Deep Learning in Flavour Tagging at the ATLAS experiment

Marie Lanfermann

Abstract

A novel higher-level flavour tagging algorithm (DL1) has been developed using a deep neural network at the ATLAS experiment at the CERN Large Hadron Collider. We have investigated the potential of Deep Learning in flavour tagging using higher-level inputs from lower-level physics-motivated taggers. The resulting DL1 tagger is described and its performance presented, obtained from simulated $\ell$ events at $\sqrt{s} = 13$ TeV and the Run 2 data taking conditions where this tagger will be applied.

Motivation

A deep neural network (NN) may be better able to exploit correlations between flavour tagging input variables than the BDT approach currently used for flavour tagging in ATLAS. The application to the separation of charm and light jets is of particular interest.

Introduction to ATLAS Flavour Tagging

- Target: b-tagging, c-tagging
- Lower level variables are obtained from dedicated algorithms as discussed below
- Higher level algorithms then use those variables, along with kinematic variables (current recommendation: BDT-based (MV2))

Overview

Selection of lower level taggers:

Impact Parameter (IP) based
- IP2D, IP3D: Log-likelihood ratios using flavour hypotheses computed from summed track contributions extracted from simulation-derived templates
- RNN (recurrent neural network): Parallel approach which feeds raw tracks into a Recurrent Neural Network (RNN) and exploits correlations between the tracks

Secondary Vertex (SV) based
- SV1: Reconstructs inclusive secondary vertices
- SV2: Reconstructs inclusive secondary vertices
- JetFlavor: Exploits the topological structure of weak b- and c-hadron decays inside the jet by approximating the b-hadron or c-hadron flight path with PV, SV1 and tertiary vertex using a Kalman filter

Also used:
- Jet kinematics
- Information on muons produced in b/c decays

Optimisation Procedure

Building on the methods described previously for stabilising and improving the learning process, and preventing overtraining, a systematic grid search on the following aspects has been performed:
- Number of hidden layers
- Number of nodes in hidden layers
- Sequencing of Maxout and Dense layers
- Learning rate

for which the optimisation is based on empirical results. During the training, the loss development is monitored to check for overtraining and overall closeness of the predicted values w.r.t. the truth-label of the jets.

Performance Improvements

The DL1 final discriminant allows the background weighting to be changed. By scanning over the range of possible values, iso-efficiency curves using a fixed cut provide a figure of merit as shown below. The tuning of the final background weighting is done looking at the dependence of the performance on the kinematics. The DL1 inputs are the same as for MV2(b) (c-tagging), while MV2 for b-jet tagging does not use some JetFlavor variables originally designed for c-jet tagging.

b-jet tagging

Generally competitive performance, e.g. keeping the b-jet tagging efficiency fixed at 77% and light-flavour-jet rejection fixed at 101:
- MV2c10: 5.8 c-jet rejection
- DL1c9: 6.3 c-jet rejection

≈ 9% improvement

Validation

Significant expected improvements, e.g.:
- 25% c-jet efficiency:
  - ~110% light-flavour jet rejection at 16 b-jet rejection
  - 40% c-jet efficiency:
  - ~65% light-flavour jet rejection at 4 b-jet rejection

Input Modeling

Data/MC comparison of the IP3D log-likelihood ratio using the 15-dominated $c$-jet sample (left) and a $Z \rightarrow \mu^-\mu^+$-jets-dominated sample (right).

No systematic evaluation

Ratio plots only show a statistical error

Well modeled within 30% with some localised differences for low and high values

DL1 higher level tagger

Preprocessing
- To avoid discrimination based on kinematic differences between signal and background:
  - Reweight 2D (t, p_t) kinematics distribution of jet flavours to b-jet distribution
  - Use weights in backpropagation update
- Defaults should not disturb the learning process on physics cases
- Default values set to mean of distribution and flagged with binary check variables

NN design
- Adaptive Momentum (AdaM) optimiser to minimise categorical cross-entropy loss
- ReLU activation function (Softmax for output layer)
- Combination of Maxout and simple fully-connected layers
- Dropout
- Batch Normalisation

Construction of the Final Discriminant

b-jet tagging:

$$DL1_{b-jets} = \ln \left( \frac{P_b}{P_c + (1 - P_b)} \cdot \text{Plight-flavours} \right)$$

(1)

c-jet tagging:

$$DL1_{c-jets} = \ln \left( \frac{P_c}{P_b + (1 - P_c)} \cdot \text{Plight-flavours} \right)$$

(2)

Reduction to single dimension using a log-likelihood combination of the NN outputs, see equations (1) for b-jet tagging and (2) for c-jet tagging:
- Background weighting tunable after training
- Same training usable for b- and c-tagging

Conclusions

- Novel flexible higher level tagger ready to be used on 2017 data
- Improvements in b- and c-jet tagging
- In-depth performance studies and calibration in progress
- Promising comparison to data

References

[1] ATLAS Collaboration, b-tagging performance for release 2.1

This work was supported by grants of the Swiss National Science Foundation, grant number 156083 and 160025.

This poster was presented at the European Physical Society Conference on High Energy Physics 2017 in Venice, Italy.

marie.lanfermann@unige.ch