Identifying hadronically-decaying vector bosons and top quarks in ATLAS

Oliver Majerský on behalf of the ATLAS collaboration

Faculty of Mathematics, Physics and Informatics, Comenius University in Bratislava, Slovakia

E-mail: oliver.majersky@cern.ch

Abstract. Hadronic decays of vector bosons and top quarks are increasingly important to the ATLAS physics program, both in measurements of the Standard Model and searches for new physics. At high energies, these decays are collimated into a single overlapping region of energy deposits in the detector, referred to as a jet. However, vector bosons and top quarks are hidden under an enormous background of other processes producing jets. The ATLAS experiment has employed boosted decision trees and deep neural networks to the challenging task of identifying hadronically-decaying vector bosons and top quarks and rejecting other jet backgrounds. These discriminants are becoming increasingly complex and using more advanced machine learning techniques. The methods currently used to tag these objects are described. In order to improve the tagger performance on the signal efficiency and background rejection, new in-situ techniques are applied, thus directly evaluating the agreement between data and simulation after applying an arbitrarily complex classifier. The precision obtained by applying the in-situ techniques is presented.

ATL-SOFT-PROC-2019-001
16 May 2019

1. Introduction

The high centre-of-mass energy of 13 TeV at the Large Hadron Collider (LHC) [1] in Run 2 allows the physics programme at the ATLAS experiment [2] to probe processes with highly boosted massive decaying particles in the final state, such as vector bosons and top quarks. Of particular interest are hadronic decays of these particles, which are most abundant due to the highest branching fraction, but are mimicked by the dominant multijet background consisting of mostly light-quark/gluon (background) jets. It is thus important to identify the hadronically-decaying top quarks and W bosons (signal) and reject the multijet background. ATLAS has performed a hereby reported study [3] of algorithms for identification of these particles with focus on the application of machine learning (ML).

The standard approach to reconstruction of jets at the ATLAS experiment is to use the anti-k_t [4] sequential algorithm with the topological calorimeter clusters [5] as inputs. To capture the collimated radiation from the decay of top quarks or W bosons, the jets are reconstructed with a large radius parameter $R = 1.0$ and subsequently, trimming [6] grooming technique is applied to these jets to remove low-p_T constituents typically from pile-up and underlying event processes. The signal and background jets differ in terms of their inner radiation patterns. Since the inception of the boosted jet identification (*tagging*), many observables (high-level features) have been invented which quantify differences in substructure between signal and background jets. ATLAS has employed taggers that perform a selection typically on two observables such as the jet mass and another variable characterising the "pronginess" of the radiation in the

jet [7, 8], to suppress background jets. In this study [3], ATLAS extends the use of high-level features by combining many observables in a multivariate classifier and examines an approach of using deep neural networks (DNNs) with low-level features of jet topoclusters.

2. Optimisation of the ML algorithms using high-level features

Two approaches of combining high-level features are examined – boosted decision trees (BDTs) and DNNs, with the goal to determine if one of the ML algorithms is better suited to exploit the differences of the features and their correlations for signal and background jets. Both ML techniques are trained on jets with mass $m > 40\,\text{GeV}$ and at least three constituents. The jets used for both training and testing are from $p_{\text{T}}^{\text{true}}$ range of $[200, 2000]\,\text{GeV}$ for W-boson tagging and $[350, 2000]\,\text{GeV}$ for top-quark tagging (in simulation, the true transverse momentum of the jet $p_{\text{T}}^{\text{true}}$ is determined according to an anti-k_{t} $R = 1.0$ ungroomed jet built from stable particles with lifetimes greater than 10 ps, excluding muons and neutrinos, associated with the reconstructed trimmed anti-k_{t} $R = 1.0$ jet). Finally, the training set is reweighted such that the jet $p_{\text{T}}^{\text{true}}$ spectrum is constant across the $p_{\text{T}}^{\text{true}}$ range used, to avoid biasing the training due to kinematics. The testing set used for performance comparisons is made with $p_{\text{T}}^{\text{true}}$ distribution for signal jets weighted to match that of the multijet background sample.

The architecture used for the DNN is a fully-connected feed-forward network, with 4 (5) hidden layers for top-quark tagging (W-boson tagging), with up to 18 nodes in layer using rectified linear unit [9] activation function and sigmoid [9] activation function for the output layer. The chosen hyper-parameter were determined in a dedicated optimisation. The Adam [10] optimiser is used and the training is performed using software packages Keras [11] and Theano [12].

The BDT uses the gradient boosting technique with bagging [13], with 500 trees used in forest, and depth of individual trees up to 20, determined in a dedicated hyper-parameter optimisation. The training is performed using the TMVA [14] software package.

In addition to the optimisation of the hyper-parameters of the DNN and BDT, an optimisation of the choice of the input observables has been performed with the background rejection of the resulting ML classifier on the testing set being the figure of merit to maximise. For the BDT, a sequential addition procedure is used. At each step a single observable is added and the BDT is retrained. All combinations are tested and the observable giving the largest rise in background rejection is accepted. The procedure is stopped once the background rejection increase is compatible with statistical uncertainty of the testing set.

The input observables used in the DNN classifier are chosen by testing various groups of observables and picking the group with the largest background rejection. The observables are not added iteratively as in the case of BDT due to time requirements of training of a large number of networks. The groups differ based on selection of observables according to their dependence on the momentum scale of the jet substructure objects, the features of the substructure described, and their dependence on other substructure observables. The DNN classifier with the best performance uses 12 observables for W-boson tagging and 13 observables for top-quark tagging.

3. Optimisation of the topocluster-based deep neural network

An alternative approach to using ML techniques with high-level features is presented here, a DNN for top-quark tagging referred to as TopoDNN, which uses the vectors of 10 highest-p_{T} topoclusters of the jet in the (p_T, η, ϕ) representation as input variables [15, 3]. As a pre-processing step, the p_{T} of the clusters is normalised to bring the scale of all the input network features within the same magnitude. Secondly, the (η, ϕ) coordinates are transformed using a procedure which involves translation, rotation and a flip based on the assumed three-body topology of the top-quark decay. For jets with less than 10 topoclusters, the remaining cluster input variables are taken as null vectors.

For the training and testing set, jets with $p_T = [450, 2000]$ GeV where considered. The signal jet p_{T} distribution was modified to match the p_{T} distribution of the multijet background by using random sub-sampling on the signal jet sample to avoid biasing the DNN due to p_{T} distribution differences between the signal and the background jets.

The transformed input variables are fed to a DNN with four hidden layers composed of up to 300 nodes. The architecture as well as other hyper-parameters were tuned in a dedicated optimisation [15]. For the training of the DNN, the same software was used as for the high-level feature-based DNN taggers.

4. Performance comparison

The comparison of performance of the individual tagging techniques is shown in Figures 1 and 2 for top-quark and W-boson tagging, respectively. In these figures, simpler taggers such as those employing cuts on two sub-structure observables are compared with ML-based taggers, where the latter clearly provide higher background rejection. Both BDT and DNN using high-level features show very similar performance, hinting that for the observables and DNN architectures investigated, the DNN is not able to extract more information compared to the BDT. Finally, the difference in performance between ML-based taggers and simpler taggers is much larger for top-tagging as opposed to W-boson tagging. This can be attributed to the richer structure of the top decay leading to more distinct features in contrast to jets from multijet background.

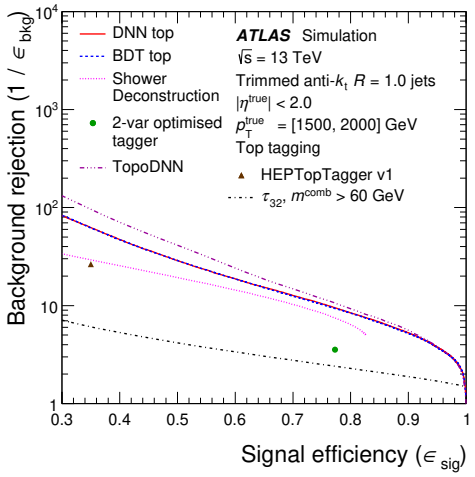

Figure 1. A comparison of the background rejection vs signal efficiency for top-quark tagging [3]. For a given signal efficiency, the higher the background rejection, the better.

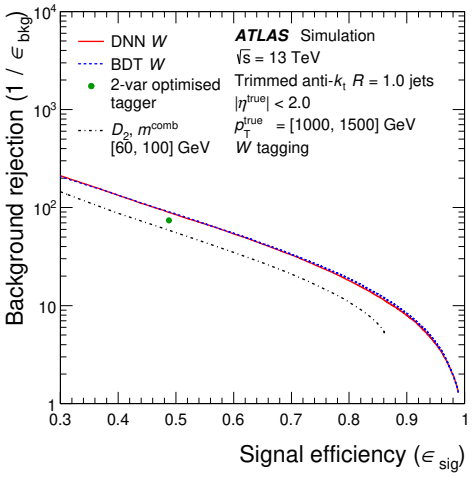

Figure 2. A comparison of the background rejection vs signal efficiency for W-boson tagging [3]. For a given signal efficiency, the higher the background rejection, the better.

5. Signal efficiency in data

The taggers described above were optimised on events from Monte Carlo (MC) simulation under the assumption that the simulation reproduces the observables and their correlations in data. To verify this assumption, the modelling of the input observables, the ML classifiers and the measurement of the signal efficiency is performed using 2015+2016 dataset of pp collisions at $\sqrt{s} = 13\,\mathrm{TeV}$ with integrated luminosity of $36.1\,\mathrm{fb}^{-1}$. The tag-and-probe method is employed in a sample enriched with top-quark pair ($t\bar{t}$) events with single triggered isolated electron or muon in the final state. The method is based on imposing selection enhancing the fraction of $t\bar{t}$ events using the relatively clean semileptonically-decaying top-quark signature (the tag).

Consequently, the presence of a trimmed anti-k_t $R = 1.0$ jet with $p_T > 200\,\text{GeV}$ isolated from the semileptonic top-quark decay products is required (the probe). Exact selection details are presented in [3]. The selected large-R jet with highest p_T is used to examine the modelling of observables and measure the tagging efficiency.

The signal efficiency measurement of the TopoDNN tagger is illustrated in Figure 3, demonstrating the agreement between data and the MC prediction well within the uncertainties. For all the investigated taggers in general, the uncertainties in the $t\bar{t}$-enriched sample are dominated by the $t\bar{t}$ modelling systematic uncertainties.

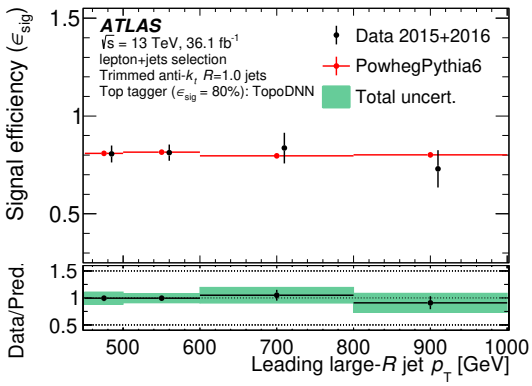

Figure 3. The signal efficiency of the TopoDNN top-quark tagger as a function of the large-R jet p_T in data and MC in the $t\bar{t}$-enriched sample [3].

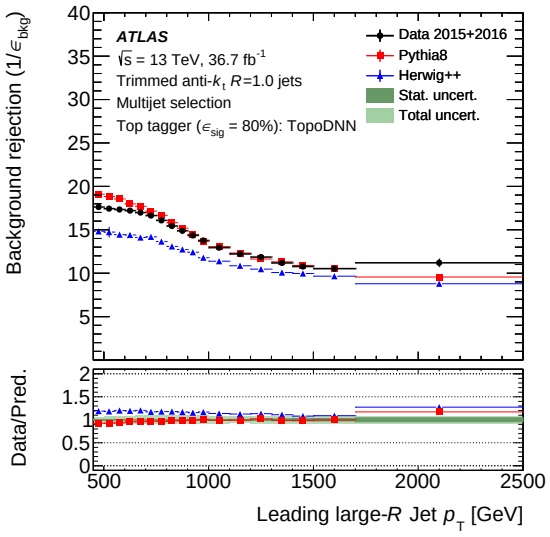

Figure 4. The background rejection of the TopoDNN top-quark tagger as a function of the large-R jet p_T in data and MC in the multijet sample [3].

6. Background rejection in data

Similarly to the signal efficiency measurement, the modelling of the input observables, the ML classifiers and the measurement of background rejection is performed in multijet and γ+jet samples, using the 2015+2016 data of pp collisions at $\sqrt{s} = 13\,\text{TeV}$. The multijet sample is obtained by requiring exactly zero isolated electrons or muons, fired single large-R jet trigger and at least one large-R jet with $p_T > 450\,\text{GeV}$. To obtain the γ+jet sample, the requirement of exactly zero isolated leptons is the same, and the event must trigger the single-photon trigger and at least one high-p_T photon must be present. In addition, a large-R jet isolated from the photon with $p_T > 200\,\text{GeV}$ is required. These selection criteria ensure negligible signal contamination in the background data samples. The two background topologies differ in terms of the contribution of quark-induced and gluon-induced jets, and additionally the γ+jet allows to study the background rejection modelling at lower jet p_T due to the lower photon trigger threshold compared to the single-jet trigger.

Figures 4 and 5 show the comparison of background rejection measurement in data and the MC prediction for the TopoDNN in multijet sample and DNN W in γ+jet sample, respectively. The level of agreement between the prediction and data in the multijet sample is reasonably good with some tension observed for the modelling of Herwig++ generator, while in the γ+jet sample, mismodelling in the rejection by Pythia8 is observed. The observed mismodelling motivates further studies towards an in-situ calibration of the background rejection.

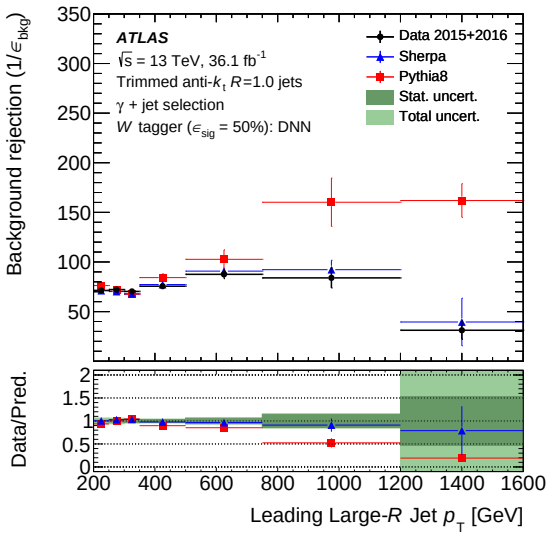

Figure 5. The background rejection of the DNN W-tagger as a function of the large-R jet p_{T} in data and MC in the γ+jet sample [3].

7. Conclusion

Advanced methods for identification of boosted hadronically-decaying top quarks and W bosons using BDTs and DNNs by combining multiple high-level features, or using low-level information from jet clusters are examined by ATLAS. Based on performance studies in MC simulation, it is shown that significant gains in background rejection can be achieved in contrast to using limited number of individual features. In addition, the signal efficiency and background rejection measurement in 2015+2016 ATLAS data is performed, obtaining comparison between the efficiencies calculated from MC and measured in data, as well as deriving the associated uncertainties on the measurement. This study demonstrates the possibility of performing in-situ calibration of complex ML classifiers in data to correct for mismodelling in the MC simulation.

References

[1] Evans L and Bryant P 2008 *JINST* **3** S08001
[2] ATLAS Collaboration 2008 *JINST* **3** S08003
[3] ATLAS Collaboration 2018 *Eur. Phys. J. C* **79** 375
[4] Cacciari M, Salam G P and Soyez G 2008 *JHEP* **04** 063
[5] ATLAS Collaboration 2017 *Eur. Phys. J. C* **77** 490
[6] Krohn D, Thaler J and Wang L T 2010 *JHEP* **02** 084
[7] ATLAS Collaboration 2016 *Eur. Phys. J. C* **76** 154
[8] ATLAS Collaboration 2016 *JHEP* **06** 093
[9] Goodfellow I, Bengio Y and Courville A 2016 *Deep Learning* (MIT Press) URL `http://www.deeplearningbook.org`
[10] Kingma D P and Ba J 2014 (*Preprint* `1412.6980`)
[11] Chollet F 2015 Keras URL `https://github.com/fchollet/keras`
[12] Theano Development Team 2016 (*Preprint* `1605.02688`)
[13] Höcker A *et al.* 2007 URL `https://cds.cern.ch/record/1019880`
[14] Speckmayer P, Hocker A, Stelzer J and Voss H 2010 *J. Phys. Conf. Ser.* **219** 032057
[15] Pearkes J, Fedorko W, Lister A and Gay C 2017 (*Preprint* `1704.02124`)