TrackML : a High Energy Physics particle tracking challenge*

Paolo Calafiura, Steven Farrell, Heather Gray
Lawrence Berkeley National Laboratory
Berkeley, Ca, USA

Jean-Roch Vlimant
CalTech
Pasadena, CA, USA

Vincenzo Innocente, Andreas Salzburger
CERN
Geneva, Switzerland

Sabrina Amrouche, Tobias Golling, Moritz Kiehn
Université de Genève
Geneva, Switzerland

Victor Estrade, Cécile Germaint, Isabelle Guyon
LRI
Orsay, France

Ed Moyse
University of Massachussetts
Boston, MA, USA

David Rousseau, Yetkin Yilmaz
LAL
Orsay, France

Vladimir Vava Gligorov
LPNHE
Paris, France

Mikhail Hushchyn, Andrey Ustyuzhanin
Yandex, HSE
Moscou, Russia

To attain its ultimate discovery goals, the luminosity of the Large Hadron Collider at CERN will increase so the amount of additional collisions will reach a level of 200 interaction per bunch crossing, a factor 7 w.r.t the current (2017) luminosity. This will be a challenge for the ATLAS and CMS experiments, in particular for track reconstruction algorithms. In terms of software, the increased combinatorial complexity will have to harnessed without any increase in budget. To engage the Computer Science community to contribute new ideas, we organized a Tracking Machine Learning challenge (TrackML) running on the Kaggle platform from March to June 2018, building on the experience of the successful Higgs Machine Learning challenge in 2014. The data were generated using [ACTS] (http://acts.web.cern.ch/ACTS/latest/doc/index.html), an open source accurate tracking simulator, featuring a typical all silicon LHC tracking detector, with 10 layers of cylinders and disks. Simulated physics events (Pythia ttbar) overlaid with 200 additional collisions yield typically 10000 tracks (100000 hits) per event (see display in Fig 1).

The task to be performed by participants in the challenge is the pattern recognition: associate the hits to tracks corresponding to the original charged particles. The participants are given 9000 events (including truth information) to train their algorithm, while the evaluation is run on 125 other events.

The "Accuracy phase" is running 1st of May to mid August 2018 on Kaggle platform. The score used to rank the candidates is the fraction of hits correctly assigned, with a weighting mechanism to favor higher momentum tracks and hits in the innermost and outermost detector layers. In this challenge, there is no CPU constraint.

contact author : David Rousseau trackml.contact@gmail.com

The "Throughput phase" is running 1st of July to end October on Codalab platform. It has a strong incentive on the CPU time of the evaluation (i.e. track reconstruction).

The emphasis of the challenge is to explore innovative Machine Learning approaches, rather than hyper-optimising known combinatorial approaches. In preliminary discussions with the ML community, Convolutional Neural Network, LSTM, Deep Neural Nets, Monte Carlo Tree Search, geometric Deep Learning have been mentioned.

In this talk, the first lessons from the Accuracy phase of the challenge will be discussed. At the time of writing (18th June 2018), 430 teams are participating and the score reached is already 70%. Some indication on the Throughput phase (which will just have been completed) will also be discussed. What algorithms have emerged and are the most promising ? What are their detailed performances beyond the score ?

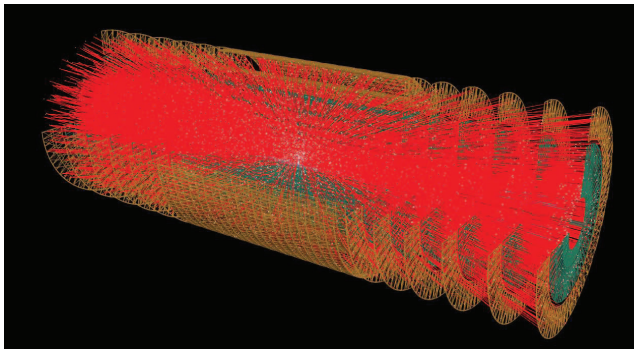

Fig. 1. Display of one event with truth trajectories (red) overlaid to measured points (white)

978-1-5386-9156-4/18/$31.00 ©2018 IEEE
DOI 10.1109/eScience.2018.00088