Uncertainty Aware Machine Learning Models for Particle Physics Applications

Kishansingh Rajput
Diana McSpadden, Malachi Schram
Thomas Jefferson National Accelerator Facility

Karthik Somayaji, Peng Li
University of California, Santa Barbara

Willem Blokland
Oak Ridge National Laboratory

Zhiwen Zhao
Duke University
Uncertainty Quantification in Machine Learning

- Deterministic transformation functions

\[
\text{Input(s)} \xrightarrow{\text{ML model}} \text{Output(s)}
\]

- Decision making based on predictions from ML models
  - Uncertainty Quantification is required to make an informed decision

Uncertainty Types: Aleatoric vs Epistemic uncertainties

- Aleatoric $\Rightarrow$ Data uncertainties
- Epistemic $\Rightarrow$ Model or Out of training distribution uncertainty (OOD)

Gaussian Process (GP) provide robust uncertainty quantification
Outline

- Traditional application of Gaussian Process (GP):
  - Uncertainty aware experiment control at GlueX with GP
- Scaling limitations of GP and UQ for Deep Neural Networks (DNN):
  - Uncertainty aware anomaly prediction at SNS
  - Uncertainty aware particle identification for SoLID
  - Uncertainty aware surrogate model for FNAL booster
- Conclusion
Gaussian Process (GP) for Calibration of the Central Drift Chamber (CDC)

- Requires two calibrations: gain and drift time-to-distance
  - Gain Correction Factor (GCF) [variation +/- 15%]
- Has one control: operating voltage
- CDC is gas filled, sensitive to
  - Atmospheric pressure
  - Temperature
- Experimental conditions change, i.e. beam current
  - model uses high voltage board current as a proxy for beam current

Motivation: Online calibration to save time and computing resources required for post calibration

Conventional
- Calibrate: calibration values iteratively produced after the experiment
- No Control: CDC operating voltage is fixed at 2125 V

Online and ML
- Online, calibration and Control:
  Stabilize detector response to changing environmental/experimental conditions by adjusting CDC HV
GP for Calibration of the Central Drift Chamber (CDC)

**Cosmic Ray Experiment**
- Sorted the CDC into 2 halves
- Leave one side at a fixed HV (conventional)
- Let the ML control the other
- **Autonomously** adjust HV every 5 min

**Experimental Control Requires Uncertainty Quantification**
- Training data rarely has complete coverage: 601 runs, 2 run periods
- Do experts trust model predictions?

**Uncertainty Surface Mesh:**
- Threshold of standard deviation, visualize feature space
- **Uncertainty > threshold**
  - Approach-1: use prediction for the closest point on the surface
  - Approach-2: return CDC to traditional high voltage setting, 2125 V and collect more data to train
Large data sets with higher dimensions and GP

- GP provides robust UQ but scales poorly with increase in data samples ($O(n^3)$)
- For large datasets and/or in large feature space, **GP approximation** is required
- Deep Neural Networks (DNN) are very expressive and scales with size of dataset
- DNN can deal with different types of data including images, text, and timeseries
- What about UQ for DNN?
UQ in Deep Learning

(a) MC Dropout

(b) Ensemble

(c) Quantile Regression

(d) Conformal Predictions

GP takes into account distance between input samples explicitly

Spectral Neural Gaussian Process

- Distance preservation via Spectral Norm on each hidden layer
  - Reduced expressivity, harder to learn
- Distance preservation via bi-lipschitz

\[ L_1 \times ||x_1 - x_2|| \leq ||h_{x_1} - h_{x_2}|| \leq L_2 \times ||x_1 - x_2|| \]

\( L_1 \) and \( L_2 \) are hyper-parameters, \( h_{x_1}, h_{x_2} \) are hidden layer outputs corresponding to inputs \( x_1, x_2 \) respectively.

https://arxiv.org/abs/2107.07511
Anomaly Prediction at SNS Accelerator

- Accelerators are complex multi-system machine
- Failure in any equipment can cause errant beams
- Fault prediction is beneficial in many ways including reduced downtime

**Goal:** To predict errant beam pulses (with uncertainty quantification) before they occur to avoid potential damage to the equipment(s) and reduce the downtime

**Dataset:**
- We used the macro-pulse before an errant beam pulse (and labeled it as anomaly) and macro-pulses from the normal operation (and labeled them as normal) for our studies
- Our hypothesis: there is a sign about upcoming anomaly in macro-pulses even before it happens
- We also need to forecast the fault within a **short time window** to be actionable
Uncertainty Aware Siamese Model and Results

- We use Siamese Model to predict similarities with normal reference waveforms
- Attached GP approximation layer at the end to provide UQ
- The ROC curves bands are produced by smearing the predictions with uncertainty
- To evaluate the OOD uncertainty robustness
  - Introduced a different anomaly type (not included in training)
  - The model predicts OOD anomaly reasonably well with higher uncertainty

Maximize TPR at FPR < 0.5%
Uncertainty Aware PID for SoLID

Cherenkov Readout:

Each 5cmx5cm MAPMT can have

- 1 pmt sum output with Number of photoelectrons (Npe)
- 4 2.5cm x 2.5cm quad sum output with Npe
- 64 6mm x 6mm pixels with 0/1
- Cost increase with more readout channels and increasing resolution

Goal:

Use ML to understand what level of readout is needed to achieve desired performance

Dataset:

- Readout hits on the sensors from PMT having the lowest to Pixel readout having the highest resolution
Uncertainty Aware PID for SoLID

- Pion vs Kaon
- The PID data is converted to images
- The deep learning model is designed using conv2D layers to learn the patterns
- GP approximation layer to provide UQ
- Simulation data is used with higher background noise to mimic the realistic scenario

- DNN based PID model has shown much better performance than traditional cuts
- Next Step: Evaluate the UQ robustness by introducing third/noise particle

<table>
<thead>
<tr>
<th>Section</th>
<th>PMT</th>
<th>Quad</th>
<th>Pixel</th>
<th>Npe cut</th>
</tr>
</thead>
<tbody>
<tr>
<td>P 2.5 GeV theta 8.0 degree</td>
<td>0.978</td>
<td>0.991</td>
<td>0.994</td>
<td>0.5</td>
</tr>
<tr>
<td>(most difficult for pion acceptance)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P 7.5 GeV theta 14.0 degree</td>
<td>0.991</td>
<td>0.995</td>
<td>0.997</td>
<td>0.6</td>
</tr>
<tr>
<td>(most difficult for kaon rejection)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

True prediction rate at False Positive rate of 5%
Uncertainty Aware Surrogate Model for FNAL Booster

The Booster accelerates 400 MeV beam to 8GeV with the help of booster cavities and Combined-function bending and focusing electromagnets known as gradient magnets.

These magnets are powered by the gradient magnet power supply (GMPS)

Fluctuations in GMPS electrical current due to neighboring electrical loads and thus fluctuations of the magnetic field in the Booster gradient magnets

Problem definition:
Develop a surrogate model of the GMPS that can be leveraged to train a modern control system such as reinforcement learning agent to stabilize the GMPS power supply

A GMPS regulator is required to stabilize the accelerating magnetic field
Uncertainty Aware Surrogate Model for FNAL Booster

- Surrogate model is developed using ResNet Conv1D blocks followed by fully connected layers
  - UQ is important to know which areas are not well modeled by the surrogate model
- We evaluated three different methods of UQ for DNN regression
  - DQR provide robust UQ in-distribution but not on OOD
  - BNN method does a better job at estimating OOD uncertainty
  - GP approximation model provide the best OOD estimation and is calibrated by design

In distribution data


The high frequency with higher fluctuation in amplitude are removed from training to make it OOD

One of the input variable is monotonically increased manually to make it go OOD
Conclusion

• GP provides robust uncertainty quantifications for both in-distribution and OOD samples
• Presented application of GP to autonomously calibrate CDC in GlueX detector
• With larger datasets and higher dimension, GP does not scale; approximation required
• DNN can handle large multi-model datasets with ease
• GP can be attached to DNN models to provide UQ
• Distance between input samples need to be preserved through hidden layers
• Presented uncertainty aware anomaly prediction for SNS accelerator with GP approximation
• Presented comparison of PID in SoLID with different readout resolutions
• Presented uncertainty aware surrogate model for FNAL booster

Acknowledgement  This work was supported by the DOE Office of Science, United States under Grant No. DE-SC0009915 (Office of Basic Energy Sciences, Scientific User Facilities program) and DE-FG02-87ER40315. ORNL is managed by UT-Battelle, LLC for the US Department of Energy (DOE) under contract DE-AC05-00OR22725. Jefferson Laboratory is managed by Jefferson Sc. Assoc., LLC for the US Department of Energy (DOE) under Contract No. DE-AC05-06OR2317. This research used resources at the Spallation Neutron Source, a DOE Office of Science User Facility operated by the Oak Ridge National Laboratory, resources at GlueX experiment, and upcoming SoLID experiment at Jefferson lab, and open sourced data from Fermi National Accelerator Laboratory.
References


