

Explainable AI for ML jet taggers using expert variables and layerwise relevance propagation

Tuesday, July 13, 2021 5:15 PM (15 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 or decision. 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 part of the network performance. These XAUG variables can also be added to low-level networks as a guide to improve performance.

*This work was supported under NSF Grants PHY-1806573, PHY-1719690 and PHY-1652066. Computations were performed at the Center for Computational Research at the University at Buffalo.

Are you are a member of the APS Division of Particles and Fields?

No

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

Presenter: AGARWAL, Garvita (The State University of New York SUNY (US))

Session Classification: Computation, Machine Learning, and AI

Track Classification: Computation, Machine Learning, and AI