SAND2021-9483C

Sandia National Laboratories

Exceptional service in the national interest

Impact of Safeguards Measurement Errors on Deep Neural Networks

Nathan Shoman*,[1], Tom Burr[2]

*nshoman@sandia.gov, [1]SNL, [2]LANL

July 28, 2021

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

Motivation: Previous work has shown machine learning is sensitive to measurement error

- ML[1] for MC&A has been shown to rapidly degrade in presence of measurement error
  - Considers bulk nuclear facilities (i.e. subject to MC&A rather than simple item accounting)
- Traditional safeguards also degrade, but to a lesser degree
- Why does this behavior occur?
- Can this reduction in performance be mitigated?

[1] Only unsupervised approaches examined

Cornerstone of traditional MC&A: The material balance

The material balance (MB), shown below and sometimes called Material Unaccounted For (MUF) or Inventory Difference (ID), is a simple and straightforward way to perform "audits" of nuclear material at facilities.

Generic material balance

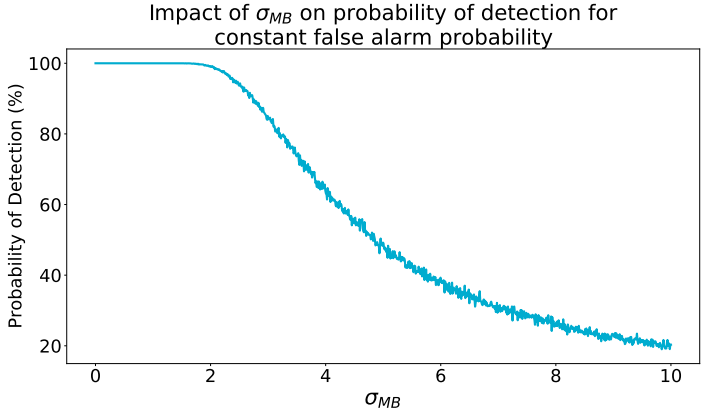
$$\mathrm{MB}_t = \left(\sum_{i=1}^{n_\mathrm{I}} I_{i,t-1} + \sum_{i=1}^{n_\mathrm{in}} Tin_{i,t} - \sum_{i=1}^{n_\mathrm{out}} Tout_{i,t} \right) - \sum_{i=1}^{n_\mathrm{I}} I_{i,t} \tag{1}$$

Under normal conditions $\mathrm{MB}_t = 0$, but measurement error causes a non-zero MB as follows below.

Material balances are approximately normally distributed

$$\mathrm{MB}_t \sim \mathcal{N}(\mu_t, \sigma^2_{\mathrm{MB}}) \tag{2}$$

Increases in MB uncertainty (i.e. σ_{MB}) degrade detection probabilities



Impact of measurement error on detection probability

$$\lim_{\sigma_{\text{MB}} \to \infty} PD(\mathcal{N}(\mu_t \to \mu_t^*, \sigma_{\text{MB}}^2)) = \text{FAP} \quad (3)$$

Where's the error come from?

Imperfect measurements result in uncertainty. Safeguards measurements often has a multiplicative error structure, shown below.

Safeguards error model

$$
\begin{aligned}
&M_{i,t} = G_{i,t}(1 + S_i + R_{i,t}) \\
&\text{where} \\
&S_i \sim \mathcal{N}(0, \delta_S^2) \\
&R_{i,t} \sim \mathcal{N}(0, \delta_R^2)
\end{aligned}
\tag{4}
$$

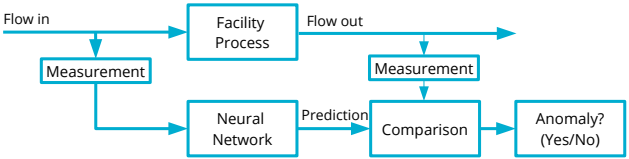
- S_i: Short-term systematic (i.e. epistemic) error from measurement conditions or calibrations
- $R_{i,t}$: Random (i.e. aleatory) error that is unpredictable but reducible through repeated conditions
- $M_{i,t}$: Measured quantity
- $G_{i,t}$: True (but unknown) quantity

Key points about traditional MC&A

- Some techniques exist to reduce impact of measurement errors
  - SITMUF which explicitly accounts for error
  - GEMUF
- Ultimately at mercy of measurement error
- Operate on single set of data
- Learns measurement error structure over time

Alternative machine learning approach

Goal: Learn $f(x, \theta)$ that can approximate a facility process using machine learning (neural network).



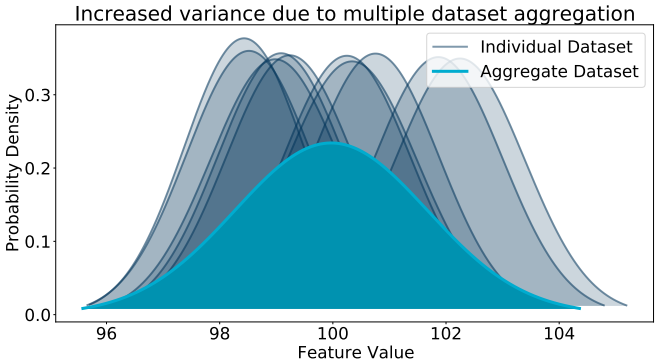
Example

- $f(x, \theta)$: Function to be approximated (facility process) with input measurement x and learned weights θ
- y: Observed measurement of process output
- $\hat{y} - y$: Objective function to be minimized, difference between prediction and observed value

Problem description

Machine learning algorithms are data hungry which comes with a cost

It is reasonable to assume any training dataset would, in practice, be comprised of multiple measurement campaigns with different biases. Aggregating these datasets results in a larger training dataset variance due to the systematic error.



Intuition: How do errors impact training of machine learning algorithms?

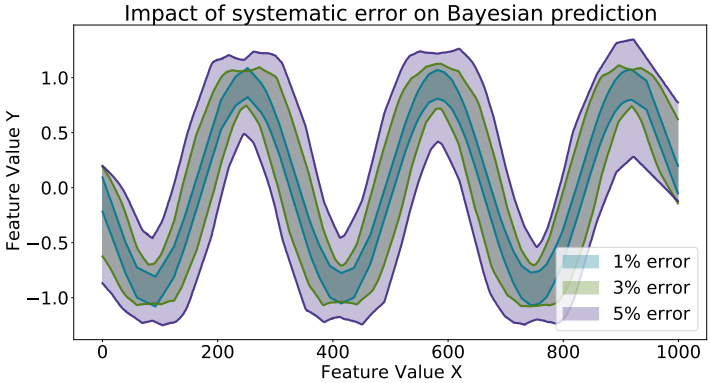
Thought experiment

Training a Bayesian neural network (weights are distributions not vectors or scalars) to learn a sine wave with increasing levels of error.

Example

$$
\begin{aligned}
x_{\text{true}} &\in [-3\pi, 3\pi] \\
y_{\text{true}} &= \sin(x_{\text{true}}) + 10 \\
x_{\text{observed},t} &= x_{\text{true}}(1 + R_t + S) \\
y_{\text{observed},t} &= y_{\text{true}}(1 + R_t + S)
\end{aligned}
\tag{5}
$$

Increased systematic error lowers the confidence of machine learning algorithms

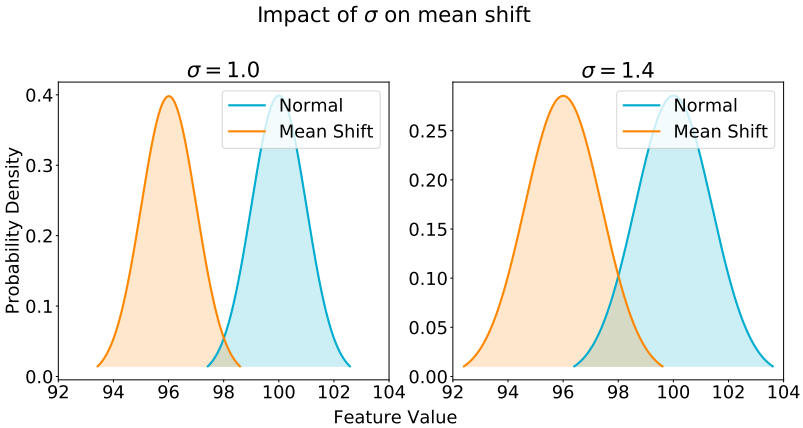


Predictions made within uncertainty bounds cannot be used for reliable detection of anomalous behavior.


Theoretical guarantee for poorer ML results compared to traditional approaches

Thinking statistically: comparing normal and loss distributions

Material losses can be thought of as a shift in an observed distribution of measurements. Larger variances lead to a more difficult mean shift detection problem.



Thinking statistically: machine learning training

In this work, the machine learning training objective is defined to be the mean squared error (MSE), that is, the algorithm is trained to provide more accurate predictions of a process output. It can be shown that the MSE objective is essentially negative log-likelihood (i.e. cross-entropy) between the empirical distribution and a Gaussian model (i.e. the learned distribution, assumed to be normal).

MSE and KL-divergence

$$\underset{\theta}{\operatorname{argmin}} \frac{1}{N} \sum_{i=1}^{N} (y_i - x_i\theta)^2 \equiv \underset{\theta}{\operatorname{argmin}} -\mathbb{E}_{x \sim \hat{p}_{\text{data}}} [\log p_{\text{model}}(x)] \tag{6}$$

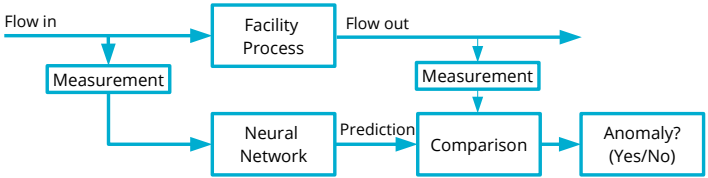
Key takeaway

During training, the proposed ML approach attempts to have a distribution $p(y|x, \theta)$ that closely matches the training distribution

Machine learning is disadvantaged due to training dataset requirements

- Machine learning algorithms usually require large datasets
  - Reasonable to assume safeguards datasets will be from multiple measurement campaigns
  - Aggregation of multiple measurement campaigns leads to a larger training dataset variance (in both x and y)
  - Larger variance leads to poorer loss detection
- Statistical methods for safeguards operate on a single dataset at a time (i.e. single set of systematic errors)
  - Consequently, variance in dataset is smaller
- Traditional safeguards also mitigate error through some transformations

Key takeaway



Finding

As a consequence of aggregating multiple measurement campaigns, machine learning methods are guaranteed[a] to have poorer performance than traditional methods. **The systematic error requires treatment independent of traditional pre-processing techniques like scaling**.

[a] For the proposed architecture (i.e. unsupervised regression)

Potential future work

- Reduce required training dataset size
  - Few-shot learning concepts (restriction of hypothesis search space, dataset augmentation, etc)
- Experimental mitigation strategies for systematic error
- Use of models that are not deeply parameterized (i.e. classical statistics and ML approaches)

Acknowledgments

Funding for this work was provided by the U.S. Department of Energy National Nuclear Security Administration's Office of Nonproliferation & International Security. The authors would also like to thank Benjamin Cipiti, Philip Honnold, and Richard Fields (Sandia National Laboratories) for their contributions to this work.

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.