B-tagging on ATLAS

Michael Kagan

SLAC

Workshop on New Techniques In Particle Reconstruction for VBS
October 23, 2018
Bottom Quark Jet Identification

- Goal: Discriminate b-jets from c-jets, light-jets, tau

- B-tagging algorithms utilize the long lifetime and displaced decays of b-hadrons to look for secondary vertices and displaced tracks
Baseline B-tagging Algorithms on ATLAS

- **IP3D / IP2D:**
  - Impact parameter algorithm
  - Exploit (in)compatibility of track with PV

- **SV:**
  - Inclusive Secondary vertexing
  - Determination of single inclusive weak b-hadron decay vertex

- **JetFitter:**
  - PV→B→D decay chain finding
  - More detailed determination of decay vertex topology
ML based Combination – “High-Level” taggers

- Train models utilizing the outputs of impact parameter taggers and properties of vertices from vertexing algorithms
- MV2 – Boosted Decision tree based model
- DL1 – 8-hidden layer Deep Neural Network, multi-class classification
Tagger Performance

- Excellent performance for both MV2 and DL1
- Multi-class output for DL1: flexible definitions of b-tagging and c-tagging

\[
\text{for } b\text{-tagging} \quad DL1 = \ln \left( \frac{p_b}{f_{c\text{-jets}} \cdot p_c + (1 - f_{c\text{-jets}}) \cdot \text{Plight-flavour}} \right)
\]

\[
\text{for } c\text{-tagging} \quad DL1 = \ln \left( \frac{p_c}{f_{b\text{-jets}} \cdot p_b + (1 - f_{b\text{-jets}}) \cdot \text{Plight-flavour}} \right)
\]
Tagger Performance

- Tagger performance, especially at high-$p_T$, can be improved with good choice of training samples
  - ttbar for low $p_T$ jets
  - $Z' \rightarrow \text{bb}$ for high $p_T$ jets
• Calibrations scale factors to correct tagging efficiency differences between data and MC in key kinematic variables

• Pure samples of b, c, and light-jets in data used to measure unbiased estimate of tagging (mistag) rate
  – Assumption: data/MC differences measured in one sample applicable elsewhere
  – Measuring SF as function of key factors that impact performance differences (p_T, η, distance to nearest jets, etc.) help to mitigate topological dependence

\[
SF(p_T) = \frac{\epsilon_{\text{data}}(p_T)}{\epsilon_{\text{MC}}(p_T)}
\]
Calibration Results

- Calibration samples:
  - For b-jets: ttbar dilepton events
  - For c-jets: Hadronic W-decays in ttbar single lepton events
  - For light-jets: Z+jets events, looking for “negative” tags behind jet

- Increasing precision
  - Improved physics modeling
  - Improved simulations of material interactions
  - Improved analyses to constrain estimates of non-b backgrounds
Pseudo-Continuous b-tagging

- Can we use the full MV2/DL1 score distribution to define different “quality” of b-tags?

- **Pseudo-continuous b-tagging** provides calibrations in bins of $p_T$ and b-tag score

- Allows for defining multiple signal regions of varying purity

- Could be used with ML classifier
  - Propagate systematics (including correlations) “bottom up”
Soft Muon Tagging

- Exploit large semi-leptonic b-decay fraction (~20%) and large b-hadron mass to find muons for b-tagging

- **SMT**: Soft-muon tagger BDT utilizes:
  - Muon kinematics and impact parameter
  - Track quality, to reject fakes and decays in flight
Impact Parameter Tagging

- Impact Parameter based taggers:
  \[ p(\text{jet flavor} \mid \text{tracks in jet}) \]
  - Large number of tracks in jet \( \rightarrow \) Dimensionality too high for easy density estimation
  - Often make naïve Bayes assumption that tracks independent
Sequence learning and Recurrent NN

• Instead of considering the tracks as individual objects, treat them as sequence
  – Make use of sequence classification techniques in ML!
  – Naturally unordered sequence, but can impose a physics-inspired ordering:
    Order here by largest impact parameter significance

• Neural network approach to analyzing sequences is Recurrent Neural Networks
  – Used in sentence classification, Natural Language Processing, time-series analysis, etc.
Jets and Sequence Processing

Recurrent Neural Network

Image from I. Goodfellow, *Deep Learning Book*
• Order tracks by impact parameter

• RNN can learn inter-track dependencies
• Order tracks by impact parameter

• RNN can learn inter-track dependencies
New High-Level taggers

**MV2 Family:**
- MV2 = Baseline
- MV2mu = MV2 + SMT inputs
- MV2murnn = MV2 + SMT + RNN inputs

**DL1 Family:**
- DL1 = Baseline
- DL1mu = DL1 + SMT inputs
- DL1murnn = DL1 + SMT + RNN inputs

†: BDT Based
*: NN Based
Performance in ttbar sample
Performance in high-$p_T$ $Z' \rightarrow bb$ sample
ML improvements to single b-jet tagging

![Graph showing light-flavour jet rejection for ε_0 = 77% as a function of Jet p_T [GeV]. The graph includes data points for different configurations: MV2 - 2016 configuration, MV2 - 2017 configuration, MV2Mu - 2017 configuration, and MV2MuRnn - 2017 configuration. The CMS = 13 TeV, Z'].

ATLAS Simulation Preliminary

ATL-PHYS-PUB-2017-013
ML improvements to single b-jet tagging

High-Level Tagger Data/MC comparisons in ttbar sample
- Full calibrations in progress
Conclusion

• Excellent performance from ATLAS b-tagging algorithms allows for use in a variety of analyses with variety of tagging efficiency and mis-tag rate working points

• Precision calibrations available for tagging and mis-tag rates
  – With continued improvements in MC modeling, simulations, and analysis methods, expect much improved precision of these calibrations

• New ML driven algorithms can have large impact on improving taggers
  – Exploration of ML driven algorithms will continue!
Boosted B-Tagging
Track Jet based Boosted Higgs Tagging

• Jet clustering on tracks using small-R to resolve fine features

• Match track jets to larger-R jets

• Pros
  – Excellent track resolution allows for small-R clustering
  – Pileup insensitive
  – Independent of large-R jet
    • Insensitive to grooming algorithm
    • b-tagging calibration independent of large-R jet

• Cons:
  – No access to neutral particle information in finding direction
H-Tagging with Fixed Radius Track Jets

**ATLAS Simulation Preliminary**

- $p_T > 250 \text{ GeV}, 76 \text{ GeV} < m_{\text{jet}} < 146 \text{ GeV}$
- Double b-tag
- Asymm. b-tag (70% wp)
- Single b-tag
- Leading subjet b-tag

**ATLAS Simulation Preliminary**

MV2c10 b-tagging at 77% WP

- 1 b-tag, Loose $m_{\text{jet}}$ window
- 2 b-tags, No $m_{\text{jet}}$ selection
- 2 b-tags, Loose $m_{\text{jet}}$ window
- 2 b-tags, Tight $m_{\text{jet}}$ window, $D_2$ sel.

**ATLAS Simulation Preliminary**

MV2c10 b-tagging at 77% WP

- 1 b-tag, Loose $m_{\text{jet}}$ window
- 2 b-tags, No $m_{\text{jet}}$ selection
- 2 b-tags, Loose $m_{\text{jet}}$ window
- 2 b-tags, Tight $m_{\text{jet}}$ window, $D_2$ sel.
Performance in Data

- **Left:** Single jet b-tagging calibration in dilepton t-tbar events

- **Right:** Muon matched large-R jets in multijet events for double b-tagging validation in gluon → bb enriched sample
  - Useful to check for topological effects or “cross talk” between among the two b-tagged jets

![Graph showing data and MC comparison for jet transverse momentum](image)

- ATLAS
- \( \sqrt{s} = 13 \text{ TeV} \times L \text{ dt} = 3.2 \text{ fb}^{-1} \)

- LH Method
  - MV2c10, \( \varepsilon_b = 70\% \), single-cut OP
  - Anti-\( k_t \), \( R=0.2 \) track-jets
  - Total Uncertainty
  - Stat. Uncertainty

- Data/MC
  - Events/40 GeV
  - Large-R Jet \( p_T \) [GeV]
VR Track Jet Based b-tagging

- At high Higgs jet $p_T$, fixed radius track-jets will merge
  - One option: use single and double track-jet selections

- Shrink track jet $R$ with $p_T$
  - $R \rightarrow R_{eff}(p_T) = \frac{\rho}{p_T}$

- Pro:
  - Resolve and tag multiple b-jet region of interests up to higher Higgs jet $p_T$

- Cons:
  - Need to be careful about overlapping track jets
VR Track Jet Based b-tagging

- At high Higgs jet $p_T$, fixed radius track-jets will merge
  - Necessitates using single and double track-jet selections

- Shrink track jet R with $p_T$

- Pro:
  - Resolve and tag multiple b-jet region of interests up to higher Higgs jet $p_T$

- Cons:
  - Need to be careful about overlapping track jets
Exclusive subjet guided b-tagging

- Identify two exclusive regions of interest using exclusive-$k_T$ declustering of large-R jet
  - ATLAS: standard b-tagging on ExKt subjets
  - CMS: Tagger using RoIs based on $\tau$ axes (found with exclusive-$k_T$)

- Alternatively boost into center of mass frame and do exclusive clustering
  - E.g. with EECambridge algorithm, $\min y_{ij} = 2(1-\cos\theta_{ij})$
Note: Trainings are based on ttbar samples, and no specific training has yet been performed using jets in boosted topologies.
B-jet properties and algorithms
Fragmentation of b-quarks in b-hadrons

- Roughly 80% of the time, b-quarks in the final state fragment into excited B hadrons ($B^*$, $B^{**}$)

- Excited B hadrons decay to weakly decaying B hadrons (plus additional particles)

- The b-quark fragmentation is hard: on average most of the original b-quark energy (~70%) goes into the B-hadron

B hadron types

<table>
<thead>
<tr>
<th>Mesons:</th>
<th>$B^+_u = \bar{b}u$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$B^+_d = \bar{b}d$</td>
</tr>
<tr>
<td></td>
<td>$B^+_s = \bar{b}s$</td>
</tr>
<tr>
<td></td>
<td>$B^+_c = \bar{b}c$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Baryons:</th>
<th>$\Lambda^0_b = bud$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\Xi^0_b,^- = bus, bds$</td>
</tr>
<tr>
<td></td>
<td>$\Omega^-_b = bss$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>b-hadron</th>
<th>Production rate (LEP, Tevatron, LHCb, average)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$B^+$</td>
<td>$40.6 \pm 0.5%$</td>
</tr>
<tr>
<td>$B^0$</td>
<td>$40.6 \pm 0.5%$</td>
</tr>
<tr>
<td>$B_s^0$</td>
<td>$10.5 \pm 0.5%$</td>
</tr>
<tr>
<td>b-baryon</td>
<td>$8.3 \pm 1.0%$</td>
</tr>
</tbody>
</table>

Heavy Flavor Averaging Group
http://arxiv.org/abs/1412.7515

ATL-PHYS-PUB-2014-008
**B-hadron decays**

- Weakly decaying b-hadron: \( \tau \sim 1.5 \times 10^{-12} \) s and mass \( \sim 5 \) GeV
  - A b-hadron with \( p_T \sim 30 \) GeV has decay length of:

  \[
  L = \beta \gamma c \tau = \left( \frac{p}{m} \right) c \tau \sim 3 \text{mm} \quad \Rightarrow \quad \text{Measurable displace vertex!}
  \]

- \(| V_{cb} | \gg | V_{ub} | \rightarrow \text{most b-hadrons decay to c-hadrons}
  - c-hadron \( \tau \sim 0.4 - 1.0 \times 10^{-12} \) s \rightarrow \text{tertiary vertex} in b-hadron decay chain

- Approximately 40% of b-hadron decays are semi-leptonic
  - \( \sim 10\% (b \rightarrow l) \) directly and \( \sim 10\% (b \rightarrow c \rightarrow l) \), where \( l = e, \mu \)

- Use all this information to design b-taggers!

Note: b/c decay model very generator dependent
[J. Brosamer, M. Shapiro, LBNL], Introduce EvtGen to ATLAS
Fakes from Light jets

• Most tracks in light jets come from fragmentation of the quark

• Displaced vertices can be produced from:
  – Hadronic interactions in material
  – Photon conversion
  – Long lived \((K_s, \Lambda)\) decays \((c\tau \sim 3-8 \text{ cm})\)
  – Poorly measured tracks / tracks with shared hits in B-layer leading to fake tracks / fake vertices
Inputs to b-jet identification

• **Jets** (based on calorimeter or track information)
  – **Direction**: seed b-tagging algorithms and assign lifetime sign to tracks
  – **$P_T$ and $\eta$**: exploit correlation with physics properties and detector resolution

• **Tracks**
  – Perigee representation of trajectory: $(d_0, z_0, \phi, \theta, \frac{q}{p})$
    • Impact parameters are very important for b-tagging
  – Select tracks using $p_T$-dependent cone around jet axis
    \[ \Delta R(\text{jet}, \text{track}) < R_{\text{cut}}(p_T) \]

• **Leptons**
  – $\mu/e$ to identify semi-leptonic b-decays

![Diagram showing the process of b-jet identification](image-url)
Impact Parameter Algorithms

- **Signed impact parameter significance**
  
  $$S_{d_0} = \text{sign}_{d_0}\left(\frac{d_0}{\sigma_{d_0}}\right)$$
  
  $$S_{z_0} = \text{sign}_{z_0}\left(\frac{z_0}{\sigma_{z_0}}\right)$$

- **Build 2D PDF’s of IP significance** to compute track likelihood

- **Discriminator based on** likelihood ratio of different flavor hypotheses
Impact Parameter Algorithms

- **Signed impact parameter significance**

  \[ S_{d_0} = \text{sign}_{d_0} \left( \frac{d_0}{\sigma_{d_0}} \right) \]
  \[ S_{z_0} = \text{sign}_{z_0} \left( \frac{z_0}{\sigma_{z_0}} \right) \]

- **Discriminator based on likelihood ratio of different flavor hypotheses**

  \[ \text{sign}_{r\phi} = \text{sign} \left( \sin(\phi_{jet} - \phi_{trk}) \cdot d_{0,\text{trk}} \right) \]
  \[ \text{sign}_{3D} = \text{sign} \left( [\vec{p}_{trk} \times \vec{p}_{jet}] \cdot [\vec{p}_{trk} \times \Delta \vec{r}_{IP}] \right) \]

- **Build 2D PDF’s of IP significance to compute track likelihood**

- **Discriminator based on likelihood ratio of different flavor hypotheses**

  \[ \text{LR}(IP_1, IP_2, ..., IP_N) = \frac{\prod_{i=1}^{N} \text{PDF}_b(IP_i)}{\prod_{i=1}^{N} \text{PDF}_l(IP_i)} \]
Inclusive Secondary Vertexing Algorithms

• Fit a single inclusive vertex to B decay
  1. Find all 2 track vertices within jet with good $\chi^2$ prob. and decay length significance
  2. Remove vertices with mass consistent with $K_s$, $\Lambda$, photon conversion
  3. Remove vertices at pixel layers
  4. Fit all tracks from good vertices into single vertex

• Form likelihood discriminant using vertex properties
JetFitter

• Avoid fitting all tracks to single vertex

• Constrain track stemming from B/D vertices to **intersect single B-flight axis**

• Use Kalman filter to fit multiple vertices along flight axis
  – Based on “ghost track” method from SLAC-PUB-8225 (1999)

• **Harder for light jet to fake a 2 vertex decay topology**

• Can provides slightly better separation between b- and c- jets
RNN b-tagging

- RNN captures correlations not seen by IP3D
  - even with only impact parameter significance and track category inputs