Comparative performance of ATLAS boosted W taggers using different AI/ML algorithms

M. LeBlanc (Manchester), on behalf of the ATLAS Collaboration
@matt.leblanc@cern.ch

BOOST 2023 (31 July - 4 August), Lawrence Berkeley National Lab, California, USA

Introduction

There are many ways to represent the information in a jet, e.g.:
- A set of precomputed features (e.g. JJS observables)
- An unordered set of particles (‘point-cloud’)
- A structured graph, based on the constituent kinematics, or the jet clustering history (e.g. Lund jet plane, ‘LJP’)

What are the advantages and disadvantages of each for W tagging? What gives the best performance?

Model-dependence

- While more complicated taggers are more accurate, they can also be more model-dependent!
  - Important to understand the origin of model-dependence, or else improved classification performance may not generalise to data!
  - Large data-to-MC scale factors, uncertainties can erode ML gains.

Classification Performance

- Complex architectures often result in better classification performance.
- Full ATLAS detector simulation is used:
  - R=1.0 CS+SK Soft-Drop (β=1, zcut=0.1) UFO jets
  - Consistent picture as a function of p_T
- LJP network outperforms others despite having fewer trainable parameters ∼ O(10^5).
- Physics-driven pre-processing provides clean picture for ML.
- Similar conclusions when studying the background rejection at a fixed tagging efficiency.

<table>
<thead>
<tr>
<th>Model</th>
<th>AUC</th>
<th>ACC</th>
<th>ɛ_{rel} @ ɛ_{tag} = 0.5</th>
<th>ɛ_{rel} @ ɛ_{tag} = 0.8</th>
<th># Params</th>
<th>Inference Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>EFN</td>
<td>0.920</td>
<td>0.835</td>
<td>35.1</td>
<td>7.95</td>
<td>56.73k</td>
<td>0.065 ms</td>
</tr>
<tr>
<td>PFN</td>
<td>0.931</td>
<td>0.833</td>
<td>44.7</td>
<td>9.50</td>
<td>57.13k</td>
<td>0.11 ms</td>
</tr>
<tr>
<td>ParticleNet</td>
<td>0.983</td>
<td>0.826</td>
<td>46.2</td>
<td>9.76</td>
<td>366.16k</td>
<td>0.36 ms</td>
</tr>
<tr>
<td>ParticleTransformer</td>
<td>0.951</td>
<td>0.880</td>
<td>77.9</td>
<td>14.6</td>
<td>2.14M</td>
<td>0.28 ms</td>
</tr>
</tbody>
</table>

Parton Shower Modelling

- Varying the Parton Shower algorithm in Herwig 7 MC samples causes large variations.
- W-tagging algorithms sensitive to modelling of perturbative emissions.
- LundNet differences anti-correlated from others.

Hadronisation Modelling

- AHADIC Cluster vs. Lund string hadronisation in Sherpa v2.2.5
- Most W-taggers not sensitive to soft emissions.
- LundNet tagger more sensitive: does the LJP picture over-emphasise soft emissions?