

Jet classification in t-tbar decays of heavy BSM resonances in ATLAS using ML techniques

The search for heavy resonances with diverse mass and width values is the ultimate goal of the ATLAS physics group where we are developing this study. The analysis selects events where one of the top quarks decays hadronically and the other one semileptonically. The reconstruction of the different components of the decay allows us to build an invariant mass distribution which would show some excess around the mass of the resonance if it exists.

The improvement in the reconstruction of the ttbar decay is the main purpose of our study. Therefore we are replacing a traditional Chi2 technique based on intermediate invariant mass values and transverse momentum balance with a novel approach making use of Machine Learning algorithms with any variable that might contribute to a classification of the jets of the event.

After a selection of appropriate events we store a set of variables for each jet in the event including kinematic variables, angular distances, invariant masses and tagging variables. Our classification involves 4 different classes: b jet from hadronic decay; b jet from semileptonic decay; jets coming from the desintegration of the W boson of the hadronic top quark decay and any other jet present in the event but not related to the resonance decay process. Given the fact that we prefer to characterise the jets in a non-binary way, our problem would be termed as multiclass multilabel. We have used several algorithms capable of providing classification for that kind of problem, namely Deep Neural Networks, Random Forest and eXtreme Gradient Boosting in order to train our dataset of jets.

As a first step we study the relevance of the variables with the help of the Permutation Importance method (DNN) and the Boruta method (RF, XGB) and discard those that do not contribute to any algorithm and resonance mass. Then we proceed with the optimization of the hyperparameters for each kind of algorithm. The training is followed by an assignment of jet roles within each event that gives our final reconstruction efficiency.

Preliminary results show a reconstruction efficiency at $M(Z')=1\text{TeV}$ of around 77% of the events where all jets were matched (reference jets) at Monte Carlo level for XGB and slightly smaller efficiencies for RF and DNN. The traditional Chi2 method gives values just above 70% for the same mass value. Current work involves the addition of the Chi2 assignment to the set of variables and the use of a clustering algorithm to allow further optimization of the ML classifying algorithms.

Authors: MARTINEZ DE LEJARZA SAMPER, Jorge Juan (Univ. of Valencia and CSIC (ES)); LOZANO BAHILLO, Julio (Univ. of Valencia and CSIC (ES)); SALT, Jose (Instituto de Fisica Corpuscular (IFIC) - Universidad de Valencia)

Presenter: MARTINEZ DE LEJARZA SAMPER, Jorge Juan (Univ. of Valencia and CSIC (ES))

Session Classification: COMCHA

Track Classification: COMCHA