Cornering charming Higgs decays

Joseph Walker, Frank Krauss

Machine Learning for Jets Physics 2020
New York University
17th January 2020
Contents

• Why look at H -> cc
• Processes considered
• Data collection
• Observables, Images and Particle Flows
• Training case study
• FIFO Rivet <-> Python
• Network Voting
• Results and outlook
Why H -> cc

Can we pin down the yukawa coupling for light fermions?

\[ \kappa_c = \frac{y}{y_{SM}} \]

Current upper bound from CMS [1]

\[ \frac{\sigma(VH)Br(H \rightarrow c\bar{c})}{\sigma_{SM}(VH)Br_{SM}(H \rightarrow c\bar{c})} = \kappa_c^2 < 70(37^{+16}_{-10}) \]

Charming higgs decays dominated by huge backgrounds

Signals

We consider three signal processes

Vector Boson Fusion

W Higgs-Strahlung

Z Higgs-Strahlung

This gives us 3 channels;

0 Isolated Leptons

1 Isolated Leptons

2 Isolated Leptons
Signals

We consider three signal processes

Vector Boson Fusion

W Higgs-Strahlung

Z Higgs-Strahlung

This gives us 3 channels;

0 Isolated Leptons
1 Isolated Leptons
2 Isolated Leptons

Using fat jets as a gateway to background elimination
In total 19 background runcards are used:
- Higgs couplings turned off
- MEPS(2)
- Decay of one vector boson is enforced in the hard matrix element
Data

Data is generated using Sherpa 2.7.7 at MEPS(2) for a selection of Backgrounds. Analysis performed by Rivet 2.7.0 and Fastjets 3.3.2

Many Runcards are used to boost the efficiency of backgrounds being generated.

We find fatjets with $R=1$ with the Anti-Kt algorithm with $p_T > 250$ GeV

We apply a set of simple cuts to identify candidate fatjets

All surviving events are analysed and three data structures are saved;
1. Observables [2]
2. Jet Images [3]

Cutflows

We have 3 cut flows / 3 channels.

0. Eta and pT cuts on event

1. A Missing Transverse Energy condition
   0: < 10 GeV
   1: > 80 GeV
   2: < 10 GeV

2. Demand N isolated leptons

3. Demand at least 1 R = 1.0 fatjet, pT > 250 GeV

4. Pick fatjet with suitable properties
   a. C tagged
   b. |JetMass - 125 GeV| < 30 GeV

5. Demand 2 R = 0.4 jets OR
   Demand Isolated leptons and candidate jet back-to-back.

Cuts based on arXiv:1912.01662
Example accepted event

Two fat jets observed

Cuts Identify best candidate

[5] pyjets
### Chosen Processes

We collect data from any process with an expectation of surviving events for $\int \mathcal{L} dt = 3000 fb^{-1}$

<table>
<thead>
<tr>
<th>0 Isolated Leptons</th>
<th>1 Isolated Leptons</th>
<th>2 Isolated Leptons</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. VBF (11%)</td>
<td>1. W Higgs-Strahlung (10%)</td>
<td>1. Z Higgs-Strahlung (17%)</td>
</tr>
<tr>
<td>2. WZ[qq] (0.90%)</td>
<td>2. t[bw]t[bw] (0.016%)</td>
<td>2. WZ[ll] (0.16%)</td>
</tr>
<tr>
<td>3. WW[qq] (0.14%)</td>
<td>3. WW[lv] (0.0098%)</td>
<td>3. Z[ll] (0.0020%)</td>
</tr>
<tr>
<td>4. t[bw]t[bw] (0.017%)</td>
<td>4. W[lv]Z (0.0057%)</td>
<td>4. Z[qq] (0.0020%)</td>
</tr>
<tr>
<td>5. WZ[ll] (0.012%)</td>
<td>5. W[lv] (0.0006%)</td>
<td></td>
</tr>
<tr>
<td>6. W[qq] (0.0034%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Z[ll] (0.0002%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. Z[qq] (0.0001%)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

These runcards are boosted to generate balanced classes.
# Chosen Processes - Cross sections

We collect data from any process with an expectation of surviving events for $\int \mathcal{L} dt = 3000 fb^{-1}$

<table>
<thead>
<tr>
<th>0 Isolated Leptons</th>
<th>1 Isolated Leptons</th>
<th>2 Isolated Leptons</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. VBF 0.19(4) pb</td>
<td>1. $W$ Higgs-Strahlung 0.008(2) pb</td>
<td>1. $Z$ Higgs-Strahlung 0.00123(4) pb</td>
</tr>
<tr>
<td>2. $WZ[qq]$ 20(1) pb</td>
<td>2. $t[bw]t[bw]$ 5.4(7) pb</td>
<td>2. $WZ[ll]$ 1.8(1) pb</td>
</tr>
<tr>
<td>3. $WW[qq]$ 0.12(2) nb</td>
<td>3. $WW[lv]$ 31(4) pb</td>
<td>3. $Z[ll]$ 3.3(1) nb</td>
</tr>
<tr>
<td>5. $WZ[ll]$ 1.8(1) pb</td>
<td>5. $W[lv]$ 29(2) nb</td>
<td></td>
</tr>
<tr>
<td>6. $W[qq]$ 90(5) nb</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. $Z[ll]$ 3.3(1) nb</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. $Z[qq]$ 27(2) nb</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$$\frac{S}{B} \simeq 10^{-6} \quad \frac{S}{B} \simeq 10^{-7} \quad \frac{S}{B} \simeq 10^{-7}$$

LO Cross sections!
Example accepted event
11 Chosen observables

1. Number of charged particles
2. Energy in Charged Particles
3. Jet mass
4. Momentum perpendicular to jet axis
5. Angularities (a = -2)
6. Planar Flow
7. Ratio of 1 and 2subjettiness
8. Subjet energy balance
9. D(2->1) / PT
10. D(3->2)/D(2->1)
11. Missing Transverse momentum

Selected based on individual classifying power.

~500 Trainable Parameters
Example accepted event
Example accepted event

Create RGB images

red ~ Energy
green ~ Perpendicular momentum to jet momentum
blue ~ Charge particle multiplicity

Jet axis lined up with (0,0) pixel
Sum up variables in each bin then rescale

~15k Trainable Parameters
Example accepted event

Cornering charming Higgs decays
Example accepted event

Particles in jet ordered in decreasing energy

Form a sequence of 10 particles to be fed into a particle flow network.

1. $\Delta \eta$
2. $\Delta \phi$
3. $\log(p_T)$
4. $\log(E)$
5. $\log(p_T/p_{T,jet})$
6. $\log(E/E_{jet})$
7. $\Delta R$
8. PID

~1.6k Trainable Parameters
The pipeline

1. Determine appropriate backgrounds for each channel
2. Boost events to give balanced classes ~10k samples
3. Normalise and rescale data
4. Learn in tensorflow
Time to learn - 2 Lepton Channel

Observables

![ROC curve for different observables and AUC values](image)

- BackgroundWZ[II] AUC: 0.99
- BackgroundZ[qq] AUC: 0.90
- SignalZH AUC: 0.98
- BackgroundZ[II] AUC: 0.91
- Naive
Time to learn - 2 Lepton Channel

Observables

- Rejection efficiency $1/\varepsilon_B$ vs Tagging efficiency $\varepsilon_S$

- Confusion matrix, without normalization:
  - Background: 3693
  - Signal: 394
  - Predicted label:
    - 1706
  - Actual label:
    - 520

17/01/2020
Time to learn - 2 Lepton Channel

Cornering charming Higgs decays
Time to learn - 2 Lepton Channel

Images

Flow

Cornering charming Higgs decays
Sherpa -> Rivet Analysis

---

```python
# Reduce max for 2.3 j j W-[Q, Q] j to 0.759666 ( eps = 0.001 )
Read in channels from directory: Results/Comix/MC_2_3 j j W-[Q, Q] j
Process Group: CalculateTotalXSec(): Calculate xs for '2.3 j j W-[Q, Q] j' (Comix)
2.3 j j W-[Q, Q] j: 4244.64 pb +/- ( 10.6767 pb = 0.237399 % ) exp. eff: 0.58381
 Reduce max for 2.3 j j W-[Q, Q] j to 0.593677 ( eps = 0.001 )
-- SHERPA generates events with the following structure --

Perturbative:
  : Signal_Processes
  : Hard_Decays
  : Jet_Evolution:CSS
  : Lepton_FS_QED_Corrections:Photons
  : Multiple_Interactions:None
  : Minimum_Bias:Off
  : Hadronization:Beam_Remnants
  : Hadronization:Hadronic
  : Hadron_Decays
  : Unknown:

Analysis:
  : Rivet

Rivet.Analysis.Handler: WARN Analysis 'FatJets0Lep' is unvalidated: be careful, it may be broken!
Rivet.Analysis.Handler: WARN Analysis 'FatJets1Lep' is unvalidated: be careful, it may be broken!
Rivet.Analysis.Handler: WARN Analysis 'FatJets2Lep' is unvalidated: be careful, it may be broken!

---

# FastJet release 3.3.1
# M. Cacciari, G.P. Salam and G. Soyez
# A software package for jet finding and analysis at colliders
# http://fastjet.fr
#
# Please cite EPJC72(2012)1896 [arXiv:1111.6697] if you use this package
# FastJet is provided without warranty under the terms of the GNU GPLv2.
# It uses T. Chan's closest pair algorithm, S. Fortune's Voronoi code
# and 3rd party plugin jet algorithms. See COPYING file for details.
#
Event 200 ( 1 s elapsed / 5 s left ) -> ETA: Sun Jan 05 15:26
XS = 107484 pb +/- ( 10097 pb = 9.39 % )
```

---

Cornering charming Higgs decays

17/01/2020
Voting

Each network returns a probability for each class.

We have a naive voting system, but can we do better?
Voting

The networks are providing complementary information! More classifying power!

![Graphs showing complementary information in particle and observable certainties](image-url)
The networks are providing complementary information! More classifying power!

Only a few 1K data points (for now!)
Time to learn

Learn how to vote!

![ROC curve diagram with AUC of 0.93 for background and signal, and a naive baseline with AUC of 0.93.](image)
Results and Conclusions

Combined networks give superior killing power than individual networks!

\[
\frac{S}{B} \sim 10^{-2}
\]

\[
\frac{S}{B} \sim 10^{-2}
\]

\[
\frac{S}{B} \sim 10^{-3}
\]
Results and Conclusions

\[
\frac{\sigma(VH)Br(H \rightarrow c\bar{c})}{\sigma_{SM}(VH)Br_{SM}(H \rightarrow c\bar{c})} = \kappa_c^2
\]

\[
Br_{SM}(H \rightarrow c\bar{c}) = 2.89 \times 10^{-2} \pm 5.5\% -2.0\%
\]
Results and Conclusions

Cornering charming Higgs decays
Results and Conclusions
Results and Conclusions

Key points
1. A proof of concept
2. Promising results
3. Pipeline ready to go!

Considerations
1. This was only a LO study
2. Possibility to consider more backgrounds
3. Detector simulation
4. More data!

![Graphs showing cuts only and cuts with ML]
Backups
1. Network Architectures
2. Chosen Observables
3. Distributions
## Architectures

<table>
<thead>
<tr>
<th>Architecure</th>
<th>Input</th>
<th>Output</th>
</tr>
</thead>
<tbody>
<tr>
<td>input_1: InputLayer</td>
<td>(None, 11)</td>
<td>(None, 11)</td>
</tr>
<tr>
<td>Dense_1: Dense</td>
<td>(None, 11)</td>
<td>(None, 10)</td>
</tr>
<tr>
<td>Dense_2: Dense</td>
<td>(None, 10)</td>
<td>(None, 10)</td>
</tr>
<tr>
<td>Drop_1: Dropout</td>
<td>(None, 10)</td>
<td>(None, 10)</td>
</tr>
<tr>
<td>Dense_3: Dense</td>
<td>(None, 10)</td>
<td>(None, 10)</td>
</tr>
<tr>
<td>Dense_4: Dense</td>
<td>(None, 10)</td>
<td>(None, 10)</td>
</tr>
<tr>
<td>Predictions: Dense</td>
<td>(None, 10)</td>
<td>(None, 2)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Architecure</th>
<th>Input</th>
<th>Output</th>
</tr>
</thead>
<tbody>
<tr>
<td>input_2: InputLayer</td>
<td>(None, 10, 5)</td>
<td>(None, 10, 5)</td>
</tr>
<tr>
<td>Conv2D_1: Conv2D</td>
<td>(None, 21, 21, 3)</td>
<td>(None, 16, 16, 16)</td>
</tr>
<tr>
<td>max_pooling2d_1: MaxPooling2D</td>
<td>(None, 16, 16, 16)</td>
<td>(None, 8, 8, 16)</td>
</tr>
<tr>
<td>Drop_1: Dropout</td>
<td>(None, 8, 8, 16)</td>
<td>(None, 8, 8, 16)</td>
</tr>
<tr>
<td>Conv2D_2: Conv2D</td>
<td>(None, 8, 8, 16)</td>
<td>(None, 6, 6, 32)</td>
</tr>
<tr>
<td>max_pooling2d_2: MaxPooling2D</td>
<td>(None, 6, 6, 32)</td>
<td>(None, 3, 3, 32)</td>
</tr>
<tr>
<td>Drop_2: Dropout</td>
<td>(None, 3, 3, 32)</td>
<td>(None, 3, 3, 32)</td>
</tr>
<tr>
<td>Conv2D_3: Conv2D</td>
<td>(None, 3, 3, 32)</td>
<td>(None, 2, 2, 64)</td>
</tr>
<tr>
<td>max_pooling2d_3: MaxPooling2D</td>
<td>(None, 2, 2, 64)</td>
<td>(None, 1, 1, 64)</td>
</tr>
<tr>
<td>Drop_2: Dropout</td>
<td>(None, 1, 1, 64)</td>
<td>(None, 1, 1, 64)</td>
</tr>
<tr>
<td>Dense_1: Dense</td>
<td>(None, 1, 1, 64)</td>
<td>(None, 64)</td>
</tr>
<tr>
<td>Flatten_1: Flatten</td>
<td>(None, 64)</td>
<td>(None, 1)</td>
</tr>
<tr>
<td>Dense_2: Dense</td>
<td>(None, 64)</td>
<td>(None, 10)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Architecure</th>
<th>Input</th>
<th>Output</th>
</tr>
</thead>
<tbody>
<tr>
<td>LSTM_1: LSTM</td>
<td>(None, 10, 5)</td>
<td>(None, 10, 10)</td>
</tr>
<tr>
<td>LSTM_2: LSTM</td>
<td>(None, 10, 10)</td>
<td>(None, 10)</td>
</tr>
<tr>
<td>Dense_1: Dense</td>
<td>(None, 10)</td>
<td>(None, 10)</td>
</tr>
<tr>
<td>Predictions: Dense</td>
<td>(None, 10)</td>
<td>(None, 6)</td>
</tr>
</tbody>
</table>
Chosen Observables

A good observable: Jet Mass

A bad observable: Jet Rapidity
Chosen Observables

A good observable: Jet Mass

A bad observable: Jet Rapidity
Distributions

Sanity checks - 2 Isolated lepton channel
Candidate fatjet distributions

Leptons invariant mass