AI for Experimental Controls

Diana McSpadden dianam@jlab.org,
Torri Jeske roark@jlab.org
Naomi Jarvis, Thomas Britton, Kishansingh Rajput, and David Lawrence
AI Town Hall, July 26th 2021

Carnegie Mellon University

GlueX

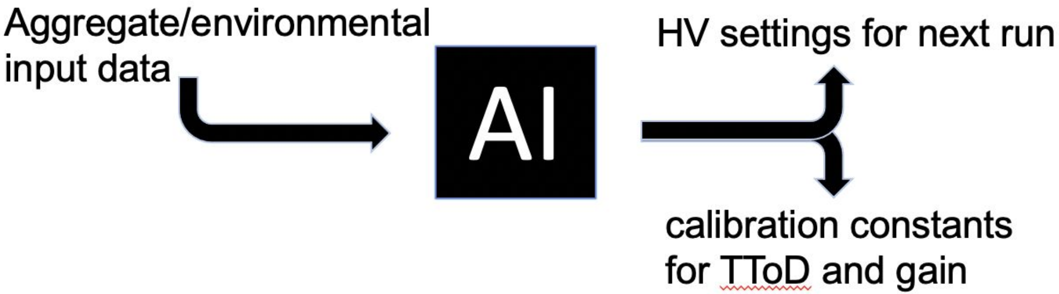
EPSCI
Experimental Physics Software and Computing Infrastructure

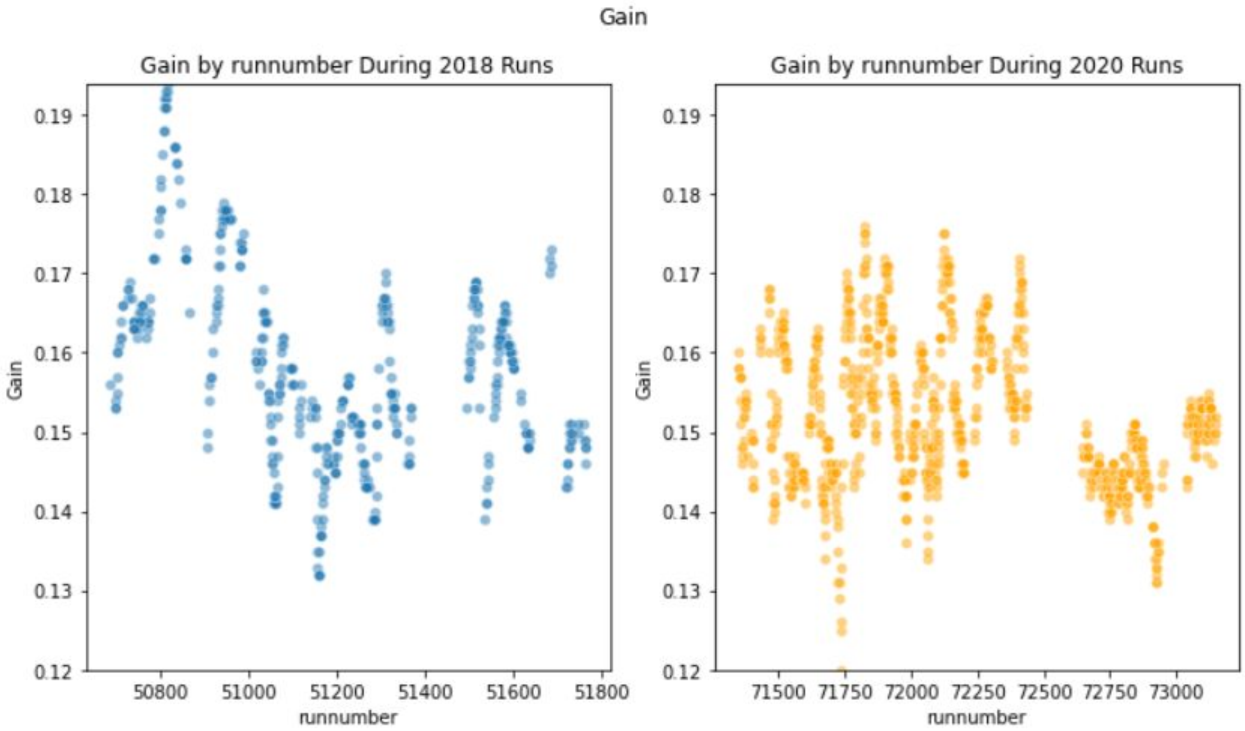
AIEC Goals:

- AI-recommended HV settings to maintain GlueX central drift chamber gain
  - Chamber gain is sensitive to atmospheric pressure

- **Have neural network determine calibration constants as quickly as possible**
  - **Reduce time for offline calibration**

- Application to other detector systems such as CLAS12 spectrometer





Can we use a neural network to predict existing gain constants to within 1%?

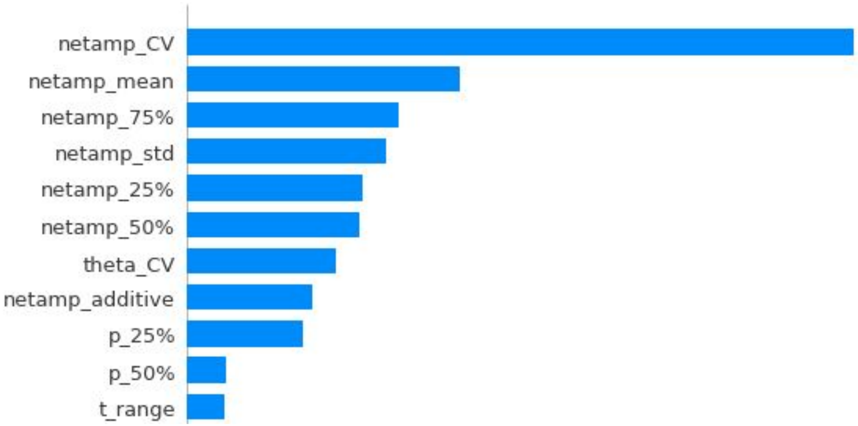
The AI Process:

Input Features:

- Aggregate features per run from experimental data and EPICS system:
    - Netamp = pulse height - pedestal, momentum, track angle, drift time

- Split data into train and test sets:
    - 438 runs from 2018
        - 350 train
        - 88 test
    - 897 runs from 2020
        - 717 train
        - 180 test

Feature Importance:



Example output of average importance of features using the **shap python library**.

Current status:

- **4 Tensorflow sequential neural networks** have been hyperparameter tuned based on each set of features
  - Contain varying number of input features
  - Contain 2 or 3 layers
  - Use sigmoid and/or relu activation functions
- Success metric: within 1% max absolute error percentage

All 2020 runs perform within 1%

97.8% of 2018 test data perform within 1%

What’s next:

- Near term:
  - Predicting time to distance calibration constants
  - Improving gain model(s)
  - Verifying predicted gain constants
- Intermediate term:
  - Collecting data at slightly different HV setting
- Long term:
  - Deployment for GlueX
  - Application to CLAS12 drift chambers

Please send all questions/comments/suggestions via email :)

backups

Training and Test Data Set Information

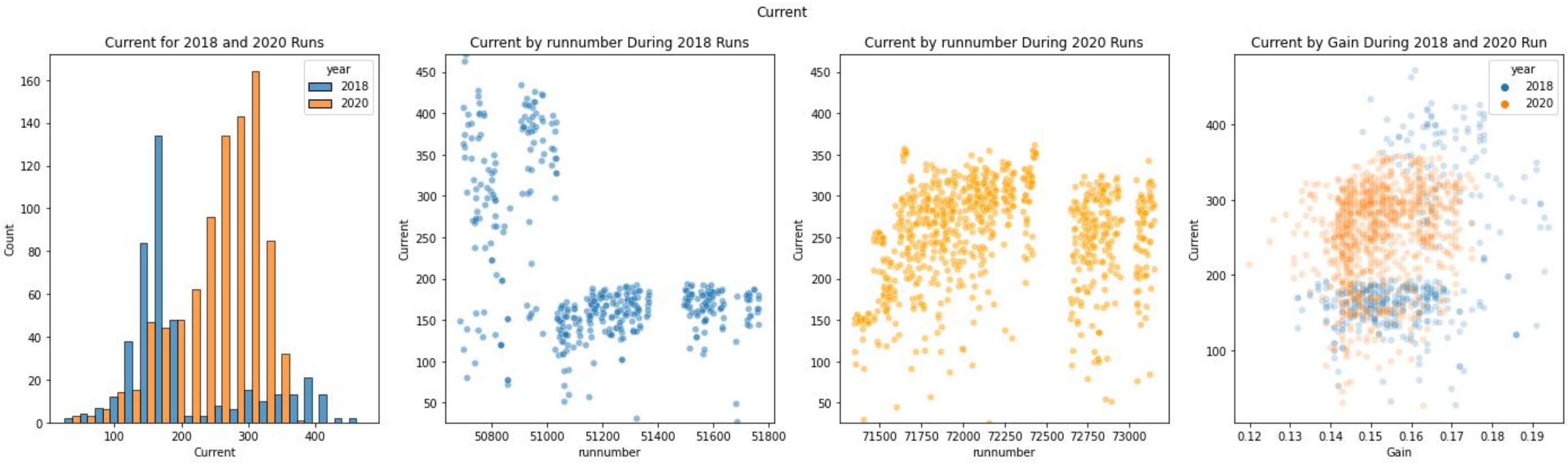
1. runs_epics_gain2018.csv == all 2018 data (count: 438)
2. runs_epics_gain2020.csv == all 2020 data (count: 897)
3. test-holdout2018.csv == final testing/holdout dataset, only 2018 data (count: 88)
4. test-holdout2020.csv == final testing/holdout dataset, only 2020 data (count: 180)
5. test-holdout20182020.csv == union of test-holdout2018.csv and test-holdout2020.csv (count: 268)
6. train-validate2018.csv == training dataset for training and validation, only 2018 data (count: 350)
7. train-validate2018w.csv** == training dataset for training and validation, only 2018 data (count: 412)
8. train-validate2020.csv == training dataset for training and validation, only 2020 data (count: 717)
9. train-validate20182020.csv union of train-validate2018.csv and train-validate2020.csv (count: 1067)
10. train-validate2018w2020.csv** union of train-validate2018.csv and train-validate2020.csv (count: 1129)

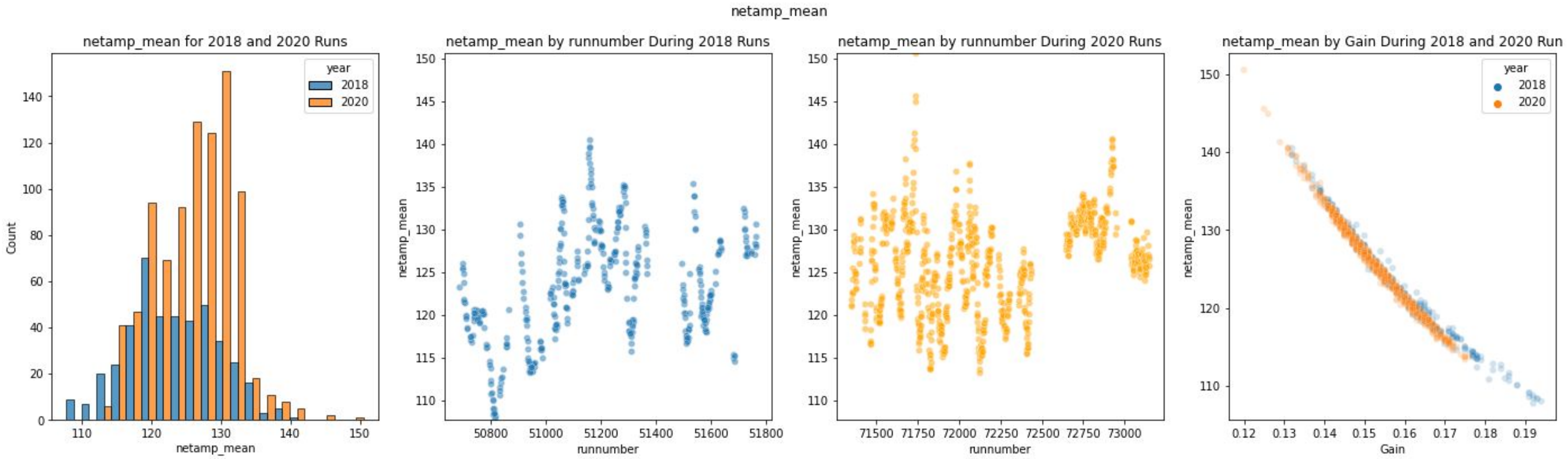
** weighted, i.e. upsampled 2018 training data

Models Discussed

- "11 Feature Vanilla LR" Linear Regression model with following features:
  - 'Pressure',
  - 'netamp_CV', 'netamp_mean','netamp_75%', 'netamp_std', 'netamp_25%','netamp_50%'
  - 'theta_std','theta_75%'
  - 'p_25%','p_50%'
- "11 Feature NN": Sequential Neural Net model:
  - 'Pressure'
  - 'netamp_CV', 'netamp_mean','netamp_75%', 'netamp_std', 'netamp_25%','netamp_50%'
  - 'theta_std','theta_75%'
  - 'p_25%','p_50%'
- "No Mode NN": Sequential Neural Net model with no mode features:
  - 'Pressure','Current',
  - 'eventCount',
  - 'netamp_std', 'netamp_25%', 'netamp_mean', 'netamp_50%', 'netamp_max', 'netamp_75%', 'netamp_CV'
  - 'theta_std', 'theta_min', 'theta_mean', 'theta_50%', 'theta_75%', 'theta_CV'
  - 'p_mean', 'p_std', 'p_min', 'p_25%', 'p_50%', 'p_75%', 'p_max', 'p_CV'
  - 't_mean', 't_std', 't_min', 'theta_25%', 't_25%', 't_50%', 't_max', 't_CV'
  - calculated features:
    - 'netamp_additive' = 'netamp_25%' + 'netamp_50%' + 'netamp_75%' + 'netamp_max'
    - 't_additive' = 't_min' + 't_25%' + 't_50%' + 't_max'
    - 't_range' = 't_max' - 't_min'
    - 'p_range' = 'p_max' - 'p_min'
- "37 Feature NN": Sequential Neural Net model - results currently **only available on 2018 data**:
  - 'Pressure', 'Current'
  - 'event_count'
  - 'netamp_std','netamp_CV','netamp_mean','netamp_75%','netamp_50%','netamp_25%','netamp_max'
  - 'theta_CV','theta_75%', 'theta_25%','theta_50%','theta_mean','theta_mode','theta_min','theta_std'
  - 'p_75%', 'p_25%','p_50%','p_CV','p_min','p_mean', 'p_range','p_max','p_std'
  - 't_CV', 't_50%','t_75%','t_25%','t_additive','t_min','t_max','t_range','t_std','t_mean'
- "39 Feature NN": Sequential Neural Net model - results currently **only available on 2018 data**:
  - 'Pressure','Current',
  - 'eventCount',
  - 'netamp_std','netamp_CV','netamp_mean','netamp_75%','netamp_50%','netamp_additive','netamp_25%','netamp_max','netamp_mode',
  - 'theta_CV','theta_75%','theta_25%','theta_50%','theta_mean','theta_mode','theta_min','theta_std',
  - 'p_75%','p_25%','p_50%','p_CV', 'p_min','p_mean','p_range','p_max','p_std',
  - 't_CV','t_50%','t_75%','t_25%','t_additive','t_min','t_max','t_range','t_std','t_mean'

Candidate Models


Current
Current for 2018 and 2020 Runs
Count
Current
year
2018
2020
Current by runnumber During 2018 Runs
Current
runnumber
Current by runnumber During 2020 Runs
Current
runnumber
Current by Gain During 2018 and 2020 Run
Current
Gain
year
2018
2020


netamp_mean
netamp_mean for 2018 and 2020 Runs
Count
netamp_mean
year
2018
2020
netamp_mean by runnumber During 2018 Runs
netamp_mean
runnumber
netamp_mean by runnumber During 2020 Runs
netamp_mean
runnumber
netamp_mean by Gain During 2018 and 2020 Run
netamp_mean
Gain
year
2018
2020

Understanding importance of an individual feature in a multivariable, non-linear function?

For regression problems, there are a number of available evaluation methods.

We implemented Shapley values.

“The Shapley value is a framework originally proposed in the context of game theory to determine individual contributions of a set of cooperating players” - [Explaining Deep Neural Networks and Beyond: A Review of Methods and Applications | IEEE Journals & Magazine | IEEE Xplore]