

# **HEPML workshop at NIPS14**

## **Rapport sur les contributions**

ID de Contribution: **0**

Type: **Non spécifié**

## Welcome

*samedi 13 décembre 2014 08:30 (15 minutes)*

**Orateur:** KÉGL, Balázs (LAL)

**Classification de Session:** Session 1

ID de Contribution: 1

Type: **Non spécifié**

## HEP&ML and the HiggsML challenge

*samedi 13 décembre 2014 08:45 (35 minutes)*

We first describe the HiggsML challenge (the problem of optimizing classifiers for discovery significance, the setup of the challenge, the results, and some analysis of the outcome). In the second part we outline some of the application themes of machine learning in high-energy physics.

**Orateur:** KÉGL, Balázs (LAL)

**Classification de Session:** Session 1

ID de Contribution: 2

Type: **Non spécifié**

## **Embedding ML in Classical Statistical tests used in HEP (invited talk)**

*samedi 13 décembre 2014 09:20 (40 minutes)*

I will review the ways that machine learning is typically used in particle physics, some recent advancements, and future directions. In particular, I will focus on the integration of machine learning and classical statistical procedures. These considerations motivate a novel construction that is a hybrid of machine learning algorithms and more traditional likelihood methods.

**Orateur:** CRANMER, Kyle (New York University)

**Classification de Session:** Session 1

ID de Contribution: 3

Type: **Non spécifié**

## **Presentation of the winner of the HiggsML challenge**

*samedi 13 décembre 2014 10:30 (20 minutes)*

We describe the winning solution of the HiggsML challenge, the issues related to the evaluation metric and reliable assessment of model performance. Finally, we take a stab at predicting how to achieve larger improvements.

**Orateur:** MELIS, Gábor

**Classification de Session:** Session 2

ID de Contribution: 4

Type: **Non spécifié**

## Presentation of the runner up of the HiggsML challenge

*samedi 13 décembre 2014 10:50 (20 minutes)*

High Energy Physics provides a challenging data domain with data that is highly structured, but also very noisy. I will present what I have learned analyzing this data for the HiggsML challenge, focusing on methods that are able to effectively search through a high dimensional model space while also achieving good statistical efficiency. In addition, I will discuss the role of the physicist in modelling this type of data, and I will talk about robustly applying our findings to real (not simulated) HEP data.

**Orateur:** SALIMANS, Tim

**Classification de Session:** Session 2

ID de Contribution: 5

Type: **Non spécifié**

## **Presentation of the winner of the HEP meets ML prize**

*samedi 13 décembre 2014 11:10 (20 minutes)*

In this talk, I will describe how we use principle of gradient boosting method to construct simple and effective regression trees functions for Higgs Boson detection. We take a functional space optimization framework that jointly optimize the training objective and simplicity of functions learnt. I talk about how the objective could be clearly related to the tree searching, pruning and leave weight estimation. Finally I will discuss how the framework could be modularized, to provide interface for adding physics domain knowledge into the learning algorithm.

**Orateur:** CHEN, Tianqi

**Classification de Session:** Session 2

ID de Contribution: 6

Type: **Non spécifié**

## Real time data analysis at the LHC : present and future

*samedi 13 décembre 2014 11:30 (40 minutes)*

The large hadron collider (LHC), which collides protons at an energy of 14 TeV (for non-physicists, each beam of protons carries roughly the energy of a TGV train going at full speed), produces hundreds of exabytes of data per year, making it one of the largest sources of data in the world today. At present it is not possible to even transfer most of this data from the four main particle detectors at the LHC to “offline” data facilities, much less to permanently store it for future processing. For this reason the LHC detectors are equipped with real-time analysis systems, called triggers, which process this volume of data and select the most interesting proton-proton collisions. The LHC experiment triggers reduce the data produced by the LHC by between 1/1000 and 1/10000, to tens of petabytes per year, allowing its economical storage and further analysis. The bulk of this data-reduction is performed by custom electronics which ignores most of the data in its decision making, and is therefore unable to exploit the most powerful known data analysis strategies developed by e.g. the machine learning community. In this talk I will cover the present status of real-time data analysis at the LHC, before explaining why the future upgrades of the LHC experiments will increase the volume of data which can be sent off the detector and into off-the-shelf data processing facilities (such as CPU or GPU farms) to tens of exabytes per year. This development will simultaneously enable a vast expansion of the physics programme of the LHC’s detectors, and make it mandatory to develop and implement a new generation of real-time multivariate analysis tools in order to fully exploit this new potential of the LHC. I will explain what work is ongoing in this direction and hopefully motivate why more effort is needed in the coming years.

**Orateur:** GLIGOROV, Vava**Classification de Session:** Session 2



ID de Contribution: 7

Type: **Non spécifié**

## **Machine Learning for Ultra-High-Energy Physics (invited talk)**

*samedi 13 décembre 2014 15:00 (40 minutes)*

I will describe the computational and machine learning challenges of the CRAYFIS project: a distributed cosmic ray telescope consisting of consumer smartphones and geared for the detection of ultra-high-energy cosmic rays. For more info: <http://crayfis.ps.uci.edu/>

**Orateur:** WHITESON, Daniel

**Classification de Session:** Session 3

ID de Contribution: 8

Type: **Non spécifié**

## Consistent optimization of AMS by logistic loss minimization

*samedi 13 décembre 2014 16:00 (20 minutes)*

In this paper, we theoretically justify an approach popular among participants of the Higgs Boson Machine Learning Challenge to optimize approximate median significance (AMS). The approach is based on the following two-stage procedure. First, a real-valued function is learned by minimizing a surrogate loss for binary classification, such as logistic loss, on the training sample. Then, a threshold is tuned on a separate validation sample, by direct optimization of AMS. We show that the regret of the resulting (thresholded) classifier measured with respect to the squared AMS, is upperbounded by the regret of the underlying real-valued function measured with respect to the logistic loss. Hence, we prove that minimizing logistic surrogate is a consistent method of optimizing AMS.

**Orateur:** KOTLOWSKI, Wojciech**Classification de Session:** Session 3

ID de Contribution: 9

Type: **Non spécifié**

## Weighted Classification Cascades for Optimizing Discovery Significance in the HiggsML Challenge

*samedi 13 décembre 2014 15:40 (20 minutes)*

We introduce a minorization-maximization approach to optimizing common measures of discovery significance in high energy physics. The approach alternates between solving a weighted binary classification problem and updating class weights in a simple, closed-form manner. Moreover, an argument based on convex duality shows that an improvement in weighted classification error on any round yields a commensurate improvement in discovery significance. We complement our derivation with experimental results from the 2014 Higgs boson machine learning challenge.

**Orateur:** MACKEY, Lester

**Classification de Session:** Session 3

ID de Contribution: 10

Type: Non spécifié

## Ensemble of maximied Weighted AUC models for the maximization of the median discovery significance

*samedi 13 décembre 2014 17:00 (20 minutes)*

From May 12th 2014 to September 15th 2014 took place the Higgs Boson Machine Learning Challenge. Its goal was to explore machine learning methods to improve the discovery significance of the ATLAS experiment. This talk describes the preprocessing, training and results of our model, that finished in 9th position among the solutions of 1785 teams.

**Orateur:** MORALES, Roberto Diaz (University Carlos III de Madrid)

**Classification de Session:** Session 4

ID de Contribution: 11

Type: **Non spécifié**

## Deep Learning In High-Energy Physics (invited talk)

*samedi 13 décembre 2014 17:20 (40 minutes)*

We will provide a brief overview of the challenges and opportunities facing machine learning in the natural sciences, from physics to biology, and then focus on the application of deep learning methods to problems in high-energy physics. In particular we will describe the results obtained on three different problems (Higgs boson detection, Supersymmetry, and Higgs boson decay).

**Orateur:** BALDI, Pierre**Classification de Session:** Session 4

ID de Contribution: **12**

Type: **Non spécifié**

## **Panel discussion**

*samedi 13 décembre 2014 18:00 (30 minutes)*

**Classification de Session:** Session 4