



Contribution ID: 10

Type: not specified

Explainable AI for ML Jet Taggers

Thursday, July 8, 2021 4:40 PM (20 minutes)

A framework is presented to extract and understand decision-making information from a deep neural network classifier of jet substructure tagging techniques. The general method studied is to provide expert variables that augment inputs (“eXpert AUGmented” variables, or XAUG variables), then apply layerwise relevance propagation (LRP) to networks that have been provided XAUG variables and those that have not. The XAUG variables are concatenated to the classifier’s intermediate input to the final layer. The results show that XAUG variables can be used to interpret classifier behavior, increase discrimination ability when combined with low-level features, and in some cases capture the behavior of the classifier completely. The LRP technique can be used to find relevant information the network is using, and when combined with the XAUG variables, can be used to rank features, allowing one to find a reduced set of features that capture a majority of network performance. These identified XAUG variables can also be added to low-level networks as a guide to improve performance.

Academic Rank

Postdoctoral researcher

Affiliation

University at Buffalo

Primary authors: MCLEAN, Christine Angela (SUNY Buffalo); AGARWAL, Garvita (The State University of New York SUNY (US)); HAY, Lauren Meryl (SUNY Buffalo); RAPPOCCIO, Salvatore (The State University of New York SUNY (US)); MORRIS, Margaret (University at Buffalo); IASHVILI, Ia (The State University of New York SUNY (US))

Presenter: MCLEAN, Christine Angela (SUNY Buffalo)

Session Classification: Interpretability, Robustness, and Uncertainties