10th International Workshop on Boosted Object Phenomenology, Reconstruction and Searches in HEP (BOOST 2018)

Contribution ID: 33

Type: Talk

JUNIPR: a Framework for Unsupervised Machine Learning in Particle Physics

Thursday 19 July 2018 11:35 (25 minutes)

In applications of machine learning to particle physics, there is a persistent tension between interpretability and performance. In this paper, this tension is allayed by introducing a novel framework for unsupervised machine learning in particle physics, in which the neural network architecture is built as a scaffolding around a leading-order description of the physics under study. This approach not only reduces the complexity and increases the efficiency of the machine learning models it inspires, but it can also lead to interpretable models, as opposed to deep black boxes. In particular, we present JUNIPR, a framework for studying "Jets using UNsupervised Interpretable PRobability models". Within this framework, we use a deep neural network to construct a probability model for jet physics, i.e. a function that computes the relative differential cross section of individual jets in a sample. Jets at colliders are defined by sequentially clustering final-state particles together; the resulting tree structure augmenting each jet provides the scaffolding for the deep neural network architecture in the JUNIPR framework, enabling predictions that are easy to visualize and interpret. Although neural network architectures in the JUNIPR framework leverage a user-specified sequential tree structure for jets, training such models is unsupervised and unrestricted: the network could decide that the chosen tree structure has little to do with the training data. To test this, both physically-motivated and unphysical trees are considered. JUNIPR-based probability models are shown to perform powerful discrimination through the statistically optimal likelihood-ratio test, and to permit visualizations of this discrimination power at each branching in a jet's tree. Samples from such probability models can also be drawn, providing a data-driven Monte Carlo generator for computing arbitrary physical observables. It is further demonstrated that JUNIPRbased models can efficiently re-weight jets from one (e.g. simulated) data set to agree with jets from another (e.g. experimental) data set. We elaborate significantly on the inner workings of our approach in an attempt to place a foundation for future work in this novel direction.

Authors: SCHWARTZ, Matthew; Dr ANDREASSEN, Anders (Harvard University); Dr FEIGE, Ilya (ASI Data Science); FRYE, Christopher (Harvard)

Presenter: FRYE, Christopher (Harvard)

Session Classification: Machine Learning