Long Lived Machine Learning

G. Watts (UW/Seattle)
For the ATLAS Collaboration
Join us for talks, round tables, and discussions on
• Machine Learning in Physics
• New algorithms for Particle Physics
• Advances in algorithms for theory computations
• Tools and infrastructure for computing in physics

There is a fundamental shift occurring in how computing is used in research in general and data analysis in particular. The abundance of cheap, powerful, easy to use computing power in the form of CPUs, GPUs, FPGAs, etc., has changed the role of computing in physics research over the last decade. The rise of new techniques, like deep learning, means the changes promise to keep coming.
Longed Lived Particle Signatures

A Little Physics Context
The Standard Model

The Hidden Sector

Communication Mechanism

Hidden Sector does interact via Gravity
Why haven’t we seen anything yet?

G. Watts (UW/Seattle)
Why haven’t we seen anything yet?

- LHC (Run 2)
- Mixing
- Tunneling/Mixing
- Physics of the Hidden Sector/Valley
- Standard Model Particles

G. Watts (UW/Seattle)
Energy

How long it lives $\tau$

What it decays to, how many of them will you expect, kinematics, etc.
The mixing means the end result will be Standard Model Particles

- Couples to leptons ($M_X$ small)
  - “lepton-jets” – jets of leptons
- Couples to heavy fermions (quarks)

Which means two (displaced) jets
Detection strategy depends on where in the detector the decay happens.

Let's concentrate on the calorimeter.
Identifying Displaced Jets in the ATLAS Calorimeter
Decays in the Calorimeter

1. Look for the “appearance” of energy the Calorimeter
2. Little or no activity in the tracker
“CalRatio”

ATLAS Simulation

$m_H = 126$ GeV - $m_\tau = 10$ GeV

Signal Like

Background Like
ATLAS

\[ L = 15 \, \text{nb}^{-1}, \, \sqrt{s} = 8 \, \text{TeV} \]

- Multi-jets
- Signal: decay in HCal
- Signal: decay in ID

Fraction of Jets

\begin{align*}
\text{Fraction of Jets} & = 1 - \text{Background} \\
& = 1 - \text{Multi-jets} \\
& = 1 - \text{Signal: decay in HCal} \\
& = 1 - \text{Signal: decay in ID}
\end{align*}

\[ n_{\text{tracks}} \]

G. Watts (UW/Seattle)
Can We Do Better?

There are lots more attributes connected with displaced jets.
13 Jet Shape Variables
Boosted Decision Tree

We used TMVA, and an old version at that.

Training:

- Jets with $p_T > 40$ GeV, $|\eta| < 2.4$
- Signal if $\Delta R < 0.2$ to a $\pi_v$

We are just finding out we are missing out here!

Reweight to flatten $p_T$ spectra to avoid model dependence.
Training

A few parameter modifications from the defaults:
- Minimum Leaf Size: 1%
- Maximum Training Depth: 20
- Signal Training: 944,638
- Background Training: 1,499,996

Measuring performance:
- What BDT cut gives you a straight-cut background rate?
- Calculate the signal efficiency

Systematic Errors:
- Many input variables depend on Jet Energy Scale
- Required re-running on full sample many many times

![Graph showing performance of 13Lxy BDT](image)
## Ordered Variables

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\log_{10}(E_{\text{H}}/E_{\text{EM}})$</td>
<td>The base-10 logarithm of the ratio of the energy in the HCal to ECal. If no energy is present in the ECal, a large number is used as input to the BDT instead.</td>
</tr>
<tr>
<td>Jet Width</td>
<td>The $p_T$-weighted sum of the $\Delta R$ between each cluster the jet is built from and the jet axis.</td>
</tr>
<tr>
<td>Leading cluster Longitudinal length</td>
<td>How far the highest $p_T$ cluster has spread in the longitudinal direction.</td>
</tr>
<tr>
<td>Jet $p_T$</td>
<td>Training events are reweighted so that the jet $p_T$ distribution is flat. This allows the BDT to look for correlations between other variables and the jet $p_T$ without using jet $p_T$ directly as a discriminating variable.</td>
</tr>
<tr>
<td>Leading cluster lateral width</td>
<td>How far the highest $p_T$ cluster has spread in the latitudinal direction.</td>
</tr>
<tr>
<td>Leading Jet Cluster Shower Center</td>
<td>The distance from the inner edge of the ECal to the highest $p_T$ cluster’s center along the jet axis. A zero would mean this was on the edge of the ECal.</td>
</tr>
<tr>
<td>$\Delta R$ to Closest 2 GeVTrack</td>
<td>The $\Delta R$ between the jet axis and the closest 2 GeV track. Or 0.4 if there is no such track.</td>
</tr>
<tr>
<td>Radius of the leading cluster</td>
<td>The radius from the beam spot to the highest $p_T$ cluster.</td>
</tr>
</tbody>
</table>

### Cluster Energy Density

- **Number of tracks**
- **Sum $p_T$ of all tracks**
- **Layer 1 Energy Fraction**
- **Max $p_T$**

The $\frac{\sum E_i^2/V_i}{\sum E_i}$, where $E_i$ is the energy of each cell, and $V_i$ is the volume of the cell in the highest $p_T$ cluster in the jet. Only positive energy cells are considered.

- The number of tracks with $p_T > 2$ GeV within a $\Delta R < 0.2$ of the jet axis.
- The sum $p_T$ of all tracks within $\Delta R < 0.2$.
- The fraction of the hadronic calorimeter’s energy located in layer 1. The later in the calorimeter a LLP decays, the smaller value this variable will be.
- The maximum $p_T$ of all tracks within $\Delta R < 0.2$.

### Notes
- **This result released as a conference note**
- **Same sensitivity with 3.2 fb$^{-1}$ of data as the previous one with 20 fb$^{-1}$**
  - Note the $\sqrt{s}$ change!
- **Low mass sensitivity not as good in Run 2!**
What is next?

We had to limit the sensitivity of the last BDT artificially.
• Beam induced background is ugly and looks exactly like our signal.
• Currently training a multi-class BDT to look for Multijet, signal, and Beam Induced Background

Further in the future:
• The calorimeter is just a 3D CCD
• These jets have particular shapes
• Use Deep Learning to do image recognition.
• Some work along these lines in other contexts has already been done
Peaches and Cream
Pain...

GRID Samples (xAOD) → Group Flat Ntuples → Training Ntuples → Training

I am a professor! How the heck can I tell the Internal Review Board what dataset I used on Tuesday, 6 months ago?!

Partly ATLAS’ Fault
Partly the tools
Partly the analysis team

G. Watts (UW/Seattle)
Testing the training required a huge amount of repeated runs. Also a disaster to track.
Automated Everything

- Tracks from GRID to TMVA training XML
- Scripts to download and scan log files automatically to generate performance tables

Command line configuration parameters
TMVA Training Simpler?

```javascript
var training = flatSignalTrainingData
    .AsSignal(isTrainingEvent: e => !(e.EventNumber % 3 == 1))
    .Background(flatBackgroundTrainingData, isTrainingEvent: e => !(e.EventNumber % 3 == 1))
    .UseVariables(varList);

// Build options (like what we might like to transform.
var m1 = training.AddMethod(ROOTNET.Interface.NTMVA.NTypes.EMVA.kBDT, "BDT")
    .Option("MaxDepth", options.BDTMaxDepth.ToString())
    .Option("MinNodeSize", options.BDTLeafMinFraction.ToString())
    .Option("nCuts", "200")
    ;

// Do the training
var trainingResult = training.Train("JetMVATraining");
```

Does the required flexibility really require the existing TMVA complexity?

Obviously I should be exploring some of the tools mentioned here
Conclusions

• How to give theorists object efficiency plots for reinterpretation of the results

• The Community White Paper processes
  • The groups are not islands
  • Analysis complexity is rising too fast
  • For us: ML is too small a part of the overall experiment

• Automate all tools

• Make sure their use can easily be moved to the “cloud”
  • Config files, network access required, keep to a minimum

• I’ve been involved in ML for a long time
  • Since DZERO single top discovery days

• This is the first time I did it
  • Rather than a student or post-doc of mine. 😊

• I’m bullish in that given my limited analysis time I was able to contribute to an analysis
  • Obvious measure of TMVA and ML’s success and accessibility

• But, I am a “pet” analyzer

G. Watts (UW/Seattle)