



Highlights from the ML session

PhysTev 2019

Les Houches - 28 June 2019



Riccardo Torre
CERN & INFN Genova



on behalf of the

ML Convener: Sascha Caron

ML activities

- ML session in LH for the first time: very successful!
- Clear program of things to do (some of which are proof of principle)
- Just need to gather interest from different communities and manpower

3 big projects

1. Event generation with Generative models
2. Database of Networks (regression, classification)
3. Anomaly detection

Many interested people

Melissa v. Beekveld, Wolfgang Woltenberger, Richard Ruiz, Sydney Otten, Andrea Coccaro, Roberto Ruiz, Riccardo Torre, Sascha Caron, Sezen Sekmen, Sanmay Ganguly, Giovanni Zevi, Bob Stienen, Maurizio Pierini, Sabine Kraml, Jan Heisig, Luca Silvestrini, Seung Lee, ...

Event generation with generative models

Topics we aim to study

1. Interpolation and extrapolation properties of generative models and transfer learning
2. Comparing different architectures and new ideas from the data science community
3. Detector simulation and showers

One of the most challenging topics, but also the one which could change the game in view of the HL-LHC challenges

Organization:

Mailing list: [lh2019-ML-generative-models](#) (static e-group with self subscription)

GitLab repo under a [GitLab group](#)

Meetings organized every 2(?) weeks through a dedicated [Indico page](#)

Database of Networks (regression, classification)

To do list:

1. Library of regression and classification networks
2. Keras/Tensorflow and Scikit-learn to start
3. Library of Training data, zenodo group available

Train regression on the following: cross sections, branching ratios, efficiencies, LLPs, likelihoods, loop integrals, EFTs, ...

Scope:

- MachineLearning.Likelihood(Model, Modelparameters)
- MachineLearning.CS(Process, Parameters)
- ...

Organization:

Mailing list: [lh2019-ML-regression-models](#) (static e-group with self subscription):

GitLab repo under a [GitLab group](#)

Meetings organized every 2(?) weeks through a dedicated [Indico page](#)

Data stored and shared through a [Zenodo group](#)

Anomaly detection

Three different approaches

1. Finding events which are not in the training data (real outliers, example two electron (charge=-2) event)
2. Finding over-densities (e.g. more events than expected by SM, example is $Z' = 500$ GeV)
3. Generally supervised approaches (train on ensemble of BSM theories)

Idea: Maybe publish a network for new physics

Project:

- Make „Standard Model“ and toy “data sets“ (need BSM signals!) → Paper
- Organise a challenge to compare different approaches for anomaly detection

Organization (joined with the DarkMachine unsupervised project):

Mailing list: [darkmachines-unsupervised](#) (static e-group with self subscription):

[Google doc](#) with minutes of the DarkMachines meetings

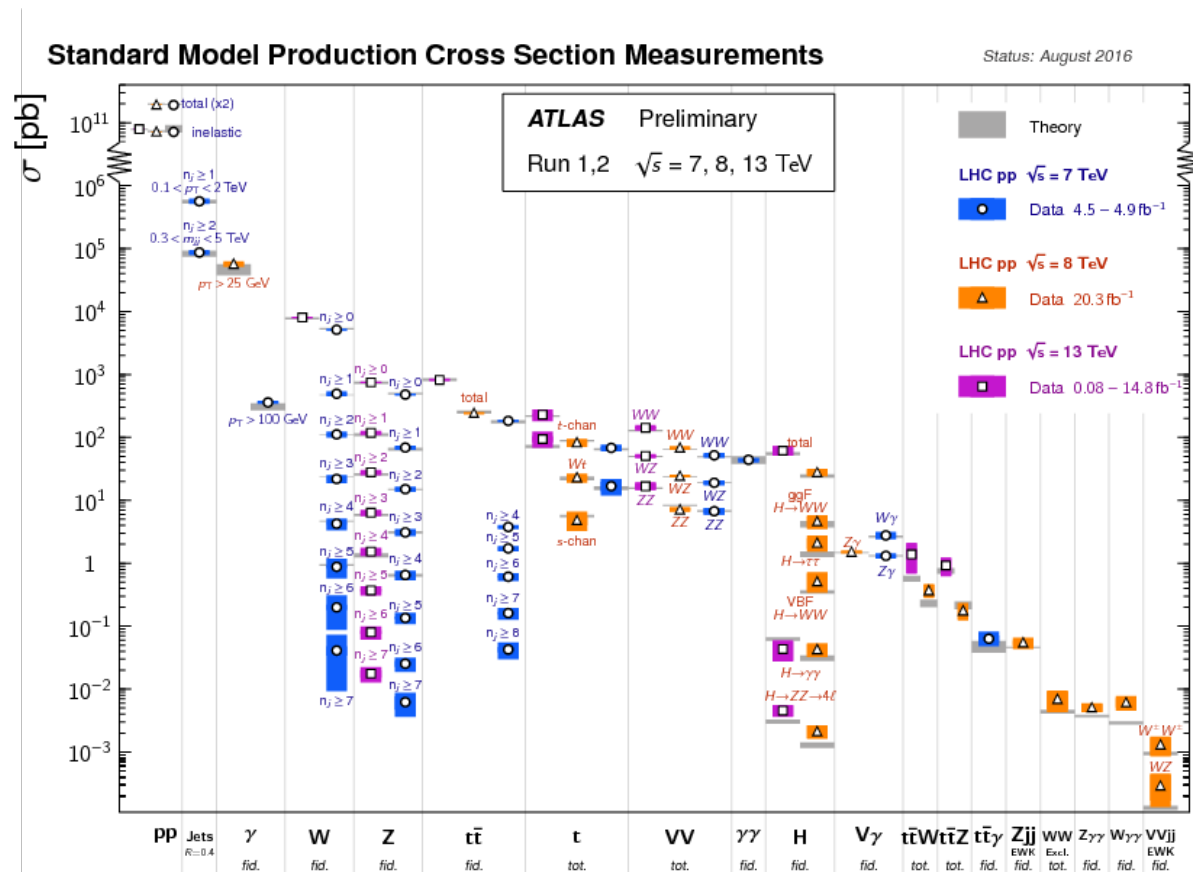
[Google doc](#) with description of the datasets (Overleaf available soon)

[GitHub repo](#) with code for the generation of events:

„Standard Model“ datasets

Dataset already available ([Google doc](#) dataset explanation)

- O(100) Million events, corresponding to 30 fb⁻¹ (LHC)
- All SM processes (~20), "trigger cuts", OR of jets>X, MET>Y, leptonPt>Z
- Text file (csv) including 4-vectors high-pt objects, weights, etc.
- Aim to release this with "Dataset & anomaly detection paper"



Conclusion

Join us and tell us what Machine Learning can do for you!



Thank you to all ML 'professionals' and 'enthusiasts' in LH and wish to see again a ML session in the next edition!