CDS-Paris Saclay pitching day 2016 : Tracking Machine Learning Challenge

Authors : David Rousseau (LAL-Orsay), Cecile Germain (LRI-Orsay), Isabelle Guyon (LRI and ChaLearn),

The Large Hadron Collider at CERN, where the Higgs boson has been discovered, is poised for a major upgrade for possible discovery of new particles, super-symmetric particles, dark matter or signs of extra-dimensions of space. The increase in yearly number of proton collisions recorded comes at a cost of a large increase of the recorded event complexity.

For each collision, about 10.000 space tracks (helicoidal trajectories originating approximately from the origin), will leave about 10 precise 3D measurements (see picture next page). The core pattern recognition tracking task is to associate the 100.000 3D measurements into tracks. Preliminary studies show that traditional algorithms suffer from a combinatorial explosion of the CPU time.

To reach out for Computer Science specialists, a Tracking Machine Learning challenge (trackML) is being set up for 2017, building on the experience of the successful Higgs Machine Learning challenge in 2014, which associates physicists from the 4 LHC experiments (ALICE, ATLAS, CMS and LHCb) with Computer Scientists (some from CDS).

A dataset consisting of a simulation of a typical full Silicon LHC experiments is being created, listing for each event the measured 3D points, and the list of 3D points associated to a true track. The data set is large to allow the training of data hungry Machine Learning methods. The orders of magnitude are : one million event, 10 billion tracks, 1 terabyte. Typical CPU time spent by traditional algorithms (already very optimised by arch-experts) is 100s per event, which should be reduced by one order of magnitude.

The emphasis is on discovering innovative approaches, rather than hyper-optimising known approaches. What is asked to CDS :

- 2-months of an engineer (preferably from CDS core) to finalise the challenge starting-kits
- use the CDS channels to advertise the challenge
- master internship for post-challenge analysis
- build post-challenge collaboration with CDS ML scientists on innovative approaches to the tracking problem as revealed by the challenge.
- possibly, collaboration on the visualization