

21st International Conference on Computing in High Energy and Nuclear Physics (CHEP2015)



Contribution ID: 150

Type: oral presentation

The ATLAS Higgs Machine Learning Challenge

Thursday 16 April 2015 09:00 (15 minutes)

High Energy Physics has been using Machine Learning techniques (commonly known as Multivariate Analysis) since the 1990s with Artificial Neural Net and more recently with Boosted Decision Trees, Random Forest etc. Meanwhile, Machine Learning has become a full blown field of computer science. With the emergence of Big Data, data scientists are developing new Machine Learning algorithms to extract meaning from large heterogeneous data.

HEP has exciting and difficult problems like the extraction of the Higgs boson signal, and at the same time data scientists have advanced algorithms: the goal of the HiggsML project was to bring the two together by a “challenge”: participants from all over the world and any scientific background could compete online to obtain the best Higgs to tau tau signal significance on a set of ATLAS fully simulated Monte Carlo signal and background.

Instead of HEP physicists browsing through machine learning papers and trying to infer which new algorithms might be useful for HEP, then coding and tuning them, the challenge has brought realistic HEP data to the data scientists on the Kaggle platform, which is well known in the Machine Learning community.

The challenge has been organized by the ATLAS collaboration associated to data scientists, in partnership with the Paris Saclay Center for Data Science, CERN and Google.

The challenge ran from May to September 2014, drawing considerable attention. 1785 teams participated, making it the most popular challenge ever on the Kaggle platform. New Machine Learning techniques have been used by the participants with significantly better results than usual HEP tools.

This presentation has two parts: the first one describes how a HEP problem was simplified (not too much!) and wrapped up into an online challenge, the second what was learned from the challenge, in terms of new Machine Learning algorithms and techniques which could have an impact on future HEP analysis.

Author: ROUSSEAU, David (LAL-Orsay, FR)

Co-authors: KÉGL, Balázs (LAL-Orsay, FR); GERMAIN-RENAUD, Cecile (Laboratoire de Recherche en Informatique); ADAM BOURDARIOS, Claire (Laboratoire de l'Accelérateur Lineaire (FR)); COWAN, Glen (Royal Holloway, University of London); GUYON, Isabelle (Chalearn)

Presenter: COWAN, Glen (Royal Holloway, University of London)

Session Classification: Track 2 Session

Track Classification: Track2: Offline software