Kucherenko Viacheslav

17/06/22

IRIS-HEP Project Proposal

Introduction

The upcoming sPHENIX experiment generates petabytes of data by detections. The unprecedented data rate generated by the background and interesting events from the modern nuclear-energy physics detectors overwhelms state-of-the-art data storage and computing facilities. The complex physics detectors act like cameras to record what happens right after particle collisions. Physicists often use hardware triggers to make real-time decision to collect/ discard events and filter out background events, and only retains a small fraction of data that include triggers. The hardware triggers are rigid and rely on simple physics properties to recognise events. The events of sPHENIX experiment have more complex patterns and invalidate the assumption associated with the hardware trigger. We propose to use AI/ML to perform software-based triggering processing and recognise complex patterns of particle decays.

Here we use a collection of hits that are lighted up by passing particles from the collision locations to describe events. The pipeline process of the raw events hit consists of three steps: 1) cluster, 2) tracking, and 3) tree. I propose to focus on Step 2 and improve the tracking performance. Tracking algorithms attempt to connect hits into the trace of particles and plays a vital role in event detection

Planned Work

This project will focus on tracking algorithms, which can effectively identify tracks of the particles, having the phase-space distribution of points on detectors. Using geometrical constraints and a graph neural networks approach it is possible not only to identify tracks but also to apply trigger detection classification on the preprocessed tracks. After having a model, the goal would be to fine-tune the best approaches and compare their final metrics, such as AUC and accuracy. This project will be under the mentorship of Oksana Shadura and Dantong Yu.

Software Deliverables

The programming language that will be used for the project is Python with machine learning frameworks, such as Pytorch, Tensorflow, and Keras. The workspace that the project will be contained in a private Github repository with the project code and model weights used for training.

Timeline

The period of the project will be lasting for 10 weeks starting from 27 June 2022 to 26 August 2022.

Week 1-2: Literature review for the project background. sPHENIX detector data characteristics studying, understand ground knowledge and math of the Graph Neural Networks. Become familiar with simulation packages Pythia 8.3 and Geant4. Make an introductory presentation for this part of the project.

Week 3-4: Dataset Exploratory Data Analysis among the simulation packages and simulation features exploring. Test and apply feature engineering and encoding methods in order to reduce data dimensionality and prepare an encoded version of the dataset, ready for training of the model. The mid-project review will be made so far in order to adjust the timelines and goals accordingly.

Week 5-8: Constructing the GNN and establishing a model pipeline in order to form the continuous input-output process of the tracking and trigger detection. Training model on the prepared dataset, hyper-parameters tuning, retrieving metrics for the training and validation distributions.

Week 9-10: Preparing the summary of the project work. Compare results of the models with the metrics of the analogues, document code, upload it to a repository, preparing and submit the final report to the mentors.