Contribution ID: **53** Type: **Standard Talk**

Yggdrasil Conifer: Latency and resource-aware decision trees for faster FPGA inference at the LHC

Monday 25 September 2023 16:00 (15 minutes)

Decision Forests are fast and effective machine learning models for making real time predictions. In the context of the hardware triggers of the experiments at the Large Hadron Collider, DF inference is deployed on FPGA processors with sub-microsecond latency requirements. The FPGAs may be executing many algorithms, and many DFs, motivating resource-constrained inference. Using a jet tagging classification task representative of the trigger system, we optimise the DF training using Yggdrasil Decision Forests with fast estimation of resource and latency cost from the Conifer package for FPGA deployment. We use hyperparameter optimisation to select the optimal combination of DF architecture, feature augmentation, and FPGA compilation parameters to achieve optimal trade-off between model accuracy and inference cost under realistic LHC hardware constraints. We compare this Hardware/Software Codesign approach to other methods.

Co-authors: OLIVER, Andrew George; PFEIFER, Jan (Google); GUILLAME-BERT, Mathieu (Google); STOTZ, Richard (Google); SUMMERS, Sioni Paris (CERN); AARRESTAD, Thea (ETH Zurich (CH))

Presenter: OLIVER, Andrew George

Session Classification: Contributed Talks

Track Classification: Contributed Talks