Contribution ID: **234**   Type: **Poster**

ROOT Machine Learning Ecosystem for Data Analysis

Thursday, 27 October 2022 11:00 (30 minutes)

Through its TMVA package, ROOT provides and connects to machine learning tools for data analysis at HEP experiments and beyond. In addition, ROOT provides through its powerful I/O system and RDataFrame analysis tools the capability to efficiently select and query input data from large data sets as typically used in HEP analysis. At the same time, several existing Machine Learning tools exist in a diversified landscape outside of ROOT.

In this talk, we present new developments in ROOT that bridge the gap between external tools and ROOT, by providing better interoperability in a common software ecosystem for Machine Learning in data analysis.

We present recently included features in TMVA allowing for generating batches of events for ROOT I/O and RDataFrame to train efficiently machine learning models using Python tools such as Tensorflow and PyTorch. This will facilitate direct access to the ROOT input data when training using external tools. Another focus is put on fast machine learning inference, which enables analysts to deploy their machine learning models rapidly on large scale datasets. A new tool has been recently developed in ROOT, SOFIE, allowing for generating C++ code for evaluation of deep learning models, which are trained from external tools. This provides the capability to better integrate Machine Learning model evaluation in HEP data analysis.

The new developments are paired with newly designed C++ and Python interfaces for TMVA supporting modern C++ paradigms and providing full interoperability in the Python ecosystem.

Significance

This presentation covers some novel results, teh development of a batch generator for better integration of ROOT RDataFrame with external Machine tools for training models.

Furthermore it will contain an update on other TMVA developments such as SOFIE which has greatly developed since the last ACAT presentation, being able to parse complex ML models used by LHC experiments.

References

Experiment context, if any

Primary author: MONETA, Lorenzo (CERN)

Co-authors: AN, Sitong (CERN, Carnegie Mellon University (US)); ZAPATA MESA, Omar Andres (University of Antioquia & Metropolitan Institute of Technology); SENGUPTA, Sanjiban; HAMDAN, Ahmat

Presenter: MONETA, Lorenzo (CERN)

Session Classification: Poster session with coffee break