Contribution ID: **906** Type: **Talk**

Automated selection of particle-jet features for data analysis inHigh Energy Physics experiments

Tuesday, July 28, 2020 5:10 PM (20 minutes)

In high-energy physics experiments, the sensitivity of selection-based analyses critically depends on which observable quantities are taken into consideration and which ones are discarded as considered least important. In this process, scientists are usually guided by their cultural background and by literature.
Yet simple and powerful, this approach may be sub-optimal when machine learning strategies are envisaged and potentially all features are usable. On the other hand, training multivariate algorithms with all available features is often impossible, due to lack of calibration or computing power limitations. How to robustly choose the set of observables to use in a modern high-energy physics analysis?
We show here that it is possible to rank the relative importance of all available features in an automated fashion by engineering a fast and powerful classification model.
Features are sorted with the Random Forest algorithm, then selected as input quantities for a Deep Learning Neural Network. We make it explicit the relation between Random Forest importance ranking and signal-to-background ratio increase, varying the number of features to feed the Neural Network with. We benchmark our procedure with the case of highly boosted di-jet resonances decaying to two $b\bar{b}$ quarks, to be selected against an overwhelming QCD background. Promising results from Monte Carlo simulation with HEP pseudo-detectors are shown.

Secondary track (number)

14

Primary authors: Mr DI LUCA, Andrea (Universita degli Studi di Trento and INFN (IT)); FOLLEGA, Francesco Maria (Universita degli Studi di Trento and INFN (IT)); Dr CRISTOFORETTI, Marco (Universita degli Studi di Trento e INFN (IT)); IUPPA, Roberto (Universita degli Studi di Trento and INFN (IT))

Presenter: Mr DI LUCA, Andrea (Universita degli Studi di Trento and INFN (IT))

Session Classification: Computing and Data Handling

Track Classification: 14. Computing and Data Handling