EUROPEAN AI FOR FUNDAMENTAL PHYSICS CONFERENCE
EuCAIFCon 2024

Contribution ID: **182**   Type: **Flashtalk with Poster**

b-hive: a modular training framework for state-of-the-art object-tagging within the python ecosystem at the CMS experiment

Wednesday, 1 May 2024 16:06 (3 minutes)

In high-energy physics (HEP), neural-network (NN) based algorithms have found many applications, such as quark-flavor identification of jets in experiments like the Compact Muon Solenoid (CMS) at the Large Hadron Collider (LHC) at CERN. Unfortunately, complete training pipelines often encounter application-specific obstacles like the processing of many and large files of HEP data format such as ROOT, the data provisioning to the model, and a correct evaluation of performance.

We have developed a framework called "b-hive" that combines state-of-the-art tools for HEP data processing and training in a Python-based ecosystem. The framework uses common Python packages like law, Coffea and pytorch bundled in a conda-environment, aimed for an uncomplicated setup. Different subtasks like dataset conversion, training, and evaluation are implemented inside the workflow management system "law", making the reproduction of trainings through built-in versioning and parametrization straightforward.

The framework is designed in a modular structure so that single components can be exchanged and used through parameters, making b-hive suited for not only production tasks but also network development and optimization. Further, fundamental HEP requirements as the configuration of different physics processes, event-level information, and kinematic cuts can be specified and steered in a single configuration without touching the code itself.

Primary author: EICH, Niclas

Presenter: EICH, Niclas

Session Classification: 4.4 Explainable AI

Track Classification: Session B