

TrackML : the High Energy Physics Tracking Challenge

David ROUSSEAU (LAL-Orsay)

Seminar EP-IT Data Science at CERN, Wed 7th March 2018

Short transcript of Q/A session by DR

Q do you accept non ML solution ? A: of course yes

Q on the weighting, what about giving more weights to track in busy environment ? A: we did not want to complexify the scoring too much. We will be able to study offline the detailed performance of the algorithm, efficiency, fake rate, jet core etc....

Q is there a penalty for fake rate ? Can you not list many solutions ? A: no, because one point can only be assigned to one track.

Q this seems to need a lot of domain knowledge ? A: good point, we'll alleviate this by a good documentation, and releasing pieces of software encapsulating it, like the error parameterisation. Note : this seminar was meant for a HEP audience; another version for Computer Scientists would not go as deep in some details.

R: Markus Elsing: good that the dataset will stay.

R: Jean-Roch Vlimant: we accept sponsors from industry, but also academic. If you're interested, get in touch, your institution will have its logo on the web page and poster

Q: you will not release the field map ? : A: right, too much domain knowledge needed to use it properly, and slow. We think the shape of the tracks can be rebuilt from the large amount of data we release.

Q: 15k€ enough to motivate ML scientist to invest time and acquiring specific knowledge ? A: money is not the main driver for ML scientists, being top ranked in a kaggle challenge is a bigger driver. Plus the investment in ML techniques is probably reusable.

Q: will there be a simplified dataset ? A : indeed, we already have events with just 3 tracks, we also have code snippet that given an event, can reduce it to N tracks.