Search for Single Top Quark Production Using Bayesian Neural Networks

Daekwang Kau
Florida State University
May 17, 2007
BNL
Outline

• The Tevatron and DØ Detector
• The Top Quark
• Single Top Quark Production
• Backgrounds
• Event Selection
• Analysis Strategy
• Neural Networks and Bayesian Neural Networks
• Cross Section Measurement
• Conclusions
The Tevatron and DØ Detector

- Collide protons and antiprotons at $\sqrt{s} = 1.96$ TeV
- The only place to directly study the top quark
- Used 0.9 fb$^{-1}$ of data
The Top Quark

- Top quark was discovered in 1995 at the Tevatron by DØ and CDF
- Heaviest quark
  
  \[ M_{\text{top}} = 171.4 \pm 2.1 \text{GeV} \]

- Short lifetime
  
  \[ \sim 0.4 \times 10^{-24} \text{s} \]

  Daughter particles retain spin information

- Main production mode is QCD pair production

  - 85\% \( q\bar{q} \) annihilation,
  - 15\% \( gg \) fusion

- Cross section is 6.67 pb at cms energy = 1.96 TeV
Final states of signal include:

**s-channel**
- One high $p_T$ lepton, $E_T$,
- 2 high $p_T$ b-jets

**t-channel**
- One high $p_T$ lepton, $E_T$,
- 1 high $p_T$ b-jet, high $p_T$ light jet

\[ \sigma_{s\text{-channel}} = 0.88 \text{pb} \pm 8\% \]
\[ \sigma_{t\text{-channel}} = 1.98 \text{pb} \pm 25\% \]

Zack Sullivan et al. PRD 66 (02) 054024

Signal MC samples are generated with CompHEP generator
Partons are hadronized using Pythia
Why Is Single Top Interesting?

- Standard Model predicts single top production
- W-t-b vertex in production
  - Direct access to $|V_{tb}|$
  - Test unitary of CKM matrix
- Sensitive to new physics
  - s-channel: heavy $W'$ boson resonance
  - t-channel: flavor changing neutral currents
  - 4th generation of quarks
**Backgrounds**

**W+Jets**
- Contain a lepton from W decay
- Estimate shapes from MC
- Normalize to data

**Top-pair production**
- Lepton + jets
  - One W decays hadronically
  - One W decays leptonically
- Dilepton
  - Both Ws decay leptonically
- Estimate from MC

**Multijet events**
- Contain fake isolated lepton
- Estimate from data
Event Selection

• **Purpose**
  - Select W events
  - Maximize acceptance for the signal

• **Requirements**
  - 1 isolated lepton with $p_T > 15$ GeV
  - $E_T > 15$ GeV
  - Jets: $p_T$(jet1) > 25 GeV, $|\eta| < 2.5$, $p_T$(jet2) > 20 GeV, $|\eta| < 3.4$,
    $p_T$(others) > 15 GeV, $|\eta| < 3.4$

• **B-tagging**
  - One or two jets tagged as b-jets
Event Selection

Divide the dataset into independent channels

Fractions of expected signal and S:B ratios are significantly different

<table>
<thead>
<tr>
<th>Electron + Muon</th>
<th>1 jet</th>
<th>2 jets</th>
<th>3 jets</th>
<th>4 jets</th>
<th>≥ 5 jets</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 tags</td>
<td>10%</td>
<td>25%</td>
<td>12%</td>
<td>3%</td>
<td>1%</td>
</tr>
<tr>
<td></td>
<td>1:3,200</td>
<td>1:390</td>
<td>1:300</td>
<td>1:270</td>
<td>1:230</td>
</tr>
<tr>
<td>1 tag</td>
<td>5%</td>
<td>21%</td>
<td>11%</td>
<td>3%</td>
<td>1%</td>
</tr>
<tr>
<td></td>
<td>1:100</td>
<td>1:20</td>
<td>1:25</td>
<td>1:40</td>
<td>1:53</td>
</tr>
<tr>
<td>2 tags</td>
<td>3%</td>
<td>2%</td>
<td>1%</td>
<td>0%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1:11</td>
<td>1:15</td>
<td>1:38</td>
<td>1:43</td>
<td></td>
</tr>
</tbody>
</table>

S:B ratio is at most 9%

More sophisticated techniques are required for further signal and background discrimination
Event Selection

W transverse mass (muon)

(light) blue : signal
green: W+jets
red: ttbar
gray: multijet

1 tag

2 tags

2 jets

3 jets

4 jets
Analysis Strategy

Event selection (12 channels)

- Decision trees (DT)
- Bayesian neural networks (BNN)
- Matrix element (ME)

Selecting variables

Training BNN

Cross section measurement
The Main Idea of Neural Networks

Goal: Given data, $x$, achieve best separation between signal and background events

Requires computing

$$P(\text{signal} \mid x) = \frac{P(x \mid \text{signal})P(\text{signal})}{P(x \mid \text{signal})P(\text{signal}) + P(x \mid \text{background})P(\text{background})}$$

Use the discriminant

$$D(x) = \frac{P(x \mid \text{signal})}{P(x \mid \text{signal}) + P(x \mid \text{background})}$$
Neural Networks

• Neural networks are non-linear functions that can map a vector of n real-valued inputs

\[ X = (x_1, x_2, \ldots, x_n), t \]

into one output

\[ y(X; W) \]

• Input vector X is assigned to either signal or background

• Target t is a binary classification label (1,0)

  1: signal

  0: background

• Estimate parameter (network weight) W
Neural Networks

Example)

\[ X_1 = (x_1, x_2, x_3), 1 \]
\[ X_2 = (x_1, x_2, x_3), 1 \]
\[ X_1 = (x_1, x_2, x_3), 0 \]
\[ X_2 = (x_1, x_2, x_3), 0 \]

\[ X_m = (x_1, x_2, x_3), 1 \]

\[ X_m = (x_1, x_2, x_3), 0 \]

2m events
(signal + background)

Neural Networks

\[ y(X; W) \]
Neural Networks

Example)

\[ X_1 = (x_1, x_2, x_3), 1 \]
\[ X_2 = (x_1, x_2, x_3), 1 \]
\[ \ldots \]
\[ X_m = (x_1, x_2, x_3), 1 \]

\[ X_1 = (x_1, x_2, x_3), 0 \]
\[ X_2 = (x_1, x_2, x_3), 0 \]
\[ \ldots \]
\[ X_m = (x_1, x_2, x_3), 0 \]

- W starts from some arbitrary values
- Outputs are computed for event 1 to event 2m
- Error function is calculated

\[ Err = \frac{1}{2m} \sum_{k=1}^{2m} (y_k - t_k)^2 \]

- W is updated to minimize Err
- 1 cycle is called 1 epoch

<table>
<thead>
<tr>
<th>target</th>
<th>output</th>
</tr>
</thead>
<tbody>
<tr>
<td>( t_1 )</td>
<td>1</td>
</tr>
<tr>
<td>( t_2 )</td>
<td>0</td>
</tr>
<tr>
<td>( \ldots )</td>
<td>( \ldots )</td>
</tr>
<tr>
<td>( t_{2m} )</td>
<td>1</td>
</tr>
</tbody>
</table>
Neural Networks

**Goal:** produce a system which makes good predictions for new data

Exhibits good generalization

Prepare independent data (testing data)

Apply the neural network function to the testing data and calculate the error function

To prevent neural networks from over-fitting or over-training, stop training when error function values of testing set starts increasing

\[
Err = \frac{1}{2m} \sum_{k=1}^{2m} (y_k - t_k)^2
\]
Neural Networks

**Training data**
- Blue: generalize to testing data
- Red: perfect fit to training data but poor representation of the testing data

**Testing data**
Bayesian Neural Networks

Find posterior probability density over all possible weight sets

\[ P(W|t) = \frac{P(t|W)P(W)}{P(t)} \]

To make prediction for new events

\[ \bar{y}(X) = \int y(X; W)P(W|t)\,dW \]

\[ \approx \frac{1}{M} \sum_{i=1}^{M} y(X; W_i) \]
Bayesian Neural Networks

\[ P(W|t) = \frac{P(t|W)P(W)}{P(t)} \]

Not possible to calculate the posterior density analytically

Draw a sample of points \( W_i \) from the posterior density \( P(W|t) \) using a Markov Chain Monte Carlo method
Selecting Variables

Construct 50 input variables based on three categories

Single Object Kinematics

- Leading, second Jet
- Tagged, untagged Jet

Angular Variables

- $\Delta R$ (jet1,jet2)
- $\cos$(alljets,Tagged Jet)

Global Event Variables

- Invariant mass top quark
- Center of mass energy ($\hat{s}$)
Selecting Variables

Check modeling of variables

Compute distribution of a discrepancy measure similar to K-S test

Order the variables according to their importance

Build discrimination rules using sequence of if statements and measure how often a particular variable appears

Select best 20 variables out of 50
Selecting Variables

**e+2jets/1tag**

![Graph showing variable selection for e+2jets/1tag](image1)

**mu+2jets/1tag**

![Graph showing variable selection for mu+2jets/1tag](image2)
Training

- Number of inputs \( \sim 20 \) and the number of hidden nodes = 20
- Training sample: 10,000 signal + 10,000 background
BNN Outputs (Electron)

1tag, 2jet

1tag, 3jet

1tag, 4jet

2tag, 2jet

2tag, 3jet

2tag, 4jet
BNN Outputs (Muon)

1tag, 2jet

1tag, 3jet

1tag, 4jet

2tag, 2jet

2tag, 3jet

2tag, 4jet
BNN outputs

Sum of all 12 channels

Near a BNN output of 1
Cross Section Measurement

Bayesian posterior probability density

\[ \text{Posterior}(d \mid D) \equiv P(\sigma, a, b \mid D) \propto \text{likelihood}(D \mid d) \text{prior}(d) \]

\[ d = \varepsilon L \sigma + \sum_{i=1}^{n} b_i \equiv a \sigma + \sum_{i=1}^{n} b_i \]

\[ P(D \mid d) = P(D \mid \sigma, a, b) = \prod_{j=1}^{\text{bins}} \frac{\exp(-d_j) d_j^D_j}{\Gamma(D_j + 1)} \]

\[ P(\sigma \mid D) \propto \int\int \text{likelihood}(D \mid d) \text{prior}(d) dadb \]

Final result

\[ \sigma_{s+t} = 4.4^{+1.6}_{-1.4} \text{ pb} \]
Significance of Signal

Generate 72000 background only ensemble

Test static: cross section

p-value: probability to measure a cross section equal to or higher than reference value
Significance of Signal

The p-value computed from the SM signal + background ensemble (y-axis) versus the p-value computed from the background-only ensemble (x-axis)
Conclusions

We measure a single top quark production cross section of $4.4^{+1.6}_{-1.4}$ pb

This analysis results in a p-value of 0.08%, corresponding to a 3.1 standard deviation significance

$$\sigma_{s+t} = 4.4^{+1.6}_{-1.4} \text{ pb}$$

3.1 standard deviation significance
Back Up