Benjamin Nachman

with Sergei Gleyzer, Gregor Kasieczka, and Gordon Watts

Fifth workshop of the LHC LLP Community, May 2019
Deep learning* has great potential to improve all areas of HEP.

Now is the time to ask what is the potential for LLPs!

*This just means “machine learning”, but emphasizes that I’m talking about modern methods - today’s NN’s/BDTs/etc. are not those of LEP and the Tevatron…
Motivation

Deep learning* has great potential to improve all areas of HEP. Now is the time to ask what is the potential for LLPs!

I don’t have to tell you that LLPs have unique challenges. What I hope to tell you is about how deep learning may be able to help.

*This just means “machine learning”, but emphasizes that I’m talking about modern methods - today’s NN’s/BDTs/etc. are not those of LEP and the Tevatron…

Image from J. Antonelli - please let me know if this is not the original source!
The toolkit is large!

...and it is important to use the right tool for the job! (may not be a NN)
How can we use deep learning?

Classification

- provide examples for training
- arbitrarily many categories

Regression

- map noise to structure

Generation

+ Extra credit: weak/unsupervised learning and anomaly detection
Classification: signal vs. background

Low-level inputs + multiple detector systems.

Q: How to validate when simulations are not fully reliable, as is often true for LLP (more on this later)?

Image credit: ATLAS + BPN + Josh Lin (see also J. Lin et al., JHEP 10 (2018) 101)
Can we automate the energy estimation of non-standard objects?

Augment standard algorithms with non-standard information?

**Warning**: need to worry about prior dependence. This is a ~solved problem with "generalized numerical inversion"
Can we design a fast sim that captures full-sim level information about strange showers using generative NN’s?
Extra Credit 1: No labels (“weak supervision”)

CWoLa
Classification Without Labels

One solution: Train directly on data using mixed samples

N.B. one of these could be sim and one could be data

for more, see L. Dery, BPN, F. Rubbo, A. Schwartzman, JHEP 05 (2017) 145,
Extra Credit 2: No simulation (s or b)

Ideas → Reality

Image from J. Antonelli - please let me know if this is not the original source!
An actual example (for classification)

HCP can be identified with high dE/dx in pixel detectors. Often, people use truncated mean. What is the best way to combine charge info?

ML is good for asking questions like this, even if the ultimate answer is simpler than “use a NN”
Trigger strategies

Reconstruction Techniques
- displaced jet calibration
- odd jet tagging
- non-standard tracking

Simulation
- calorimeter/tracker information
- delayed signals

Analysis Strategy/Interpretation
- supervised / semi-supervised
Plan for the workshop

...informal discussion of what is already being done

...discussion of census and start discussion of techniques

...continue brainstorming techniques → LLP ideas; brief tutorial possible time/interest permitting
There are many non-ML studies out there that maybe can be reused for ML studies.

Example: quirks


...we may also be able to put together public collaboration datasets - please keep an open mind!

*I did not ask Simon, Tim, Michele, or Jack if we could use their simulation - this is only an example of something that may be useful and if there is interest, we should ask them!
Conclusions and Outlook

There has already been great work, but a lot of ML potential untapped - an exciting future is ahead!

Image from J. Antonelli - please let me know if this is not the original source!
Questions?