>
<tr>
<td>Failure Time</td>
<td>FT</td>
</tr>
<tr>
<td>Failure Count</td>
<td>FC</td>
</tr>
<tr>
<td>Epsilon</td>
<td>ε</td>
</tr>
<tr>
<td>FGSM percentage</td>
<td>FGSM%</td>
</tr>
<tr>
<td>PGD percentage</td>
<td>PGD%</td>
</tr>
<tr>
<td>Test Accuracy</td>
<td>TeA</td>
</tr>
<tr>
<td>Test Loss</td>
<td>TeL</td>
</tr>
<tr>
<td>F1-Score</td>
<td>F1</td>
</tr>
</tbody>
</table>

Epsilon ranges considered:
- 0.1-0.4
- 0.2-0.5
- 0.3-0.6
- 0.4-0.7

Process for ML data collection to assess reliability and resilience
- ML Model: CNN
- Generative Adversarial Models: FGSM and PGD
- Data set: CIFAR-10
RESULTS ANALYSIS – RELIABILITY

Best-fitting models without covariates

NHPP SRGMs incorporating covariates tracked and predicted more accurately than models without covariates.

Best-fitting models with covariates
RESULTS ANALYSIS – RELIABILITY

Frequency of covariate inclusion in the NHPP SRGM with covariates

Number of failures in each interval of failure count data sets and zoomed-in view of iterations 4 to 33 as well as predictions made by models of best fit with covariates
RESULTS ANALYSIS – RELIABILITY

Reliability growth curves of NHPP SRGM models without covariates

Reliability is higher for higher $\varepsilon$ ranges
RESULTS ANALYSIS - RESILIENCE

Best-fitting resilience models

Larger $\varepsilon$ ranges exhibited smoother performance curves
CONCLUSION

Summary
• Demonstrated the applicability of quantitative reliability and resilience assessment methods of ML models equipped with defense measures and subject to adversarial attacks.

Results
• Software reliability growth models incorporating covariates more accurately track and predict the number of defects detected than models without covariates.
• Resilience models considering negative and positive factors characterizing the deterioration and recovery of a system are also able to precisely track and predict the resilience of defensive measures for ML subject to specific attacks.

Future research
• Explore the application of these models to cyber-physical systems.
• Incorporate these reliability and resilience models into maintenance models.
This research was supported by the National Science Foundation under Grant Number 1749635 and the Homeland Security Community of Best Practices (HS CoBP) through the U.S. Department of the Air Force under contract FA8075-18-D-0002/FA8075-21-F-0074. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation, U.S. Department of Homeland Security, or the U.S. Department of the Air Force.