



Contribution ID: 77

Type: Oral

Optimizing ANN-Based Triggering for BSM events with Knowledge Distillation

Tuesday, 12 March 2024 11:30 (20 minutes)

In recent years, the scope of applications for Machine Learning, particularly Artificial Neural Network algorithms, has experienced an exponential expansion. This surge in versatility has uncovered new and promising avenues for enhancing data analysis in experiments conducted at the Large Hadron Collider at CERN. The integration of these advanced techniques has demonstrated considerable potential for elevating the efficiency and efficacy of data processing in this experimental setting.

Nevertheless, one frequently overlooked aspect of utilizing Artificial Neural Networks (ANNs) revolves around the imperative of efficiently processing data for online applications. This becomes particularly crucial when exploring innovative methods for selecting intriguing events at the trigger level, as seen in the pursuit of Beyond Standard Model (BSM) events. The study delves into the potential of Autoencoders (AEs), an unbiased algorithm capable of event selection based on abnormality without relying on theoretical priors. However, the distinctive latency and energy constraints within the Level-1 Trigger domain necessitate tailored software development and deployment strategies. These strategies aim to optimize the utilization of on-site hardware, with a specific focus on Field-Programmable Gate Arrays (FPGAs).

This is why a technique called Knowledge Distillation (KD) is studied in this work. It consists in using a large and well trained “teacher”, like the aforementioned AE, to train a much smaller student model which can be easily implemented on an FPGA. The optimization of this distillation process involves exploring different aspects, such as the architecture of the student and the quantization of weights and biases, with a strategic approach that includes hyperparameter searches to find the best compromise between accuracy, latency and hardware footprint.

The strategy followed to distill the teacher model will be presented, together with consideration on the difference in performance of applying the quantization before or after the best student model has been found. Finally, a second way to perform KD will be introduced called co-training distillation which sees the teacher and the student models trained at the same time.

Significance

References

Experiment context, if any

CMS experiment

Primary author: LORUSSO, Marco (Universita Di Bologna (IT))

Presenter: LORUSSO, Marco (Universita Di Bologna (IT))

Session Classification: Track 1: Computing Technology for Physics Research

Track Classification: Track 1: Computing Technology for Physics Research