

Thoughts on using python, numpy, and scikit-learn for HEP analysis

Tuesday 10 July 2018 16:40 (20 minutes)

The HEP community has voted strongly with its feet to adopt ROOT as the current de facto analysis toolkit. It is used to write out and store our RAW data, our reconstructed data, and to drive our analysis. Almost all modern data models in particle physics are written in ROOT. New tools in industry have are making appearance in particle physics analysis, however, driven by the massive interest in Machine Learning. Further, datasets in industry rival or exceed those in particle physics. Given the large number of people outside HEP devoting time to optimizing these tools, it makes a lot of sense for HEP to adopt what it can, and devote resources to things unique to HEP. There are several external toolkits –the most popular are based on the R language and Python. Python seems to have the most interest within the HEP community, and parts of the community are devoting serious resources to its incorporation. This work discusses some of the high-level differences between the ROOT approach and the numpy/scikit-learn approach, the technical details driving them, along with some ideas of what could be adopted in the long run for use in our community. It is based on work implementing an analysis in the ATLAS experiment.

Primary author: WATTS, Gordon (University of Washington (US))

Presenter: WATTS, Gordon (University of Washington (US))

Session Classification: Posters

Track Classification: Track 6 –Machine learning and physics analysis