Application of Boosted Decision Trees in the search for a new boson decaying into $Z + \gamma$

Ntsoko Phuti Rapheeha

Institute for Collider Particle Physics, University of the Witwatersrand

February 1, 2019
Overview

1. Model
2. MC Samples
3. MVA Choice
   - What are BDTs
4. Preselection Cuts
5. Input Variables
   - Description of the Input Variables
   - Distributions of the Input Variables
6. Correlation Plots
7. Classifier hyper-parameters
8. Receiver Operating Characteristic (ROC) Curve
9. Overfitting Check
   - Signal-Background separation
   - Test-Train BDT Output plots
10. Conclusion
11. backup
Motivated by a strong deviation in the production of multiple leptons a the Electroweak scale,

a Beyond Standard Model search is attempted to find a resonance in the Z and photon system with association with jets and missing transverse energy ($E_T^{\text{miss}}$)
Leptons

- Electrons are reconstructed from tracks in the Inner Detector (ID) and energy deposits from the ECAL while muons are reconstructed from the tracks in the muon chambers.
- Selected leptons are required to have $p_T > 10$ GeV, passing a medium identification requirement and the Loosetrack only criteria.

Electrons

- They are to be found in $|\eta| < 2.7$ excluding the transition region $1.37 < |\eta| < 1.52$, with impact parameters $\frac{d_0}{\sigma d_0} < 5$ and $z_0 \sin(\theta) < 0.5$.

Muons

- They are to be found in $|\eta| < 2.7$, and with impact parameters $\frac{d_0}{\sigma d_0} < 3$ and $z_0 \sin(\theta) < 0.5$. 
Object reconstruction and selection

Photons
- They are reconstructed from energy deposits in the electromagnetic calorimeter (ECal).
- The reconstructed photons are required to have transverse momentum, \( p_T > 10 \text{ GeV} \), and be found in the pseudorapidity range \( |\eta| < 2.7 \) excluding the transition region \( 1.37 < |\eta| < 1.52 \).
- They also have to pass the loose identification requirement.

Jets
- They are reconstructed from energy deposits in the calorimeter systems using the anti-\( k_t \) reconstruction algorithm with \( R = 0.4 \).
- They required to have \( p_T > 25 \text{ GeV} \), be found in \( |\eta| < 4.4 \) and fulfill the LooseBad jet quality requirement.
Object reconstruction and selection

Missing Transverse Energy

- It is caused by objects that don’t interact with the detector material, the objects can either be neutrinos or signatures of new physics that is beyond the standard model.

- It is reconstructed as the net imbalance of the transverse momentum, $\rho_T$
The Monte Carlo samples used follow the 2017 Data’s pileup distribution.

- **Z+\gamma** events, full simulation via Sherpa 2.1, this accounts for 84.6% of the total background
  - **\( ll - \gamma\)** (ee - \( \gamma \) and \( \mu\mu - \gamma \))
- **tt\gamma** events, full simulation via MadGraphPythia8EvtGen 5%,
- **Z + jets** events contribute the last 10.4% of the total background
- The **qqZH** is used as the signal sample
  - It contains real missing energy
High luminosities are required for BSM physics searches, pile-up conditions get worse with increasing luminosity and this affect reconstruction of $E_T^{\text{miss}}$.

- A Multivariate Analysis (MVA) method, Decision Trees, is employed to classify between fake $E_T^{\text{miss}}$ real $E_T^{\text{miss}}$ signal

- They are a series of binary splits that optimize the separation between signal and background
A boosted decision tree algorithm builds a forest of decision trees, this makes the algorithm stable against statistical fluctuations.

Boosting grows a forest of trees according to this mathematical relation

\[ y^{(t)} = y^{(t-1)} + \epsilon f_t(x_i) \text{ where } \epsilon \text{ is the shrinkage} \]  

Using a square loss function result in a gradient boosted algorithm

Event classification is done on majority vote
160 < mllγ < 400 for the background
Number of Bjets = 0
E_{T}^{miss,sig} \geq 2.5
Three classifiers with 0, 1, and more than 2 central jets will be trained.
Description of the Input Variables

- **mu**: Average bunch crossing
- **dphifjmet**: The angular distance $\Delta(\phi)$ between forward jets and MET
- **dphisjmet**: The angular distance $\Delta(\phi)$ between soft jets and MET
- **ssumpt2**: Subleading vertex sum $\text{Pt}$ squared
- **djpt**: Scalar difference between the vectorial sum $\text{Pt}$ of all the jets and leading one/two vectorial sum $\text{Pt}$.
- **dsumpt2**: Difference between leading and subleading vertex sum $\text{Pt}$ squared
- **dphirefjetmet**: The angular distance $\Delta(\phi)$ between vectorial sum $\text{Pt}$ of all jets and MET
Distributions of the Input Variables - Signal and Background Samples

**ATLAS Preliminary**

\( \sqrt{s} = 13\,\text{TeV} : \int L dt = 44.6\,\text{fb}^{-1} \)

\( n_{c,j} \geq 2 \)

- **Leading Vertex** $\sum P_T^2$
- **Subleading Vertex** $\sum P_T^2$

- **Events**

---

- **All jets** $P_T$
- **Leading jet** $P_T$

---

- **(refjet, met)** $\phi$, $\Delta$

---

**Events**

- **qqZH**
- **Background**
Distributions of the Input Variables - Signal and Background Samples

**ATLAS Preliminary**

\[ \sqrt{s} = 13 \text{TeV} : \int \mathcal{L} dt = 44.6 \text{ fb}^{-1} \]

\[ n_{c_j} \geq 2 \]

Events

\[ \Delta \phi(fj, \text{met}) \]

\[ \Delta \phi(sj, \text{met}) \]

\[ \mu \]

Events
A Gradient boosted decision tree classifiers with the following parameters were trained,

<table>
<thead>
<tr>
<th>Parameter</th>
<th>ncj = 0</th>
<th>ncj = 1</th>
<th>ncj ≥ 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ntrees</td>
<td>2000</td>
<td>800</td>
<td>1000</td>
</tr>
<tr>
<td>MinNodeSize</td>
<td>2%</td>
<td>2%</td>
<td>2%</td>
</tr>
<tr>
<td>Shrinkage</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
</tr>
<tr>
<td>BaggedSampleFraction</td>
<td>65 %</td>
<td>70 %</td>
<td>60 %</td>
</tr>
<tr>
<td>MaxDepth</td>
<td>4</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>nCuts</td>
<td>200</td>
<td>200</td>
<td>200</td>
</tr>
</tbody>
</table>

Events with negative weights are ignored during training
The improvement on classification efficiency is given by,

\[ \text{Improvement} \equiv \frac{S}{\sqrt{S + B}} \]  \hspace{1cm} (2)
Receiver Operating Characteristic (ROC) Curve

![Receiver Operating Characteristic (ROC) Curve](image)

- $ncj = 0$
- $ncj = 1$
- $ncj \geq 2$
Conclusions

- Variables that are not physics dependent have been selected
- Selected variables have a very little correlation with each other
- The classifier performance is better than traditional paper cuts
- The BDT outputs show little hints of overtraining
The End
Backup
Backup

**ATLAS** Preliminary

$\sqrt{s}=13\text{TeV}$ : $\int Ldt=44.6\text{ fb}^{-1}$

$n_{c_j} = 0$

Events

$n_{c_j} = 1$

Events

**ATLAS** Preliminary

$\sqrt{s}=13\text{TeV}$ : $\int Ldt=44.6\text{ fb}^{-1}$

$n_{c_j} = 0$

Events

$n_{c_j} = 1$

Events