ATLAS Group @ The University of British Columbia | Department of Physics and Astronomy

Jet Clustering - a Machine Learning Feature for ATLAS Top Tagging

Alex Wen – *alex.wen@alumni.ubc.ca*

1. Top Tagging at ATLAS

The top quark, at a mass of 170 GeV, is by far the heaviest fundamental particle and therefore a critical probe of many (Beyond) Standard Model properties.

ATLAS produces top quarks by colliding ultrarelativistic protons, providing a large amount of energy. While top quarks are produced, other irrelevant particles are also part of the overall detector signal.

To learn anything about the top, it is critical to accurately distinguish a top signal from other backgrounds.

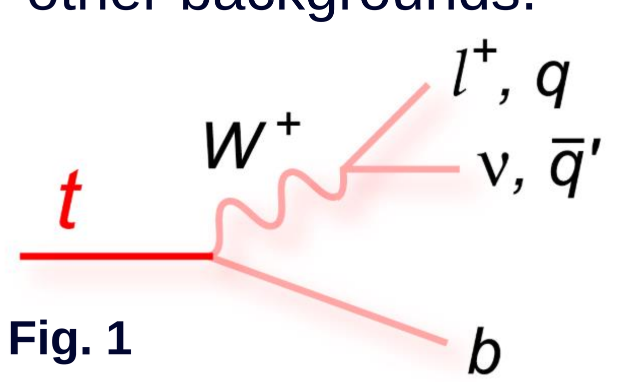


Fig. 1

Common top decay.

Fig. 2

Boosted and non-boosted clusters.

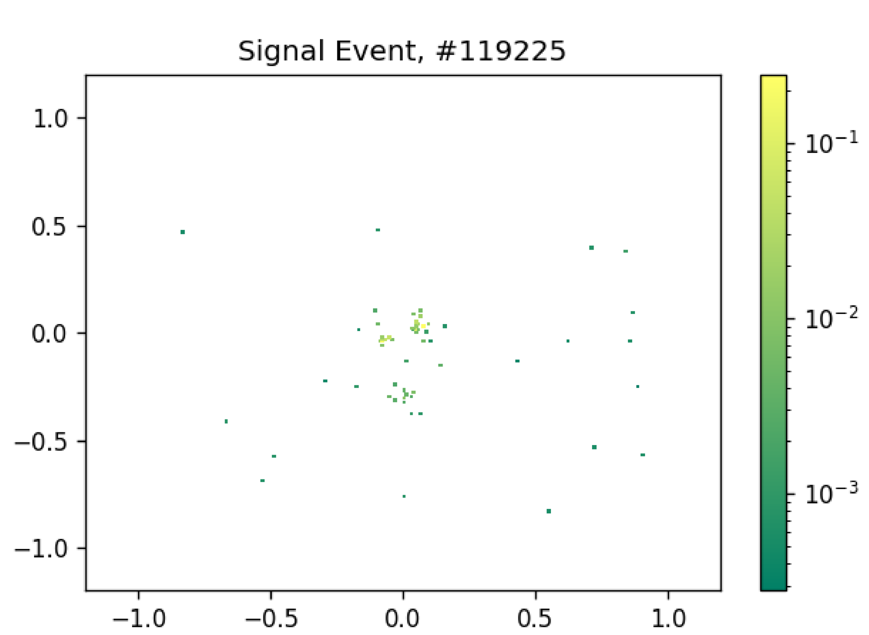


Fig. 3

An example of top decay energy deposits, scaled in color by P_t, in $\phi - \eta$ (spatial) space. Three distinct clusters are visible.

2. Decay

Top quarks predominately decay into three other quarks which then hadronize and form clusters (called jets) of energy deposits in the detector. Each individual energy deposit is a constituent.

Hence the defining feature of a top quark is three jets (Fig. 1) which add up to the top mass.

As jets get boosted, they will become collimated and overlap (Fig. 2), making identification challenging.

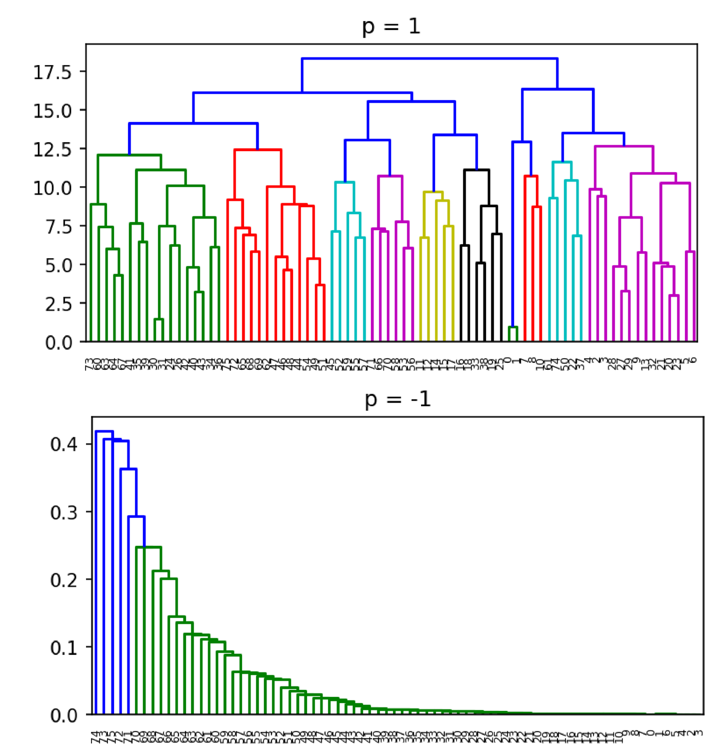


Fig. 4

Trees visualizing the clustering sequence of k_t (top) and $\text{anti}-k_t$ (bottom) algorithms for an example event. The horizontal axis is populated by constituents, and they are combined in a certain order to form the last element at the top.

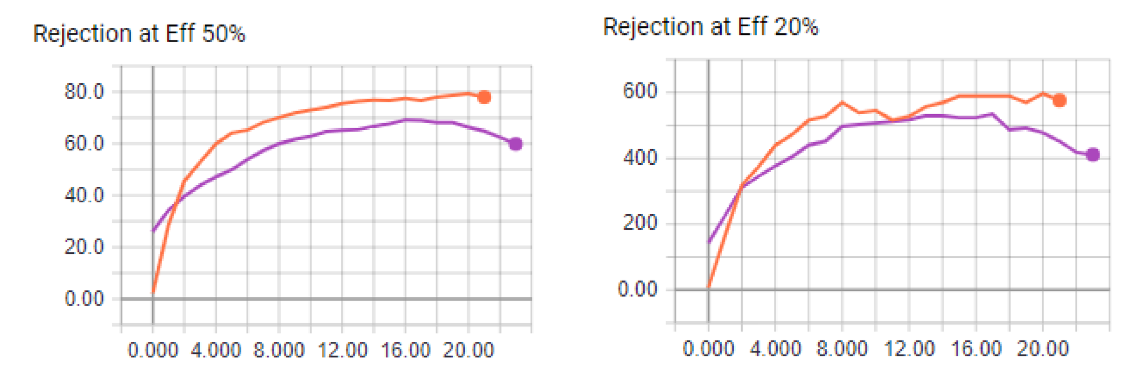
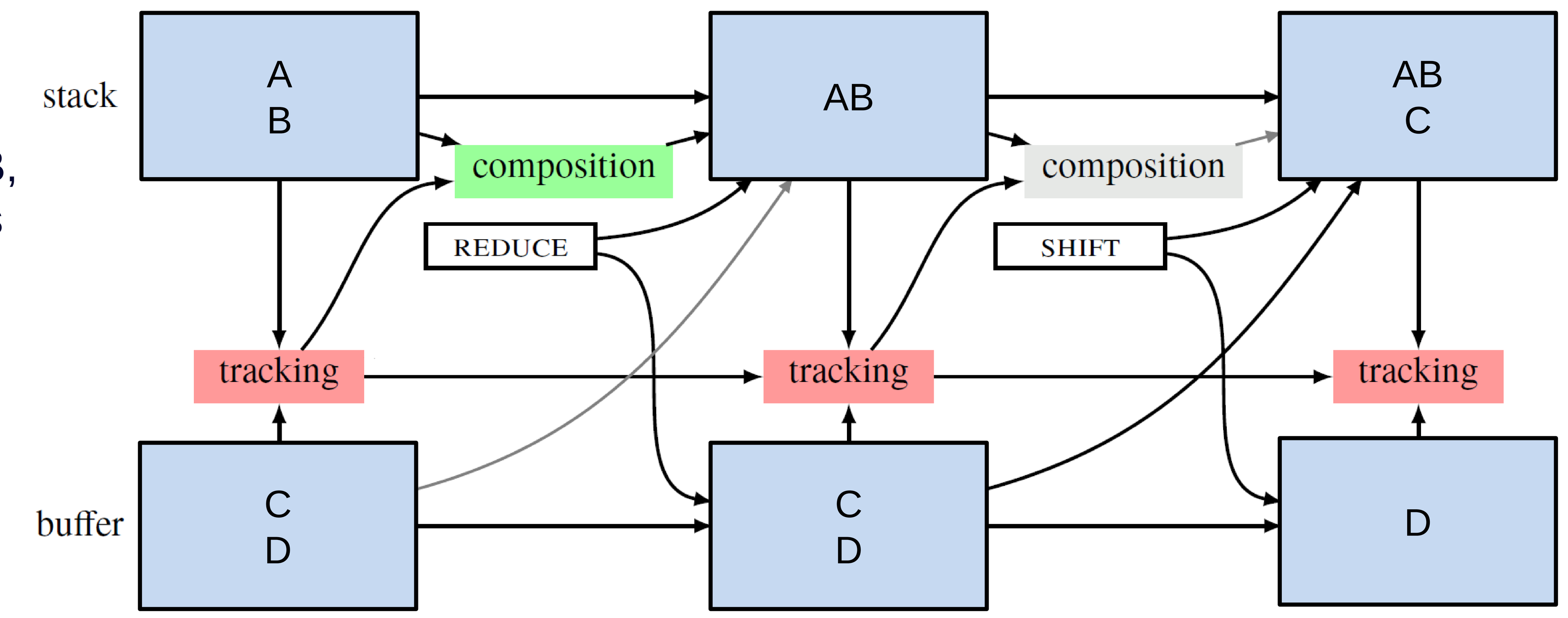


Fig. 7

Background rejection as a function of training epoch at 20% (left) and 50% (right) signal efficiency. k_t (purple) and $\text{anti}-k_t$ (orange) are shown. Background rejection is the reciprocal of the false positive rate.

Fig. 5

The network architecture of SPINN unrolled across just two steps. A, B, C, D are jet constituents, and AB is a combined element. Gray arrows indicate an inactive transition, and colored boxes are layers. SPINN follows the sequence from our clustering algorithm and combines elements in that order using a sequence of shifts and reduces.



3. Clustering & Jet Structure

Given low-level constituent data from the detector (Fig. 3), we use clustering algorithms to recursively combine constituents based on a distance metric. We can reconstruct jets from the bottom-up and create a tree-like decay sequence that would help indicate a top signal (or not).

The metric defines the clustering algorithm. Clustering jets with a well-chosen algorithm injects physically-motivated information about the event and should be a valuable feature for training a classifier, since top and background sequences should be significantly distinct. Of interest to us are the k_t and $\text{anti}-k_t$ algorithms (Fig. 6).

With the clustering sequence as a feature, the order in which we group and read constituents of an event are significant now. This has many parallels with natural language processing.

4. Neural Network

We choose a neural network that can utilize our clustering sequence information.

The Stack-Augmented Parser Interpreter NN (SPINN) is a logical choice. It builds on the LSTM structure and exactly suits our need to introduce a clustering sequence to our events. It combines events in a pre-defined way according to the sequence our clustering algorithm defines. It will learn how to best combine constituents.

It uses two auxiliary data structures, the stack and buffer, and uses shift-reduce language to encode a clustering sequence. An example with 4 constituents is in Fig. 5.

Fig. 6

Distance metrics between constituents i and j given transverse momentum p, pseudorapidity η, and azimuthal angle ϕ of the constituent.

$$ak_t\colon\ d_{ij} = \min\left(p_i^{-2}, p_j^{-2}\right)\frac{\Delta^2}{R^2}$$

$$k_t\colon\ d_{ij} = \min\left(p_i^{2}, p_j^{2}\right)\frac{\Delta^2}{R^2}$$

$$\Delta^2 = \left(\eta_i - \eta_j\right)^2 + \left(\phi_i - \phi_j\right)^2$$

5. Performance

To test performance, we used 2M event samples of Delphes-generated ATLAS data in the 600-2500 GeV range with equal numbers of top signal and dijet background and a 5-5-90% test, validation, and training split.

The same events were clustered with k_t and $\text{anti}-k_t$ and trained separately. Training was stopped after we noticed overtraining. Rejection efficiencies are plotted in Fig. 7.

Looking at Fig. 4, the k_t clustering is physically more realistic; it is surprising that $\text{anti}-k_t$ seems to have better performance.

It should be noted that this method using SPINN and $\text{anti}-k_t$ outperforms previous analyses done by our group using DNN and normal LSTM architectures [2].

6. Conclusions & Outlook

We conclude with these observations:

- Clustering sequence has a significant effect on performance.
- $\text{anti}-k_t$ starts off worse but eventually performs better. Given that k_t is the more realistic choice, we are not sure why.

References

1. Bowman et al.; *A Fast Unified Model for Parsing and Sentence Understanding,* 2016.
2. Pearkes et al.; *Jet Constituents for Deep Neural Network Based Top Quark Tagging,* 2017.
3. Cacciari et al.; *The anti-kt Jet Clustering Algorithm,* 2008.
4. Butter et al.; *The Machine Learning Landscape of Top Taggers,* 2019.
5. Images:
   - Fig. 1 – Fermilab https://www-d0.fnal.gov/
   - Fig. 5 – adapted from [1]

Acknowledgements

Thanks to everyone at UBC ATLAS, especially my supervisors Colin Gay and Alison Lister, for their tireless input.