



Contribution ID: 69

Type: **Flashtalk with Poster**

Estimation of Machine Learning model uncertainty in particle physics event classifiers

Tuesday, 30 April 2024 17:19 (3 minutes)

Particle physics experiments entail the collection of large data samples of complex information. In order to produce and detect low probability processes of interest (signal), a huge number of particle collisions must be carried out. This type of experiments produces huge sets of observations where most of them are of no interest (background). For this reason, a mechanism able to differentiate rare signals buried in immense backgrounds is required. The use of Machine Learning algorithms for this task allows to efficiently process huge amounts of complex data, automate the classification of event categories and produce signal-enriched filtered datasets more suitable for subsequent physics study. Although the classification of large imbalanced datasets has been undertaken in the past, the generation of predictions with their corresponding uncertainties is quite infrequent. In particle physics, as well as in other scientific domains, point estimations are considered as an incomplete answer if uncertainties are not presented. As a benchmark, we present a real case study where we compare three methods that estimate the uncertainty of Machine Learning algorithms predictions in the identification of the production and decay of top-antitop quark pairs in collisions of protons at the Large Hadron Collider at CERN. Datasets of detailed simulations of the signal and background processes elaborated by the CMS experiment are used. Three different techniques that provide a way to quantify prediction uncertainties for classification algorithms are proposed and evaluated: dropout training in deep neural networks as approximate Bayesian inference, variance estimation across an ensemble of trained deep neural networks, and Probabilistic Random Forest. All of them exhibit an excellent discrimination power with a model uncertainty measure that turns out to be small, showing that the predictions are precise and robust.

Primary authors: Dr HERNÁNDEZ, José M. (Centro de Investigaciones Energéticas Medioambientales y Tecnológicas (CIEMAT)); Dr CÁRDENAS-MONTES, Miguel (Centro de Investigaciones Energéticas Medioambientales y Tecnológicas (CIEMAT)); VAZQUEZ ESCOBAR, Julia (Centro de Investigaciones Energéticas Medioambientales y Tecnológicas (CIEMAT))

Presenter: VAZQUEZ ESCOBAR, Julia (Centro de Investigaciones Energéticas Medioambientales y Tecnológicas (CIEMAT))

Session Classification: 3.3 Hardware acceleration, FPGAs & Uncertainty quantification

Track Classification: Session B