Graph Neural Networks for Low-Momentum Tau Identification

Angshuk Dutta

April 13, 2021

angshukdutta543@gmail.com 84049970

Abstract

The search for new physics pertaining to high energy particles is of paramount importance. This requires accurate reconstruction of particle – particle collisions in the LHC, identification of the particles involved and understanding the causal relations between the Tracker, the ECAL and the HCAL present in the CMS. Previous work has relied on traditional machine learning methods such as boosted decision trees. With the advent of deep learning, the physics community has transitioned to these methods to find underlying patterns and relations between the data from the three components mentioned above. However, current methods require the data (collisions and decay) to have a determined structure i.e. order, thus requiring the problem to be reduced to the release of diphotons which though efficient, reduces generalisability. In this proposal we suggest the steps to the construction of an end to end Graph Neural Network for the identification of low-momentum tau particles. The choice of Graph Neural Network is made to propose an order-invariant algorithm without relinquishing the accuracy achieved by previous methods.

1 Deliverables

1. Fully Functioning Graph Neural Network Code.
   - Experimental Results from the Implementation of Simple GNN model such as GCN, GAT and GIN
   - Experimental Results from the implementation of ensemble models (combination of models)
2. A rigorous Data Analysis report.
   - Checking for Class imbalance for Classification task.
   - Identifying Correlations between different variables in the Dataset.
   - Creating the Connected Graph Dataset.
3. Documentation of the code.

2 Proposal

In this Section we introduce the proposed timeline for the entire project. Note that this is subject to change based on the progress of the project. Hence a two week buffer has been kept for any unforeseen delays.

1. Before May 17.
   - Keep in touch with the Community and Mentors
   - Go through the Source Code for CMS software on Github to familiarise myself.
2. May 17 to 26 May.
   - Literature review of existing methods in dynamical systems implementing GNN (kipf et. al)
   - Keep in touch with mentor about the specification of the application.
3. May 27 to Jun 10
   - Understand the Data and Analyse the Dataset.
   - Keep in in touch with Mentor to Understand the Data better.
4. Jun 10 to Jun 20
   - Carry out Data Cleaning and Create the Graph Dataset from the raw data.
   - Implement a Simple GNN model (2 layers) to do the Classification task
   - Explore and visualise the Results
5. Jun 21 to Jun 30
   - Understand performance bottlenecks if any and see if it meets the specification mentioned by the mentors.
   - Discuss with the mentors why the bottlenecks exist and come up with better solutions.
6. July 1 to July 12
   - Explore complicated architectures i.e. Ensemble models such as a combination of Graph-Sage and GAT.
   - Implement multiple Loss functions and Discuss the relevance of the functions with mentors
7. July 13 to July 17
   - Evaluation Period
8. July 17 to Aug 1
   - Integrate the model with the existing CMS software system and document the Code.
   - Discuss with mentors if any further modification is required and whether it meets the standards set by them.

9. Aug 2 to Aug 17
    - Kept as buffer for any unforeseen delays.
    - Test flawless execution of the model
    - Write a report on the Preprocessing and the model architecture
    - Thank the mentors for their support during the project.

References

Andrews, M. (2018, July 31). End-to-End Physics Event Classification with CMS Open Data: Applying Image-Based Deep Learning to Detector Data for the Direct Classification of Collision Events at the LHC. ArXiv.Org. https://arxiv.org/abs/1807.11916 Hamilton, W. L. (2017, June 7).

Inductive Representation Learning on Large Graphs. ArXiv.Org. https://arxiv.org/abs/1706.02216