Deep Learning in Particle Physics Research

Game Go 2016

winnings 4 : 1

Prof. Dr. Martin Erdmann, RWTH Aachen University, Rutherford Appleton Laboratory 17-Jan-2018
Topics

Deep neural network
from computer science to particle physics

Network concepts and applications

- convolutional
- fully connected
- recurrent
- autoencoder
- adversarial

Network quality
causality, uncertainty
Deep Learning in industry

humans

machine

SegNet University of Cambridge
Spectacular success

Image recognition challenge

ImageNet: 1000 categories, 1.2 million images

Classification error rate

Deep learning errors < humans

Humans 5%


WMW Jie Hu, Li Shen (Oxford), Gang Sun, 2017
New paradigm in physics education & work?

• **Up to now:** physics laws & mathematics are basis for algorithm development → machine learning with physicist’s favorite observables

• **New:** machines exploit data deeper than physicist’s algorithms so far

  • **We ought to prepare for the fundamental change to include machines in our daily work**

  • **Next:** develop physics motivated network architectures
Calorimeter: cosmic ray induced air showers

Educated physicist:
- Time offset between detectors
- Particle velocity
- Detector distance
- Plane fit of time residuals

shower direction angular resolution

Martin Erdmann, 17-Jan-2018
Calorimeter: cosmic ray induced air showers

Educated physicist:
• Time offset between detectors
• Particle velocity
• Detector distance
• Plane fit of time residuals

RAW input data:

Deep Neural Network

• Time offset only
• Black signal traces added

Sim. signal time traces

shower direction angular resolution

Deep Neural Network learns physics from data within 3h

M.E., J. Glombitza, D. Walz, 10.1016/j.astropartphys.2017.10.006
Calorimeter: cosmic ray induced air showers

Characterization of signal traces

Network extracts from training data optimized intermediate variables suited for shower direction

120k parameters

Time offset & total signal

RAW input data:

Deep Neural Network

- Time offset only
- Black signal traces added

shower direction angular resolution

Deep Neural Network learns physics from data within 3h
Deep Learning

Hierarchy of “features” with increasing level of complexity

Image recognition
- Pixel → edge → texton → motif → part → object

Text
- Character → word → word group → clause → sentence → story

Speech
- Sample → spectral band → sound → … → phone → phoneme → word

Neural Networks: 2 pages reminder
Neural Network Operations

\[ x \quad \text{multi-dimensional input data} \]
\[ W, b \quad \text{to be trained} \]
successively apply 2 operations:

\[ y = Wx + b \]
\[ h = \sigma(y) \]

activation function: departure from linear system

SeLU activation function: departure from linear system
Neural Network Training

- **Data set** \( \{x_i, y_i\} \quad i = 1 \ldots N \)

- **Define model**
  \[ y_m(x; \theta) = Wx + b \text{ with } \theta = (W, b) \]

- **Define objective** function (also called loss/cost)
  \[ J(\theta) = \frac{1}{N} \sum_{i=1}^{N} (y_m(x_i) - y_i)^2 \]

- **Train** the model by optimizing the parameters
  \[ \hat{\theta} = \arg \min J(\theta) \]
Neural Network: automated parameterization of arbitrary function

Good description (grey) of original function (black) after $10^5$ training steps
(if you don’t know the original function: don’t want to invent a suitable parameterization yourself)

T. Quast CERN/RWTH Aachen

Fully connected neural network
7 hidden units with 20 nodes each
ReLU activation function
Deep Learning Concepts

from computer science to particle physics

convolutional  fully connected  recurrent  autoencoder  adversarial
Concept 1: convolutional networks to analyse image-like data

- Slide filter over input volume
- Calculate $W \cdot x + b$
- Apply activation function
- 1 output value at each position
- *Translation invariance*

Electron neutrino identification

neural network neutrino event classifier

$\nu_e \text{ CC}$ | $\nu_e \text{ NC}$

Feature maps

<table>
<thead>
<tr>
<th>track</th>
<th>hadronic</th>
</tr>
</thead>
</table>

Method | $\nu_e$ efficiency (same purity)
---|---
Physicists algorithm | 35%
Deep learning neural network | 49%
Distinguish quark & gluon jets

- Red = $p_T$ charged particles
- Green = $p_T$ neutral particles
- Blue = multiplicity charged particles

Best performance by convolutional network (black)
CMS jet flavor tagging

Convolutional network
Recurrent network
Fully connected
Multi classification

631 input data
≤ 25 charged part. 18
≤ 25 neutral part. 6
≤ 4 sec. vert. 12
≤ 25
≤ 25
≤ 4

250k parameters

Brillant improvement with high impact on physics exploitation of data
Concept 2: *fully connected* networks to analyse fixed length data
Jet-parton assignment in ttH events

- Network distribution approximates the best possible reconstruction
- Best results: low level (4-momenta) + high level (decay angle,…) variables

<table>
<thead>
<tr>
<th>Method</th>
<th>$\epsilon(t\bar{t}H) / %$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\chi^2$ method</td>
<td>37</td>
</tr>
<tr>
<td>Boosted Decision Trees</td>
<td>45</td>
</tr>
<tr>
<td>Deep Learning</td>
<td>52</td>
</tr>
</tbody>
</table>

2,000k parameters

Martin Erdmann, 17-Jan-2018
Top tagging featuring Minkowski space

1) Combinations of particles
\[ \tilde{k}_{\mu,j} = k_{\mu,i} C_{ij} = \begin{pmatrix} E & E & \cdots & E \\ px & px & \cdots & px \\ py & py & \cdots & py \\ pz & pz & \cdots & pz \end{pmatrix} \]
\[ \text{trainable} \begin{pmatrix} \text{single particles} \\ \text{combinations of particles} \end{pmatrix} \]

2) Extract features from particle combinations respecting Minkowski metric
\[ d^2_{jm} = (\tilde{k}_j - \tilde{k}_m)_\mu g^{\mu\nu} (\tilde{k}_j - \tilde{k}_m)_\nu \]
\[ \hat{k}_j = \begin{pmatrix} m^2(\tilde{k}_j) \\ pt(\tilde{k}_j) \\ w_j \sum (E_{\tilde{k}_m}) \end{pmatrix} \]
\[ \text{trainable} \begin{pmatrix} m^2(\tilde{k}_j) \\ pt(\tilde{k}_j) \\ w_j \sum (E_{\tilde{k}_m}) \end{pmatrix} \]

- Include Minkowski metric for 4-vectors
- Let network decide on relevant particle combinations
Concept 3: *recurrent* networks to analyse input data of variable size.
Recurrent networks

Hidden state depends on previous state

\[ h(t) = \tanh(U h(t-1) + W x(t) + b) \]

Efficient parameter sharing: same \( W, U, b \) for all input data

Long Short Term Memory

E.g. detecting correlations between tracks on input far away from each other through a cell memory

Jet formation & fat jet identification

Recurrent network combines jets of event

Fully connected: fat W jet or not

network combines particle 4-momenta to jets (recursive)

30k parameters

Efficiency of finding boosted W fat jet

Physics motivated:
mapping 4-vectors into network to build & process jets
Concept 4: **autoencoder** networks to extract signal from noisy data

Radio signal from cosmic ray air showers

Noise from data + signal from simulation

Input Data
Truth
Reconstructed Signal

Signal energy & frequency spectrum approx. conserved

F. Schlüter, M.E., D. Walz, RWTH Aachen

Martin Erdmann, 17-Jan-2018
• Counterfeiter: shows police real & fake money
• Police: learns to distinguish & gives feedback
• Counterfeiter: new fake money based on feedback
• Iterate until police is fooled

Concept 5: Adversarial network to selectively modify network output
**Principle of Generative Adversarial Networks**

Real world images

Discriminator

Generator

Random input for pixels

Fake

Real

Evaluate quality

1st try

2nd try

4th try

from RWTH course for masterstudents
Deep Learning in Physics Research
Summer term 2017
S. Schipmann, D. Walz, M.E., U. Klemradt

Martin Erdmann, 17-Jan-2018
Calorimeter simulations: CaloGAN

- Simulation time faster by 10,000
- Massive parallel computing architecture required (GPU)
- ACAT 2017: Configurable framework planned for fast simulation → Integration in GEANT V

<table>
<thead>
<tr>
<th>Generation Method</th>
<th>Hardware</th>
<th>milliseconds/shower</th>
</tr>
</thead>
<tbody>
<tr>
<td>GEANT4</td>
<td>CPU</td>
<td>1772</td>
</tr>
<tr>
<td>CALOGAN</td>
<td>CPU</td>
<td>2.03</td>
</tr>
<tr>
<td></td>
<td>GPU</td>
<td>0.012</td>
</tr>
</tbody>
</table>

Electron in central calorimeter

Energy uniform between 1 GeV and 100 GeV
Research on network causality, uncertainty, data labels
Causality

Different methods to investigate network decisions

- **Input**
  - toy terrier (0.86)
  - rapeseed (1.00)
  - mongoose (0.93)
  - mailbox (0.71)

- **Back-projections to input space**

- **Heatmap: influence of pixel on output**

**Important progress opening black box**
Predictive uncertainty estimation

Use ensembles of deep neural networks (deep ensembles)

Simple averaging

Adversarial training

$\mathbf{x}' = \mathbf{x} + \text{perturbation}$

Uncertainty prediction from ensembles of networks improves with dedicated training
Decorrelation

Problem: model jet mass of background processes → decorrelate fat jet & jet mass

classifier: Fat jet or QCD?
adversary: infer jet mass from classification output

Fully connected 3 layer 300 nodes

Fully connected 1 layer 50 nodes

10 jet mass bins

$L_{tagger} = L_{classification} - \lambda L_{adversary}$

Best: Traditional network

Best: Adversarial network

Decorrelation by 2nd network reduces systematic dependency, improves discovery significance
Classification without labels

\[
\begin{align*}
\mathcal{L}_{M_1/M_2} &= \frac{p_{M_1}}{p_{M_2}} = \frac{f_1 p_S + (1 - f_1) p_B}{f_2 p_S + (1 - f_2) p_B} = \frac{f_1 L_{S/B} + (1 - f_1)}{f_2 L_{S/B} + (1 - f_2)}, \quad f_1 \neq f_2
\end{align*}
\]

Step towards training with data

Likelihood

Summary on deep learning

• Machines exploit physics contained in data deeper than before
• Modeling particle physics into deep network architecture
• Investigations of causality, stability, uncertainties