Deep learning and neutrino physics

G. N. Perdue[1], T. Golan[4], A. Himmel[1], E. Niner[1], F. Psihas[2], and A. Radovic[3]

[1]Fermi National Accelerator Laboratory, Batavia, Illinois 60510, USA
[2]Indiana University, Department of Physics, Bloomington, Indiana 47405, U.S.A.
[3]Department of Physics, College of William & Mary, Williamsburg, Virginia 23187, USA
[4]University of Wroclaw, Department of Physics and Astronomy, 50-204 Wroclaw, Poland

April 4, 2017

1 The data analysis challenge of neutrino physics

The Particle Physics Project Prioritization Panel ("P5") report [1] highlights the physics of neutrino mass as one it it's five Science Drivers for organizing the activities of high energy physics (HEP). Because neutrino masses are so anomalously small, they may provide a window to the highest energy scales in nature, allowing access to regimes far beyond what we may currently test directly. Cornerstones of this program are the efforts to establish whether neutrinos and antineutrinos oscillate differently (so-called "CP violation") and to understand the nature and ordering of the neutrino masses. Novel tools are needed to fully exploit the enormous amount of data being generated now and to be produced in the future to address these programs. Furthermore, neutrino experiments, especially in the large far detectors characteristic of long-baseline experiments, are in the interesting position of having enormous data volumes, but very few events in the flagship oscillation analysis. We face a serious problem not only in developing algorithms capable of sifting through huge volumes of data, but we also must develop algorithms that are maximally efficient to fully utilize the investments required to produce neutrino interactions in far detectors.

The past several years have witnessed a revolution in computing and machine learning [2, 3]. Using new hardware advances, particularly more capable Graphics Processing Units (GPUs), and the algorithms associated with "deep learning" (the use of neural networks with many hidden layers), computers have surpassed humans in certain pattern recognition exercises, particularly in computer vision and image recognition problems, but tremendous leaps are being made in many endeavors. These techniques have clear application to HEP event reconstruction as modern neutrino detectors are effectively imaging devices. Early results [4–6] indicate that deep learning will significantly improve the physics reach at running neutrino experiments. There is even the intriguing possibility to use deep learning and other forms of machine intelligence to discover entirely new phenomena in our data, although this is still a speculative exercise.

While this paper will largely discuss deep learning in HEP from perspective of neutrino physics, it is doubtlessly true that many of the items discussed apply just as well at colliders. Finding common synergies between needs at neutrino, collider, and other HEP experiments will be an important and valuable task in the very near future.

2 Deep learning brought to bear

But many questions must be answered first to understand how to map deep learning technologies into the particle physics domain. Almost all training for deep neural nets in HEP is done with Monté Carlo - how do we train deep networks when the model (our simulation stack) contains uncertainties? How do we understand what the network is learning - is it able to find relationships hidden under the complex interplay between particles in the detector or is it merely memorizing images? The high quality and large volume of simulation available in high energy physics make this an excellent arena to study deep neural network behavior and optimization while simultaneously expanding the reach of particle physics experiments.

While many problems in the HEP domain map onto known problems in the deep learning community, some do not and we are often concerned about different metrics. As a community we should push deep learning research to address HEP's specific concerns. For example, we may be willing to accept lower efficiency in favor of a more clearly understood *uncertainty* on that efficiency. Additionally, outside of HEP the vast majority of deep learning applications, including the image classification and object detection algorithms being adapted for HEP purposes, rely on labeled sets of data for training. In contrast, the low statistics searches in HEP necessitate training based on finely tuned simulations the community has spent decades developing. Early attempts to collaborate betweent communities have already created papers of interest to both HEP and DL communities[7], and hopefully mark the start of a wider trend.

2.1 Relationships to industry and academic computer science

Industry is having an enormous impact on deep learning. Some argue that large information technology companies like Google are rebuilding themselves around it, making enormous multi-billion dollar wagers that we will continue to see improving results [8]. What's more, industry and academic computer science have begun to share research space as for-profit companies look to hire expertise as quickly as possible and must also allow their staff to publish findings in academic journals in order to retain the best investigators. HEP physicists can take advantage of significant investments being made by industry and academic computer science to develop the most important software frameworks. Many leading frameworks have important industrial sponsors. While Google has probably the most broadly used deep learning software framework, TensorFlow [9, 10], companies like Amazon and Microsoft are hard at work on their own libraries. Meanwhile frameworks like Caffe [11] and Theano [12] thrive in academic settings. All three of TensorFlow, Caffe, and Theano are heavily used in HEP.

These frameworks will continue to improve without input from HEP, but it will be important for our community to engage with these groups. The projects are open source and it will be important to build relationships and trust with these groups so we can contribute code we need or voice support for the inclusion of models that may be somewhat esoteric for industry in general. For example, domain-adversarial neural nets (DANNs) [13] potentially provide an algorithmically simple way of reducing model bias from inadequate physics modeling in the simulation we use for training. We want to engage with the academic computer science community to stimulate research like this and it will also be important to engage with deep learning framework authors to ensure that otherwise esoteric features like DANNs remain supported and efficient.

2.2 Data formatting

Data formatting as a common, critical need across all of HEP. Deep learning frameworks are designed to consume data in very specific formats, for example HDF5 [14] and LMDB [15], with a focus almost entirely on quickly transferring data to a GPU. This is because one of the largest bottlenecks when using GPUs is getting data onto the GPU fast enough to keep up with the rate of computation. Currently, HEP experiments primarily use ROOT [16] for persistent storage and ROOT I/O is not optimized for high transfer rates to a GPU or some other co-processor card. Experiments are forced to develop ad-hoc ways of preparing samples and the community as a whole could benefit from common, high-quality tools that allowed us to fork versions of our data into formats that deep learning frameworks could consume in a fashion that was effort and computation efficient. We could also benefit from tools and developments that made it simpler for deep learning frameworks to more transparently consume data from HEP event stores directly, even at the cost of efficiency in circumstances where the convenience strongly outweighed performance concerns (for example, in the last stages of user analysis).

2.3 Democratizing access and exploiting the most powerful computation engines

Addressing the full menu of questions is only possible with leadership computing, but this itself is a process full of questions. What is the optimal way to conduct deep learning analysis on high performance computing (HPC) facilities? How do we make the process of defining a problem and submitting analysis jobs to HPC facilities widely accessible to the experiments? Finally, how do we give physicists the tools to scale up analyses from small test clusters to leadership class facilities (LCFs)? Many large machines (for example, Titan at Oak Ridge National Laboratory) prefer users that can utilize the entire machine at once. But there is a dearth of dedicated resources in HEP institutions between very small clusters and machines the size of Titan (over 18,000 nodes), and scaling to machines of that size is a non-trivial exercise. Cloud resources (e.g. Amazon AWS and Google Cloud) exist and might fill this gap, but historically computing resources like these have not fit well in HEP funding models or workflows.

2.4 Investing in knowledge transfer and training

One of the topics of common interest for all HEP experiments is that of knowledge transfer. The use and development of Machine Learning in HEP is made possible by scientists whose skill set combines physics and machine learning tools. However, as with all computational tools used in the field, some training is necessary for every generation of new scientist to fully exploit the benefits of employing said tools. In many cases it is up to individual experiments to develop and organize training, thus, utilizing manpower and potentially duplicating content and resources which already exist elsewhere in the field. A concerted effort in training young scientists in the use and development of machine learning tools would provide them with the skills needed to utilize them while relieving experiments from the burden of spending time and resources in knowledge transfer of common tools.

3 Conclusions

Understanding deep learning as it is applied to particle physics will help to fully realize the investments being made to pursue the most fundamental questions in nature. Strategic investments in the technology, understanding what makes HEP similar to and different from industrial and other academic avenues of research, and community-wide efforts to leverage our intellectual and financial resources efficiently are all crucial ingredients in this effort. Even under conservative assumptions about the trajectory for future improvements in deep learning, this will be an important technology for analyzing data at HEP experiments, and perhaps in many other tasks as well, in the years to come. But to fully realize those benefits it will important for us to engage with the creation of deep learning software, to build new relationships with the academic computer science community, develop common tools for data formatting and accessing deep learning codes, develop a pathway onto HPC facilities that all physicists can use, and find efficient ways to build knowledge and expertise within the field.

Acknowledgements

We would like to thank Peter Elmer, Oliver Gutsche, Jim Kowalkowski, Mark Neubauer, Panagiotis Spentzouris, and all the attendees of the HEP Community White Paper Meeting [1] held at Fermilab on March 23, 2017 for discussion that informed and inspired much of this paper.

References

[1] Building for Discovery: Strategic Plan for U.S. Particle Physics in the Global Context, http://www.usparticlephysics.org/p5/, 2014.

[2] A. Krizhevsky, I. Sutskever, and G. E. Hinton, Imagenet classification with deep convolutional neural networks, in *Advances in Neural Information Processing Systems*, 2012.

[3] A. S. Razavian, H. Azizpour, J. Sullivan, and S. Carlsson, CoRR **abs/1403.6382** (2014).

[4] NOvA, P. Adamson *et al.*, (2017), 1703.03328.

[5] A. Aurisano *et al.*, JINST **11**, P09001 (2016), 1604.01444.

[6] MicroBooNE Collaboration, Journal of Instrumentation **12**, P03011 (2017).

[7] G. Louppe, M. Kagan, and K. Cranmer, (2016), 1611.01046.

[8] S. Levy, How google is remaking itself as a “machine learning first” company, 2016.

[9] M. Abadi *et al.*, Tensorflow: A system for large-scale machine learning, in *Proceedings of the 12th USENIX Conference on Operating Systems Design and Implementation*, OSDI'16, pp. 265–283, Berkeley, CA, USA, 2016, USENIX Association.

[10] M. Abadi *et al.*, CoRR **abs/1603.04467** (2016).

[11] Y. Jia *et al.*, arXiv preprint arXiv:1408.5093 (2014).

[12] Theano Development Team, arXiv e-prints **abs/1605.02688** (2016).

[13] Y. Ganin *et al.*, Journal of Machine Learning Research **17**, 1 (2016).

[14] The HDF Group, Hierarchical data format version 5, 2000-2010.

[15] H. Chu, LDAP11 (2011).

[16] R. Brun and F. Rademakers, Nucl.Instrum.Meth. **A389**, 81 (1997).

[1] https://indico.fnal.gov/conferenceDisplay.py?confId=14032