IRIS-HEP Project Proposal: Geometric Machine Learning for Particle Tracking

Student: Sophia Paulino Korte
Date: 06/23/2022

Proposal

The Standard Model describes particles and their interactions via the electromagnetic, strong and weak forces. However, the Standard Model is an incomplete theory and does not explain many things, like gravity, the neutrino masses or why there is more matter than antimatter in the universe. So, scientists have proposed other theories, like supersymmetry, in which each of the known particles must have a superpartner with a 1/2 spin difference.

In order to study the particles that we know from the Standard Model and also new particles and phenomena, the Large Hadron Collider (LHC) accelerates and collides protons, which can produce other particles. At the LHC there exists two general purpose detectors to study these interactions. One of these detectors, the Compact Muon Solenoid (CMS), is a cylindrical detector with a solenoid at the center that provides a magnetic field of 3.8 T. This solenoid bends the tracks of the charged particles produced in proton-proton interactions making it possible to calculate the charged particles' momentum. Since many interactions happen at the CMS, events of interest are selected using a two-level trigger system. The first level (Level-1) composed of highly specialized electronics that employ FPGAs as their primary processors and has only 3.5 microseconds to process each event; it uses only the muons systems and calorimeters for event reconstruction and does not receive tracking information. The Level-1 trigger is responsible for the first round of selection. The second level, the High Level Trigger (HLT), is able to receive information from all sub-detectors. It consists of a computing farm and has a few milliseconds for event processing. Track reconstruction is computationally very expensive and can take a significant fraction of the HLT event processing time.

The LHC will undergo an upgrade to increase the instantaneous luminosity after Run-3: the High Luminosity Large Hadron Collider (HL-LHC) will have 5x more collisions per event. Occupancy of hits in the tracking system will increase proportionally. Traditional tracking algorithms, such as Kalman filter, increase exponentially with pileup. ML-based algorithms for track reconstruction have been proposed which scale linearly with pileup.

This project uses geometric machine learning methods, specifically Graph Neural Networks (GNNs), as a way of reconstructing the trajectories of particles. Detector data can be represented as a graph, with the hits of particles as nodes and the possible trajectories as edges in the graph. The GNN is an Interaction Network (IN) that has three steps: graph building, finding the edge weights through a process called edge classification, then building the track. For that we consider edge weights above a certain threshold and apply a clustering algorithm. Algorithms such as DBSCAN and UnionFind are then used to estimate the track kinematics. In this project I will use the CMS simulation data to evaluate the performance of the GNN tracking algorithms, compare them to the CMS algorithms currently in use, and work on optimizing the GNN.

Proposed timeline

Week 1: focus on learning more about machine learning, Pytorch, graphs, review Python syntax, and make sure I can run the Interaction Network
Week 2: optimization studies, learn more about adjacency matrices and graph building and also study different types of Neural Networks, multi-layer perceptrons (MLPs) and message passing neural networks
Weeks 3-5: study and learn about edge building algorithms, how to get edge weights and clustering
Weeks 6-8: run samples using CMS simulation
Week 9-11: optimize GNN, compare the efficiency of the traditional algorithms with the GNN algorithms
Week 12: prepare final presentation