Contribution ID: **212**   Type: **Poster**

Scaling MadMiner with a deployment on REANA

Thursday 27 October 2022 11:00 (30 minutes)

MadMiner is a python module that implements a powerful family of multivariate inference techniques that leverage both matrix element information and machine learning.

This multivariate approach neither requires the reduction of high-dimensional data to summary statistics nor any simplifications to the under-lying physics or detector response.

In this paper, we address some of the challenges arising from deploying MadMiner in a real scale HEP analysis with the goal of offering a new tool in HEP that is easily accessible.

The proposed approach streamlines a typical MadMiner pipeline into a parametrized yadage workflow in yaml files. The general workflow is split in two yadage subworkflows, one dealing with the physics dependencies and the other with the ML ones. After that, the worfklow is deployed using REANA, a reproducible research data analysis platform that takes care of flexibility, scalability, reusability and reproducibility features.

To test the performane of our method, we performed scaling experiments for a MadMiner workflow on the National Energy Research Sscientific Computer luster (NERSC) cluster with an HTCondor backend.
All the stages of the physics subworkfow had a linear dependency between resources & walltime and number of event generated. This trend has allowed us to run a typical MadMiner workflow consiting of 1M events and the generation step just used 2930 MB of memory and walltime of 2919s.

Significance

Our work is the first effort to scale MadMiner in order to be a reliable and widespread tool in the HEP community. We proved in the past that phyics analyisinvolving a multivariate method are more.

References

Experiment context, if any

ATLAS

Primary author: ESPEJO MORALES, Irina (New York University (US))

Co-authors: HURTADO, Kenyi; CRANMER, Kyle Stuart (New York University (US)); HEINRICH, Lukas Alexander (Max Planck Society (DE)); PEREZ, Sinclert (NYU)

Presenter: ESPEJO MORALES, Irina (New York University (US))

Session Classification: Poster session with coffee break

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