Data Science for Linac Controls (L-CAPE Project)

Brillina Wang, Elgin Community College — Jason St. John

FERMILAB-POSTER-23-162-STUDENT

Introduction

Linear accelerators (Linacs) accelerate H- ions and subject them through a series of electric potentials along a linear beamline. They are mainly composed of radiofrequency (RF) cavities placed in-line with one another to provide a large amount of energy gain per unit length.

Machine learning (ML) refers to the training of machines to get better at a task without explicit programming. Anomaly detection is an ML method of identifying unexpected events or observations that differ significantly from the predicted value. At Fermilab AD, L-CAPE is focused on predicting and preventing Linac beam outages with anomaly detection.

Purpose

L-CAPE (Linac Condition Anomaly Prediction of Emergence) is a project that aims to apply data science methods in order to improve information from the linear accelerator, to use machine learning techniques to automatically label unplanned beam outage types as they occur, and to identify patterns within the data that could help predict and prevent the occurrence of future outages.

In this project, outage file data was read and processed in order to be prepared for the outage labeling step. These files contain data sequences around identified beam outages.

Results

	B:BLMLAM@e,52,e,0_timestep1_10	...	L:BPV204@e,0A,e,0_timestep10_10
outage_100.parquet.snappy	0.009798	...	0.009593
outage_101.parquet.snappy	0.009392	...	0.009343 NaN
outage_102.parquet.snappy	0.009392	...	0.009695

Figure 3: Table representation of example output

Figure 1: Tevatron, a particle accelerator at Fermilab

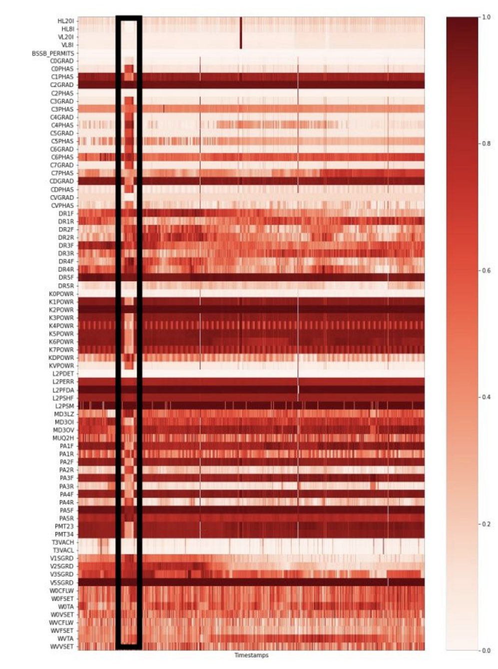

Figure 2: Heatmap of data collected by various devices (Jan Strube PNNL Private Correspondence)

Method

Read in each outage file of type .parquet.snappy as a Pandas DataFrame. In a loop for each file, create a list of column names and add each column name to a set. Ran .intersection() on the set of column names for each file until the largest set of columns common to all files is obtained. Also, find the smallest number of rows before and after the outage onset. This process allows the data to be standardized for the DataFrame.

Created two functions and ran both inside the loop. The first function, prepare_dataset_seq takes the DataFrame created from the current outage file being read and the timesteps and returns a flattened list of data per every timestep until the outage starts. The second function, flattened_col_names, takes the set of columns common to all files and the timesteps, changes the list by adding every timestep to each column's name, and returns this as a list. This would be the set of Linac devices to use. A command line was then created through argparse that takes the directory and timesteps as parameters.

In Figure 3, the table that represents the final DataFrame, 3 files were read and the timestep parameter was 10. The first column (index) of the table is the name of the current outage file that was being read. The first row (column headers) of the table contains the column names, which are the device labels from are in the format "{label}_timestep {increment}_{timesteps}", obtained from the flattened_col_names function.

When the code ran for the entire directory of outage files, there were 274 rows and 25740 columns, so there were 274 total files and 25740 device names.

Acknowledgements

The technology, software, and data associated with L-CAPE were determined to be Fund Research Exclusion eligible, which removes export control related restrictions, as of August 3, 20202, by Paul Lauper Ellison, Export/Import Control Compliance Manager, Fermilab Office of General Counsel.

This manuscript has been authored by Fermi Research Alliance, LLC under Contract No. DE-AC02-07CH11359 with the U.S. Department of Energy, Office of Science, Office of High Energy Physics.

This work was supported in part by the U.S. Department of Energy, Office of Science, Office of Workforce Development for Teachers and Scientists (WDTS) under the Community College Internships Program (CCI)