T 68: Data, AI, Computing 6 (ML utilities)

Time: Wednesday 16:00–18:00 Location: Geb. 30.34: LTI

T 68.1 Wed 16:00 Geb. 30.34: LTI

b-hive: a Model-Independent Machine Learning Training Framework for the CMS Experiment — •Mate Farkas[1], Niclas Eich[2], and Martin Erdmann[3] — [1]mate.farkas@rwth-aachen.de — [2]niclas.eich@rwth-aachen.de — [3]martin.erdmann@physik.rwth-aachen.de

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 finds application-specific obstacles like the processing of many and large root files, the data provisioning to the model, and a correct evaluation. 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 as law tasks, 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 not only suited for 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.

T 68.2 Wed 16:15 Geb. 30.34: LTI

Checkpointing of long running machine learning trainings on GPUs — •Jonas Eppelt, Matthias Schnepf, Giacomo De Pietro, and Günter Quast — Institute of Experimental Particle Physics (ETP), Karlsruhe Institute of Technology (KIT)

The rise of Machine Learning (ML) applications in High Energy Physics (HEP) analysis and reconstruction pushes for the use of GPUs in such workflows. The training of neural networks can have long runtimes, making them more susceptible to runtime constraints and failures. Checkpoints contain information about the current state of the training and therefore allow continuing the training from this state to another time and another place. This will provide resistance to failures and will allow for long runtimes while abiding to time constraints. Additionally, checkpointing will enable efforts in sustainable computing. For example, trainings can be run at times when renewable energies are available and haltet during times with limited energy supply. This talk presents a Python interface that bundles common tools from the HEP community to provide storing, transferring and restoring checkpoints for ML training.

T 68.3 Wed 16:30 Geb. 30.34: LTI

MLProf: Automated resource profiling of machine learning models in CMS production — •Nathan Prouvost, Marcel Rieger, and Peter Schleper — University Hamburg, Hamburg, Germany

With the increasing amount of data recorded by LHC experiments and collaborations, the efficient handling of computing resources is a topic of growing importance. In this regard, machine learning models, with their increasing number of applications and increasing specialization, can be difficult to include in central reconstruction workflows of an experiment.

For the CMS experiment at CERN, the awareness of this situation concerning time and memory budget has been growing in the last years. On top of these challenges, the integration of machine learning models into the CMS core software can be a very laborious task to begin with, impeding a fast feedback loop from performance measurements back to the model development. Therefore, a new tool for automating the extraction of resource consumption metrics of machine learning models within the CMS software environment, called MLProf, has been created.

This presentation introduces MLProf and its main features, including automated runtime measurements on different batch sizes, different versions of the CMS software environment and for different inference engines.

T 68.4 Wed 16:45 Geb. 30.34: LTI

Ahead-of-time (AOT) compilation of Tensorflow models — •Bogdan Wiederspan, Marcel Rieger, and Peter Schleper — Universität Hamburg

In a wide range of high-energy particle physics analyses, machine learning methods have proven as powerful tools to enhance analysis sensitivity. In the past years, various machine learning applications were also integrated in central CMS workflows, leading to great improvements in reconstruction and object identification efficiencies.

However, the continuation of successful deployments might be limited in the future due to memory and processing time constraints of more advanced models evaluated on central infrastructure.

A novel inference approach for models trained with TensorFlow, based on Ahead-of-time (AOT) compilation is presented. This approach offers a substantial reduction in memory footprints while preserving or even improving computational performance.

This talk outlines strategies and limitations of this novel approach, and presents integration workflow for deploying AOT models in production.

T 68.5 Wed 17:00 Geb. 30.34: LTI

Hog: handling HDL repositories on git — •Davide Cieri — Max-Planck-Institut für Physik, Munich, Germany

Handling HDL project development within large collaborations presents many challenges in terms of maintenance and versioning, due to the lack of standardized procedures. Hog (HDL on git) is a tcl-based open-source management tool, created to simplify HDL project development and management by exploiting git and GitLab/GitHub Continuous Integration (CI).

Hog is compatible with the major HDL IDEs from Xilinx, Intel, and Microsemi, and guarantees synthesis and placing reproducibility and binary file traceability, by linking each binary file to a specific git commit. Hog-CI validates any changes to the code, handles automatic versioning and can automatically simulate, synthesise and build the design.

T 68.6 Wed 17:15 Geb. 30.34: LTI

Improved Clustering with Graph Neural Networks on FPGAs for the Electromagnetic Calorimeter Trigger at Belle II — •Isabel Haide[1], Torben Ferber[1], and Marc Neu[2] — [1]Institute of Experimental Particle Physics (ETP), Karlsruhe Institute of Technology (KIT) — [2]Institut fuer Technik der Informationsverarbeitung (ITIV), Karlsruhe Institute of Technology (KIT)

For the Belle II experiment, beam background plays a very impactful role, especially on the hardware trigger level. Due to the maximum latency of 1.3 μs the current trigger algorithm for the Belle II electromagnetic calorimeter uses a simple clustering mechanism that, especially in high beam background, identifies a high number of fake clusters and is additionally unable to separate overlapping clusters. As Belle II plans to increase its luminosity by a factor of 40, an update of the trigger algorithm will be necessary. In this talk, we will show the application of a Graph Neural Network in the form of the object condensation algorithm applied on the hardware trigger level of the electromagnetic calorimeter at Belle II. We will show an implementation of the machine learning algorithm on FPGA level, which is necessary to guarantee a fast execution time, and the evaluation on possible Dark Sector models which would be inaccessible with the current trigger algorithm.

T 68.7 Wed 17:30 Geb. 30.34: LTI

Dealing with negatively weighted Events in DNN-based LHC Analyses — •Jörn Bach[1,2,3], Christian Schwanenberger[1,2], Peer Stelldinger[3], and Alexander Grohsjean[2] — [1]Deutsches Elektronen-Synchrotron (DESY), Hamburg, Germany — [2]Universität Hamburg, Hamburg, Germany — [3]HAW Hamburg, Hamburg, Germany

The recent decade has seen a growth of machine learning algorithms across all disciplines. In LHC physics, a multitude of applications have been tested and - in particular Deep Neural Networks (DNNs) - have been proven to be very effective in various usecases, for example in particle tagging or for separating signal from background in analyses. Since training data is primarily generated through Monte-Carlo (MC) simulation, specific challenges can emerge during DNN training due to partly negatively weighted samples. MC simulations produce negative event weights in the presence of destructive interference in the process or in the case of next-to-leading order simulations with an additive matching scheme. The negatively weighted training data impair the DNN convergence. Therefore, the current state of the art is to use reweighting methods that lead to consistently positive weights. However this alters the input distribution. We propose an alternative technique that is interpretable, computationally efficient and does not affect the input distribution. Furthermore, we show the method employed on a hypothetical search for a beyond the standard model heavy Higgs boson and discuss implications of negative weights throughout DNN based analyses.

T 68.8 Wed 17:45 Geb. 30.34: LTI

Automated Hyperparameter Optimization and Input Variable Selection for Neural Networks — •Erik Bachmann, Frank Siegert, Max Stange, and José Antolín Neumann — Technische Universität Dresden

Recent years have seen the widespread adoption of artificial neural networks in many high energy physics analyses to increase the sensitivity of measurements, largely replacing other multivariate techniques. The hyperparameters of the neural network, e.g. the number of hidden layers in a multilayer perceptron, are however often still chosen based on intuition and experience without any optimization.

Presented in this talk is a framework to perform automated hyperparameter optimization and input variable selection. During development, a special focus was put on generality and a flexible definition of metrics, allowing the framework to be configured for any supervised learning task. Being built on the Ray computing platform, the optimization can be scaled from a single computer to a multi-node HPC cluster.