Common Problems in Early-Stage Projects at the ISIS Neutron and Muon Source

Kathryn Baker

k.baker@stfc.ac.uk

ICALEPCS Data Science and Machine Learning Workshop

7th October 2023

UKRI Science and Technology Facilities Council

ISIS Neutron and Muon Source

Agenda

1. Overview of ML at ISIS Accelerators
2. Common Requests from users / clients
3. Common Problems encountered
4. Discussion

Machine Learning at ISIS Accelerators

- Serious data collection started in 2019
  - InfluxDB automatically archiving value changes from Vsystem
- Relatively new team
  - Started in 2020 with 2 x 0.5FTE (graduate level) – now at just under 2 FTE
  - All split approx. 50/50 with software development for controls
- First ~2 years building up knowledge of ML
- Support from Scientific Machine Learning (SciML) group in SCD
  - ML knowledge
  - No domain-specific knowledge
- Small-scale projects, nothing in use in the MCR yet
- Increased focus on MLOps (see **TUMBCMO26B**)

Machine Learning at ISIS Accelerators

Project	Data Source	Status
Anomaly Detection in Target Station 1 (TS1) Methane Moderator	Archive data	Shelved (TS1 upgraded)
Surrogate Model of Bunch Length Evolution	Physics simulation (in-house)	Paused
Surrogate Model of Beam Descriptors for Medium Energy Beam Transport (MEBT)	Physics simulation (ASTRA)	Deployed / Active development
Surrogate Model of Low Energy Beam Transport (LEBT)	Archive data	Active development
Anomaly Detection in Helium Recovery System	Archive data	Active development
Anomaly Detection in Controls SysLog Messages	Archive data	Deployment

Common Requests from Users

- Anomaly Detection & Prediction
    - Operators want to know what has gone wrong (ideally ahead of time) to prevent it causing down-time
- Root cause analysis (RCA)
    - If a trip does happen, they want to be able to identify what the root cause was
    - Given lots of log-book data, can we find out what the cause was for problems that caused the most down time?
- Intelligent control
    - For difficult to control systems, can data-driven approaches be used to improve control?
    - Given a description of a beam, can we predict the settings required to get us there?
- Improved machine understanding
    - Virtual diagnostics of the machine
    - Can we use data-drive approaches / ML to better understand differences between simulation and the machine?

Common Problems Encountered – Anomaly Detection / RCA

- Which anomalies?
  - Operators often don’t know which channels they want to study
  - Anomalies happen all over the machine
  - How do you find the biggest offenders?
- What counts as an anomaly?
  - Not all ‘trips’ or anomalies are noticed / recorded
- Anomalies are often infrequent but important
  - May only happen once a cycle but causes many hours of downtime
- High dimensional feature selection
  - ~35,000 channels – how do we know which ones to use?
- How do we disentangle correlation and causation?
  - Operators don’t want black boxes

Common Problems Encountered – Virtual Diagnostics

- Highly correlated outputs
  - Input – output correlation is low but output-output correlation is high (e.g. in LINAC)
  - Solved using specific model architectures?
- Trained on Physics Simulations but deployed to machine
  - Often don't match / agree (miscalibrations)
  - What if there isn't an associated measurement to calibrate against?
- Drift in machine parameters
  - Distribution no longer matches training distribution

I welcome all your thoughts!

Discussion