



Contribution ID: 208

Type: Oral

The Federation - A novel machine learning technique applied on data from the Higgs Boson Machine Learning Challenge

Tuesday, October 25, 2022 5:40 PM (20 minutes)

The Federation is a new machine learning technique for handling large amounts of data in a typical high-energy physics analysis. It utilizes Uniform Manifold Approximation and Projection (UMAP) to create an initial low-dimensional representation of a given data set, which is clustered by using Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN). These clusters can then be used for a federated learning approach, in which we separately train a classifier on the data of each individual cluster. As a requirement for this approach, we need to apply an imbalanced learning method to the data in the found clusters before the training. By using a Dynamic Classifier Selection method, the Federation can then make predictions for the whole data set. As a proof of concept for this novel technique, open data from the Higgs Boson Machine Learning Challenge is used and comparisons to results from established methods will be presented. We also investigated the issue of handling missing values and the jet-count feature for this data.

Significance

First application of UMAP and imbalanced learning methods in high-energy particle physics.

References

Experiment context, if any

Primary author: MUCHA, Maximilian (University of Bonn (DE))

Co-author: VON TORNE, Eckhard (University of Bonn (DE))

Presenter: MUCHA, Maximilian (University of Bonn (DE))

Session Classification: Track 2: Data Analysis - Algorithms and Tools

Track Classification: Track 2: Data Analysis - Algorithms and Tools