Deep Learning in High Energy Physics

Improving the Search for Exotic Particles

P. Baldi, P. Sadowski, and D. Whiteson

Department of Computer Science
Department of Physics
Center for Machine Learning and Intelligent Systems
Deep Learning in HEP

Peter Sadowski

Daniel Whiteson
Machine Learning (DL) in the Natural Sciences

• Physics
  -HEP
  -QM
  -Astronomy

• Chemistry
  -Prediction of Molecular or Material Properties
  -Prediction of Chemical Reactions

• Earth Sciences

• Biology
  -Prediction of Protein Structures
  -Prediction of gene Expression
  -Biomedical imaging
Deep Learning in Chemistry

CC(=O)O

Acetic Acid
Deep Learning Chemical Reactions

\[ \text{RCH}=\text{CH}_2 + \text{HBr} \rightarrow \text{RCH(Br)}–\text{CH}_3 \]


Deep Learning in Biology
Deep Learning in Biology: Mining Omic Data

Deep Learning in Biology: Mining Omic Data

Deep Learning

Machine Learning (DL) in the Natural Sciences

- Physics
  - HEP
  - QM
  - Astronomy
- Chemistry
  - Prediction of Molecular or Material Properties
  - Prediction of Chemical Reactions
- Earth Sciences
- Biology
  - Prediction of Protein Structures
  - Prediction of gene Expression
  - Biomedical imaging
Deep Learning in HEP

- Higgs Boson Detection
- Supersymmetry
- Higgs To Tau Tau Decay
Searching for exotic particles in high-energy physics with deep learning

P. Baldi, P. Sadowski & D. Whiteson

Collisions at high-energy particle colliders are a traditionally fruitful source of exotic particle discoveries. Finding these rare particles requires solving difficult signal-versus-background classification problems, hence machine-learning approaches are often used. Standard approaches have relied on ‘shallow’ machine-learning models that have a limited capacity to learn complex nonlinear functions of the inputs, and rely on a painstaking search through manually constructed nonlinear features. Progress on this problem has slowed, as a variety of techniques have shown equivalent performance. Recent advances in the field of deep learning make it possible to learn more complex functions and better discriminate between signal and background classes. Here, using benchmark data sets, we show that deep-learning methods need no manually constructed inputs and yet improve the classification metric by as much as 8% over the best current approaches. This demonstrates that deep-learning approaches can improve the power of collider searches for exotic particles.
Deep Learning in HEP

- Higgs Boson Detection (NC 2014)
- Supersymmetry (NC 2014)
- Higgs to Tau Tau Decay (NIPS 2014)
Deep Learning in HEP

• Higgs Boson Detection
• Supersymmetry
• Higgs Decay

• Common Features and Results:
  - dozens of features: raw + human-derived
  - millions of examples
  - classification problems
  - deep learning outperforms current methods (e.g. AUC)
  - deep learning can work without human-derived features
Machine Learning in HEP
| Quark | Mass (MeV/c²) | Charge | Spin | |---|---|---|---| | u | ≈2.3 | 2/3 | 1/2 | | c | ≈1.275 | 2/3 | 1/2 | | t | ≈173.07 | 2/3 | 1/2 | | g | 0 | 0 | 0 | | H | ≈126 | 0 | 0 | | Quarks | | | | | Higgs boson | | Photon | | Leptons | | Electron | 0.511 | -1 | 1/2 | | Muon | 105.7 | -1 | 1/2 | | Tau | 1.777 | -1 | 1/2 | | Z boson | 91.2 | 0 | 0 | | Gauge Bosons | | | | | W boson | 80.4 | ±1 | 1 | | Neutrinos | | | | | Electron neutrino | <2.2 | 0 | 1/2 | | Muon neutrino | <0.17 | 0 | 1/2 | | Tau neutrino | <15.5 | 0 | 1/2 |
Standard Model Interactions
(Forces Mediated by Gauge Bosons)

- X is any fermion in the Standard Model.
- X is electrically charged.
- X is any quark.
- U is a up-type quark; D is a down-type quark.
- L is a lepton and \( \nu \) is the corresponding neutrino.
- X is a photon or Z-boson.
- X and Y are any two electroweak bosons such that charge is conserved.
Higgs Boson Detection

Simulation tools:
- MadGraph (collisions)
- PYTHIA (showowering and hadronization)
- DELPHES (detector response)

Higgs boson decay signal

Background process

11 M examples
Higgs Boson Detection

Supervised learning problem:

- Two classes
- 11 million training examples (roughly balanced)
- 28 features
  - 21 low-level features (momenta of particles)
  - 7 high-level features derived by physicists

Data available at archive.ics.uci.edu/ml/datasets/HIGGS
Higgs Boson Detection

21 low-level features:
- 3D momentum for observed particles
- Missing transverse momentum
- Jets and $b$-tagging information

7 high-level features:
- Reconstruction of invariant masses for each particle subset.
Higgs Boson Detection

Tuning deep neural network architectures.

<table>
<thead>
<tr>
<th>Hyper parameters</th>
<th>Choices</th>
</tr>
</thead>
<tbody>
<tr>
<td>Depth</td>
<td>2,3,4,5,6 layers</td>
</tr>
<tr>
<td>Hidden units per layer</td>
<td>100,200,300,500</td>
</tr>
<tr>
<td>Learning rate</td>
<td>0.01, 0.05</td>
</tr>
<tr>
<td>Weight decay</td>
<td>0, 0.00001</td>
</tr>
<tr>
<td>Pre-training</td>
<td>none, autoencoder, multi-task autoencoder</td>
</tr>
<tr>
<td>Input features</td>
<td>low-level, high-level, complete set</td>
</tr>
</tbody>
</table>

Best:
- 5 hidden layers
- 300 neurons per layer
- Tanh hidden units, sigmoid output
- No pre-training
- Stochastic gradient descent
- Mini batches of 100
- Exponentially-decreasing learning rate
- Momentum increasing from .5 to .99 over 200 epochs
- Weight decay = 0.00001
Higgs Boson Detection

Deep network improves AUC by 8%

<table>
<thead>
<tr>
<th>Technique</th>
<th>Low-level</th>
<th>High-level</th>
<th>Complete</th>
</tr>
</thead>
<tbody>
<tr>
<td>BDT</td>
<td>0.73</td>
<td>0.78</td>
<td>0.81</td>
</tr>
<tr>
<td>NN</td>
<td>0.733 (0.007)</td>
<td>0.777 (0.001)</td>
<td>0.816 (0.004)</td>
</tr>
<tr>
<td>DN</td>
<td>0.880 (0.001)</td>
<td>0.800 (&lt; 0.001)</td>
<td>0.885 (0.002)</td>
</tr>
</tbody>
</table>

**BDT** = Bayesian Decision Trees in TMVA package

*Nature Communications, July 2014*
### Higgs Boson Detection

<table>
<thead>
<tr>
<th>Technique</th>
<th>Low-level</th>
<th>High-level</th>
<th>Complete</th>
</tr>
</thead>
<tbody>
<tr>
<td>NN 300-hidden</td>
<td>0.733</td>
<td>0.777</td>
<td>0.816</td>
</tr>
<tr>
<td>NN 1000-hidden</td>
<td>0.788</td>
<td>0.783</td>
<td>0.841</td>
</tr>
<tr>
<td>NN 2000-hidden</td>
<td>0.787</td>
<td>0.788</td>
<td>0.842</td>
</tr>
<tr>
<td>NN 10000-hidden</td>
<td>0.790</td>
<td>0.789</td>
<td>0.841</td>
</tr>
<tr>
<td>DN 3 layers</td>
<td>0.836</td>
<td>0.791</td>
<td>0.850</td>
</tr>
<tr>
<td>DN 4 layers</td>
<td>0.868</td>
<td>0.797</td>
<td>0.872</td>
</tr>
<tr>
<td>DN 5 layers</td>
<td>0.880</td>
<td>0.800</td>
<td>0.885</td>
</tr>
<tr>
<td>DN 6 layers</td>
<td>0.888</td>
<td>0.799</td>
<td>0.893</td>
</tr>
</tbody>
</table>

DNs have 300 tanh units in each hidden layer.

Deeper networks perform better and performance continues to improve after publication ....
Higgs Boson Detection

Experiment: regression on 7 high-level features

<table>
<thead>
<tr>
<th>Technique</th>
<th>Feature</th>
<th>Regression</th>
<th>MSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear Regression</td>
<td></td>
<td>0.1468</td>
<td></td>
</tr>
<tr>
<td>NN</td>
<td></td>
<td>0.0885</td>
<td></td>
</tr>
<tr>
<td>DN 3 layers</td>
<td></td>
<td>0.0821</td>
<td></td>
</tr>
<tr>
<td>DN 4 layers</td>
<td></td>
<td>0.0818</td>
<td></td>
</tr>
<tr>
<td>DN 5 layers</td>
<td></td>
<td>0.0815</td>
<td></td>
</tr>
<tr>
<td>DN 6 layers</td>
<td></td>
<td>0.0812</td>
<td></td>
</tr>
</tbody>
</table>

Deeper networks better at learning high-level features.
Supersymmetry (SUSY)

Signal

Background

Detect the production of new supersymmetric particles

6 M examples

Data available at archive.ics.uci.edu/ml/datasets/SUSY
Supersymmetry (SUSY)

Deep networks again lead to significant gains.

Data available at archive.ics.uci.edu/ml/datasets/SUSY
Higgs to Tau Tau Decay

Signal

Background
Higgs to Tau Tau Decay

10 low-level features:

- The 3D momenta, $p$, of the charged leptons
- The imbalance of momentum ($p_T'$) in the final state transverse to the beam direction, due to unobserved or mismeasured particles
- The number and momenta of particle ‘jets’ due to radiation of gluons or quarks

15 high-level:

- Axial missing momentum, $p_T' \cdot p_{\ell^+\ell^-}$
- Scalar sum of the observed momenta, $|p_{\ell^+}| + |p_{\ell^-}| + |p_T'| + \sum_i |p_{\text{jet}_i}|$
- Angular distance between leptons, $\Delta R = \sqrt{(\Delta \eta)^2 + (\Delta \phi)^2}$
- Missing mass, $m_{\text{MMC}}$
- Sphericity and transverse sphericity
- Relative missing momentum, $p_T'$ if $\Delta \phi(p,p_T') \geq \pi/2$, and $p_T' \times \sin(\Delta \phi(p,p_T'))$ if $\Delta \phi(p,p_T') < \pi/2$, where $p$ is the momentum of any charged lepton or jet;
- Difference in lepton azimuthal angles, $\Delta \phi(\ell^+, \ell^-)$
- Difference in lepton polar angles, $\Delta \eta(\ell^+, \ell^-)$
- Invariant mass of the two leptons, $m_{\ell^+\ell^-}$
- Invariant mass of all visible objects (leptons and jets).

80M examples
Higgs to Tau Tau Decay

10 low-level features: 15 high-level:

- Lepton 1 $p_T$ (GeV)
- Lepton 1 $\eta$
- Lepton 2 $p_T$ (GeV)
- Lepton 2 $\eta$
- N jets
- Missing Trans. Mom (GeV)
- Sum PT
- $\Delta R(l)$
- $m_h$
- Spher
- MMC
- $m_{h,u}$
Higgs to Tau Tau Decay

Hyper-parameters optimized with Spearmint:

- 100 deep networks trained
- 40M training examples, 100 epochs
- Hyperparameters include depth and width
- Best network:
  - Deepest network available (8 layers)
  - Rectified linear hidden units
  - ~300 units per layer

Spearmint chooses the deepest network.
Higgs to Tau Tau Decay

Optimized shallow net vs optimized deep nets:

<table>
<thead>
<tr>
<th>Technique</th>
<th>Low-level</th>
<th>High-level</th>
<th>Complete</th>
</tr>
</thead>
<tbody>
<tr>
<td>NN</td>
<td>0.789 (0.0010)</td>
<td>0.792 (0.0002)</td>
<td>0.797 (0.0004)</td>
</tr>
<tr>
<td>NN ensemble</td>
<td>0.791</td>
<td>0.793</td>
<td>0.798</td>
</tr>
<tr>
<td>DNN</td>
<td>0.798 (0.0001)</td>
<td>0.798 (0.0001)</td>
<td>0.802 (0.0001)</td>
</tr>
<tr>
<td>DNN ensemble</td>
<td>0.798</td>
<td>0.798</td>
<td>0.803</td>
</tr>
</tbody>
</table>

(1) DNN give significant performance boost
(2) Ensembles give small boost
(3) Slight gap with respect to high-level features remains (the mass of the lepton is included in the high-level features…)

AUC
Higgs to Tau Tau Decay

Improvement translates to 20% reduction in data needed for discovery.
Many Open Directions

• Apply ML to earlier stages of processing ("trigger")
• Model detector signals
• Improve performance
• Apply ML to other exotic particles and theories
• Transfer Learning
• ......Other Natural Sciences
• ......ML
THANK YOU
Higgs to Tau Tau Decay

Optimized shallow net vs optimized deep nets:

(1) DNN give significant performance boost,
(2) Ensembles give small boost
(3) Slight gap with respect to high-level features remains (Mass of lepton is contained in the high-level features)

<table>
<thead>
<tr>
<th>Technique</th>
<th>Low-level</th>
<th>High-level</th>
<th>Complete</th>
</tr>
</thead>
<tbody>
<tr>
<td>NN</td>
<td>2.57σ (0.006)</td>
<td>2.92σ (0.006)</td>
<td>3.02σ (0.008)</td>
</tr>
<tr>
<td>NN ensemble</td>
<td>2.61σ</td>
<td>2.96σ</td>
<td>3.06σ</td>
</tr>
<tr>
<td>DNN</td>
<td>3.16σ (0.003)</td>
<td>3.24σ (0.003)</td>
<td>3.37σ (0.003)</td>
</tr>
<tr>
<td>DNN ensemble</td>
<td>3.18σ</td>
<td>3.26σ</td>
<td>3.39σ</td>
</tr>
</tbody>
</table>
Figure 2 | Low-level input features for Higgs benchmark. Distributions in $hjjb\bar{b}$ events for simulated signal (black) and background (red) benchmark events. Shown are the distributions of transverse momenta ($p_T$) of each observed particle (a-e) as well as the imbalance of momentum in the final state (f). Momentum angular information for each observed particle is also available to the network, but is not shown, as the one-dimensional projections have little information.
The ATLAS Pixel Detector provides a very high granularity, high precision set of measurements as close to the interaction point as possible. The system provides three precision measurements over the full acceptance, and mostly determines the impact parameter resolution and the ability of the Inner Detector to find short lived particles such as B-Hadrons. The system consists of three barrels at average radii of ~ 5 cm, 9 cm, and 12 cm (1456 modules), and three disks on each side, between radii of 9 and 15 cm (288 modules). Each module is 62.4 mm long and 21.4 mm wide, with 46080 pixel elements read out by 16 chips, each serving an array of 18 by 160 pixels. The 80 million pixels cover an area of 1.7 m^2. The readout chips must withstand over 300 kGy of ionising radiation and over 5x10^14 neutrons per cm^2 over ten years of operation. The modules are overlapped on the support structure to give hermetic coverage. The thickness of each layer is expected to be about 2.5% of a radiation length at normal incidence. Typically three pixel layers are crossed by each track. The pixel detector can be installed independently of the other components of the ID. In the starting phase, only two of the three layers planned for will be installed.
Large Hadron Collider
Large Hadron Collider
Large Hadron Collider
Large Hadron Collider