Deep Learning in High Energy Physics

Ontologia, San Sebastian

04.10.2023

Martin L. King
MCMP Ludwigs-Maximilians-University Munich

Setting

1. No indications in favour of predictions from BSM models
2. No statistically significant deviations in SM precision measurements

Setting

1. No indications in favour of predictions from BSM models
2. No statistically significant deviations in SM precision measurements

Sounds good.

Setting

The rise of "model-independent" methods

1. SM precision measurements
2. Deep Learning
   - to analyse low-level data with fewer high-level physics assumptions

Setting

The rise of "model-independent" methods

1. SM precision measurements
2. Deep Learning
   - to analyse low-level data with fewer high-level physics assumptions

Part of a open-ended strategy for discovery

Roles and Aim

Roles DL in HEP

1. efficiency
2. performance
3. model independence

Roles and Aim

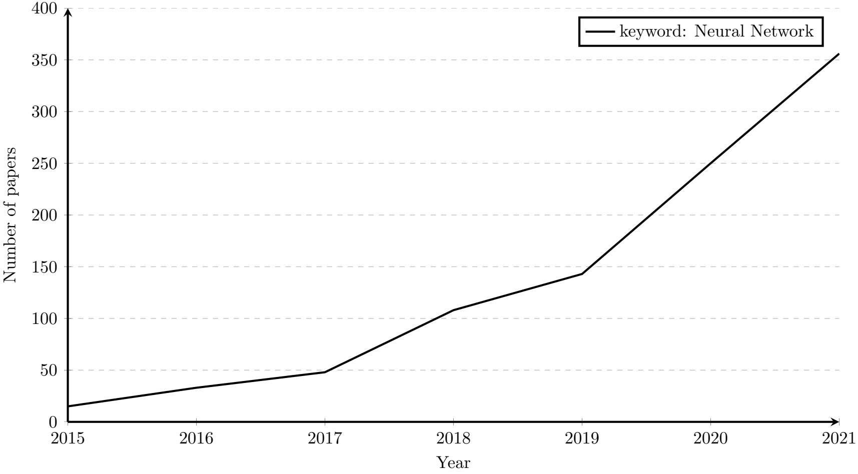
Roles DL in HEP

1. efficiency
2. performance
3. model independence

Hierarchical data representation

- well-suited to the absurdly high-dimensional data of HEP

Rise of the Machine (Learning Approach)

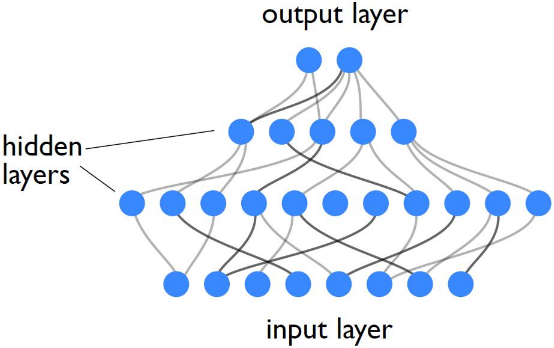


Overview

1 Deep Learning in HEP

2 Epistemic Issues

Nodes



$$y_{out} = F_w(y_{in}) \tag{1}$$

Changing Game

"In the relatively few years that modern machine learning has existed, it has already made traditional collider physics obsolete. In the past, physicists, including me, would devote their efforts to understanding signatures of particular particles or processes from first-principles: why should a stream of pions coming from a W boson decay look different than a stream coming from an energetic gluon? Now we simply simulate the events, and let a neural network learn to tell the two samples apart."

(Schwartz, 2021)

Curse of Dimensionality

The difficulty that deep learning can deal with is not the sheer amount of data, but the *high-dimensionality* of LHC data

Task in HEP

- estimate probability function $p(x|\theta)$ of observing some data x, given the model's parameters θ
- estimating is possible for low-dimensional data $(d < 5)$
- number of samples to estimate the function grows with the power of the dimensionality of data N^d

(Guest et al., 2018)

Must reduce the dimensionality of the data,

Curse of Dimensionality

The difficulty that deep learning can deal with is not the sheer amount of data, but the *high-dimensionality* of LHC data

Task in HEP

- estimate probability function $p(x|\theta)$ of observing some data x, given the model's parameters θ
- estimating is possible for low-dimensional data $(d < 5)$
- number of samples to estimate the function grows with the power of the dimensionality of data N^d

(Guest et al., 2018)

Must reduce the dimensionality of the data, $d = \mathcal{O}10^8$*

Reducing Dimensionality

Strategy

- Construct higher-level, lower-dimensional objects and analyse those

Steps

1. reconstruction algorithms are used to process raw data into objects (clusters and tracks)
2. use this to estimate the energy and momentum of particles
3. identify particles
4. build *event-level* summaries
5. perform event selection for further analysis
    - event rate $\sim$ 40kHz

Reducing Dimensionality

The reconstruction and selection are traditionally based on *physicist-identified* features of the data (having a given particle(of a certain energy), shape of the shower in the calorimeter, etc.)

Reducing Dimensionality

The reconstruction and selection are traditionally based on *physicist-identified* features of the data (having a given particle(of a certain energy), shape of the shower in the calorimeter, etc.)

- leading-order processes
- works pretty well, but not *optimal*
- DL helps optimize this process

Hierarchical Data Representation

	Image Recognition	**HEP**
High level	Shapes $\rightarrow$ Objects	Particles $\rightarrow$ Events
$\uparrow$		
Middle level	Edges $\rightarrow$ Shapes	Tracks $\rightarrow$ Particles
$\uparrow$		
Low level	Pixels $\rightarrow$ Edges	Hits $\rightarrow$ Tracks

Uses of DL in HEP

All Along Data Pipeline

1. Simulation
2. Improving triggering
3. Low-level: hit and track reconstruction
4. Mid-level: object identification
5. High-level: event classification

Uses of DL in HEP

All Along Data Pipeline

1. Simulation
2. Improving triggering
3. Low-level: hit and track reconstruction
4. Mid-level: object identification
5. High-level: event classification

The Need for Speed

At HL-LHC, there will be 7x luminosity (20x run-2 data set), not enough more computing resources

- ML has been used for decades (multivariate analysis, boosted decision trees)
- DL solutions are needed*

Object Identification

The idea

- have DL run through low-level data and identify an object (tell two objects apart) *without* telling it how to do so*

Object Identification

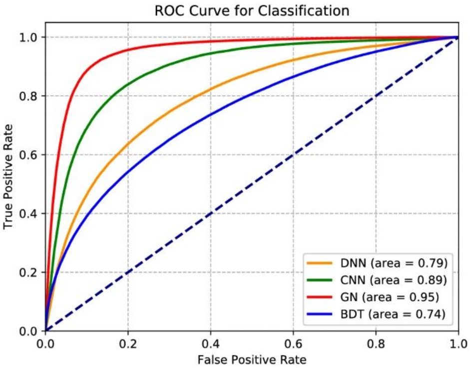


Figure: Three different deep learning classifiers, labelled *DNN*, *CNN*, and *(GN) GoogleNet* compared with the traditional BDT in the task of distinguishing γ vs. π^0. These deep networks featured between three and fifteen million trainable parameters (Belayneh et al., 2020).

Object Identification

"Our analysis shows that recent advances in deep learning techniques may lift these limitations by automatically discovering powerful non-linear feature combinations and providing better discrimination power than current classifiers—even when aided by manually-constructed features"

(Baldi, 2014, p. 9)

Object Identification

If you want to tag a top quark ($t\bar{t}$ pair production)

- don't tell the model that the top is heavy, that it decays to three other jets, that it has a slightly displaced vertex, etc.
- feed it tons of simulated data and it will have a higher success rate*

Discovery

Deep Learning as a Model Independent Strategy

- run through the entire data with minimal assumptions
- make decisions based on many variables at the same time
- identify background without knowing the SM
- identify anomalous events

Discovery

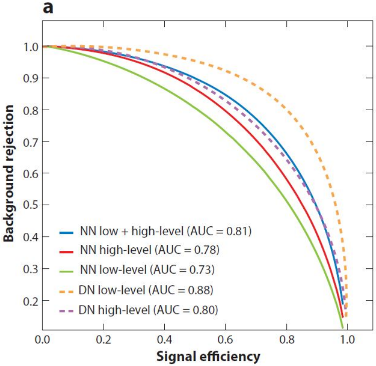


Figure: In their own words, "A comparison between the performance of deep networks (DNs) in signal–background classification and that of shallow networks (NNs) with a variety of low- and high-level features demonstrates that DNs with only low-level features outperform all other approaches" (Guest et al., 2018, p. 169)

Epistemic Difficulties in HEP

1. No theory of DL
2. (Over)Reliance on simulated data
3. Mismodelling
4. Interpretability and understanding

1. No Theory

The mathematical theory of statistical learning lags far behind the success of DL models

- little to offer in terms of design and refinement of algorithms and new techniques
- matter of heuristics and trial and error
- no theoretical guarantee of reliability or optimality

(Bahri et al., 2020; Belkin, 2021)

Simulations and simulated data is used *everywhere* in HEP

(Morrison, 2015)

- detector response cannot be analytically computed

2. Reliance on Simulation

Simulations and simulated data is used *everywhere* in HEP

(Morrison, 2015)

- detector response cannot be analytically computed

Quantum data

- The quantum nature of the real event means that simulated data is different
  - there are no ground truths
  - for a given event, you cannot say ‘Higgs’ or ‘no Higgs’ decisively
  - all one can do is collect statistics and try to reject the background only hypothesis

(Schwartz, 2021)

3. Mismodelling

Trusted simulations work in low-dimensional spaces with high-level physics features

- correlations in a high-dimensional space may not be faithful

Just because a model has a better false-positive rate doesn't mean it will perform better in particle identification

- e.g. if it works better on simulated data than real
- performance will be up to systematic uncertainties, signal-to-background ratio, etc.
    - things that are traditionally accounted for, but ignored in DL*

4. Interpretation and Understanding

"Generically, we should anticipate a trade-off between performance and interpretability"

(Guest et al., 2018, p. 175)

It is not a like an analytic function, where the structure of the solution can be analysed

- reverse engineering the classification strategy is impossible due to the high-dimensional nature of the data and the number of parameters of the models

The decisions may be faster, more accurate, more computationally tractable, but as long as we don't understand the decisions, they will be of limited benefit

Specificity

- Perhaps something learned in a network trained on top tagging could be useful in bottom tagging, but we won't know
- Sharing with other experiments may not be possible, since they run different hardware, often software*

Non-traditional

- not a cut-based analysis given in histograms and natural language description
- you get a multidimensional result that cannot even be graphically represented*

Issues but not Problems

1. No theory is kind of okay, as long as it works [1]
2. Sufficient data for reliability [2,3]
3. Post hoc interpretation—DNNs don't do everything [4]

Justification is not explanation

- justification comes from success
- its decisions can be easily justified, but not easily explained

The proof is in the pudding

- theoretical guarantees still rely on statistical arguments that are essentially about reliability of results

2. Simulated and Real Data

No need

- In principle, one can proceed in HEP entirely without simulations
  - if it can be simulated, it can be done with pen and paper
  - it would be slow

Use real data

- DL can also be run on real data (without ground truths), or mix of real and sim
- there is roughly the same amount of sim and real data at LHC (trillion events)
- simulations are rigorously validated*

Sufficient data

- in HEP there is *too much* data and computing resources
  - train and test until the cows come home*
  - statistical errors are small, *systematics* become the issue
    - DL can help with this too! $(error_{sys} \to \nu)$

Varied approaches

- data is public and there are open competitions
- new approaches and architectures are tried across the globe

Maybe its just not for that

- it is not already a model of new physics
- if an anomaly is discovered, the deep CNN (or whatever it is) will provide no understandable model in terms of intuitive physics

Maybe its just not for that

- it is not already a model of new physics
- if an anomaly is discovered, the deep CNN (or whatever it is) will provide no understandable model in terms of intuitive physics

But it will indicate where new physics is

- then traditional model-building will turn on
- similar to EFT approach
- similar to xAI

It can be tool to identify correlations, but it will not likely help in interpreting the data

Conclusion: Role for DL in HEP

- three important feature of DL: performance, efficiency, and model independence

Conclusion: Role for DL in HEP

- three important feature of DL: performance, efficiency, and model independence
- has many practical and epistemic difficulties to overcome

Conclusion: Role for DL in HEP

- three important feature of DL: performance, efficiency, and model independence
- has many practical and epistemic difficulties to overcome
- but these are *issues* not *problems*

Thank you

Martin L. King
m.king@lmu.de
www.philphys.com
MCMP Ludwigs-Maximilians-University Munich

This research was supported by the DFG.

References:

ATLAS Collaboration (2021). *JINST*, 16.

Bahri, Y., Kadmon, J., Pennington, J., Schoenholz, S. S., Sohl-Dickstein, J., and Ganguli, S. (2020). Statistical mechanics of deep learning. *Annual Review of Condensed Matter Physics*, 11(1):501–528.

Baldi, P., S. P. . W. D. (2014). Searching for exotic particles in high-energy physics with deep learning. *Nature Commun*, 5:4308.

Belayneh, D., Carminati, F., Farbin, A., Hooberman, B., Khattak, G., Liu, M., Liu, J., Olivito, D., Pacela, V. B., Pierini, M., Schwing, A., Spiropulu, M., Vallecorsa, S., Vlimant, J.-R., Wei, W., and Zhang, M. (2020). Calorimetry with deep learning: particle simulation and reconstruction for collider physics. *The European Physical Journal C*, 80(7).

Belkin, M. (2021). Fit without fear: remarkable mathematical phenomena of deep learning through the prism of interpolation.

CMS (2018). Usage of tracking in the cms level-1 trigger for the high luminosity lhc upgrade.

Guest, D., Cranmer, K., and Whiteson, D. (2018). Deep learning and its application to lhc physics. *Annual Review of Nuclear and Particle Science*, 68(1):161–181.

Morrison, M. (2015). *Reconstructing Reality: Models, Mathematics, and Simulations*. Oup Usa.

Schwartz, M. D. (2021). Modern machine learning and particle physics. *Harvard Data Science Review*, 3(2).

Triggers

CMS/ATLAS have $\sim$100 million channels sampled every $\sim$25ns

- 60Tb/s raw data

40 million events/s must be searched in **real time**

1. L1 hardware trigger $\rightarrow$ 100k/s
   - calorimeter + muon spectrometer energy threshold (+topo)
   - decision in 200 nano secs
2. L2 software (high-level) trigger $\rightarrow$ 1000/s
   - jets, missing p_T
3. ATLAS has L3 trigger $\rightarrow$ 200 events

Triggers

DL offers the possibility of triggering on events that cannot currently be stored

- acceptance 1 in 40,000 events
  - can write 15Pb/y for offline analysis
- try to keep low-energy DM signatures, b and c quark physics, etc.
- event selection based on cuts may miss a lot of events of interest

(ATLAS Collaboration, 2021; CMS, 2018)