PROCEEDINGS OF SCIENCE

Exploring Raw HEP Data using Deep Neural Networks

Evan Racah, Seyoon Ko, Peter Sadowski, Wahid Bhimji*, **Craig Tull, Sang-Yun Oh, Prabhat**
Lawrence Berkeley Lab, Berkeley, CA
Email: wbhimji@lbl.gov

High Energy Physics has made use of artificial neural networks for some time. Recently, however, there has been considerable development outside the HEP community, particularly in deep neural networks for the purposes of image recognition. We describe the deep-learning infrastructure at NERSC, and analyses built on top of this. These are capable of revealing meaningful physical content by transforming the raw data from particle physics experiments into learned high-level representations using deep convolutional neural networks (CNNs), including in unsupervised modes where no input physics knowledge or training data is used. Here we describe in detail a project for the Daya Bay Neutrino Experiment showing both unsupervised learning and how supervised convolutional deep neural networks can provide an effective classification filter with significantly better accuracy than other machine learning methods. These approaches have significant applications for use in other experiments triggers, data quality monitoring or physics analyses.

38th International Conference on High Energy Physics
3-10 August 2016
Chicago, USA

*Speaker.

PoS(ICHEP2016)889

1. Deep Learning At NERSC

Cori is the newest supercomputer at the National Energy Research Scientific Computing Center (NERSC). Cori Phase 1 consists of 1,630 32-core Intel Haswell compute nodes. Cori Phase 2 will add around 9300 Intel Knights Landing (KNL) nodes. The system also provides a 29 PB Lustre Filesystem and 1.5 PB NVRAM Burst Buffer. NERSC provides optimised versions of several deep-learning frameworks. These include, Theano: for flexibility in method development; Keras/Lasagne: Theano-based and high-level for ease-of-use; Caffe: including IntelCaffe with performance highly optimised for KNL; Neon (used here): optimised for high performance and parallel implementations; and TensorFlow: for ease-of-use and flexibility in addition to a large, growing community. These frameworks are available within a deeplearning kernel on the NERSC Jupyter service (ipython.nersc.gov) which allows for interactivity and scaling up to runs on Cori.

As well as the Daya Bay project featured here, other deep learning HEP projects are ongoing at NERSC e.g., Ice Cube: improving astrophysical neutrino detection; ATLAS (HL-LHC) Tracking: using recurrent NNs (such as LSTMs); ATLAS Calorimeter: CNNs on calo-cells for multi-jet physics analyses; Probabilistic Programming: coupling inference to simulations for new-physics detection; Cosmology images: CNNs to find clusters of galaxies and filaments in large simulations.

2. Use Case: Daya Bay

A Daya Bay Antineutrino Detector (AD) consists of 192 photomultiplier tubes (PMTs) arranged in a cylinder 8 PMTs high and with a 24 PMT circumference [1]. We use the value of the charge deposit of each of the PMTs in the cylinder unwrapped into a 2D (8 "ring" x 24 "column") array of floats that is essentially a 2-D image. For the supervised part of this analysis, and for visualizing the unsupervised results, we employ labels determined by the collaboration's physics analyses (see [1]). These label four types of events: "muon", "flasher", "Inverse Beta Decay (IBD) prompt", and "IBD delay". A label of "other" is applied to all other events.

2.1 Methods: Convolutional Neural Networks and Autoencoders

A convolutional neural network (CNN) captures our intuition about local structure and translational invariance in images [2]. Typical CNNs have several convolutional and pooling layers followed by fully connected layers. An autoencoder [3] is a neural network where the target output is the input. It consists of an encoder: layers that transform the input into a feature vector at the output of the middle layer (bottleneck or hidden layer), and a decoder: layers that attempt to reconstruct this bottleneck back to the input. When the bottleneck layer output has dimensionality smaller than that of the input (undercomplete), the network learns how to compress and reconstruct examples. A convolutional autoencoder typically consists of convolutional and max-pooling layers followed by fully-connected hidden layers (including a bottleneck) and then deconvolutional (and unpooling) layers [4]. We used a RELU [2] activation for the convolutional layers in the autoencoder. The architecture of our networks are given in Table 1. As a qualitative assessment, we use t-Distributed Stochastic Neighbor Embedding (t-SNE) [5], which maps n-dimensional data to 2 dimensions, ensuring points close together in the higher dimensional space are also near in 2-D.

Much more detail on the methods and results of this study can be found in [6].

PoS(ICHEP2016)889

type	filter size / number / stride
conv	3×3 / 71 / 1
pool	2×2 / 1 / 2
conv	2×2 / 88 / 1
pool	2×2 / 1 / 2
fc	1×5 / 26 / 1
fc	1×1 / 5 / 1

type	filter size / number / stride / pad
conv	5×5 / 16 / 1 / 2x2
pool	2×2 / 1 / 2 / 0
conv	3×3 / 16 / 1 / 1×0
pool	2×2 / 1 / 2 / 0
fc	2×5 / 10 / 1 / 0
deconv	2×4 / 16 / 2 / 0
deconv	2×5 / 16 / 2 / 0
deconv	2×4 / 1 / 2 / 0

Table 1: Architecture (layers) of the supervised CNN (left) and convolutional autoencoder (right)

Method	IBD prompt	IBD delay	Muon	Flasher	Other
k-NN	0.950	0.990	0.998	0.891	0.896
SVM	0.966	0.992	0.998	0.947	0.938
CNN	**0.977**	**0.995**	**0.999**	**0.974**	**0.962**

Table 2: Classifier accuracy for different methods and event types

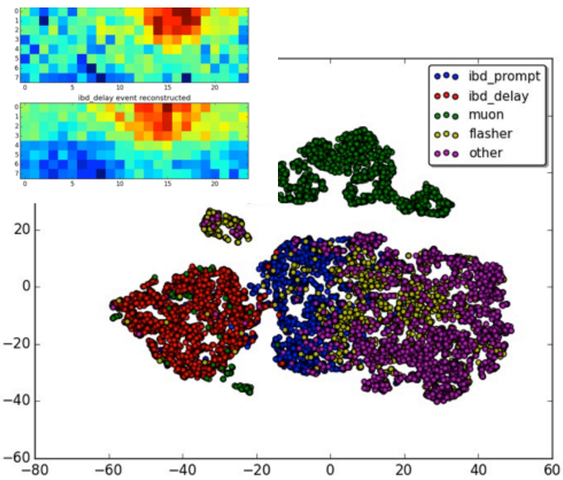

Figure 1: t-SNE representation of features learned by the convolutional autoencoder. Insert shows original and reconstructed image of one IBD delay event.

2.2 Results and Interpretation

Supervised Learning with CNN: The classification accuracies of k-nearest neighbor, support vector machine, and the CNN architecture on the test set are summarized in Table 2. These results suggest that there are patterns in the Daya Bay data that could be uncovered by the CNN. We were able to achieve high accuracy on classification using only the spatial pattern. Our results also show that deep neural networks were better than the other machine-learning techniques at classifying the images (particularly on classes "IBD prompt" and "flasher") and thus finding patterns in the data.

Unsupervised learning with Convolutional Autoencoder: The t-SNE visualization of the 10 features learned by the network is shown in Figure 1 and obtains well defined clusters without using physics-based selections. There is a clearly separated cluster that is identified with the labeled muons, and a separation between "IBD delay" and other events. We even achieve some separation between "IBD prompt" and "other" backgrounds which is mainly achieved in the physics analysis by incorporating additional timing information. In the reconstructed images, we can see the autoencoder was able to filter out input noise and reconstruct the important shape of different event types: for example for "IBD delay" in the insert in Figure 1.

2.3 Conclusions

We apply unsupervised convolutional neural nets to raw data from the Daya Bay experiment and have shown that the network can successfully learn patterns of physics relevance. Such unsupervised techniques could be used for a variety of particle physics experiments to aid in trigger decisions, in evaluating data quality, or to discover new instrument anomalies without having to engineer features. We have also demonstrated that convolutional neural networks outperform other supervised machine learning approaches running directly on raw particle physics data. These can be used in fast triggers or in final analyses. We are now focussing on more challenging backgrounds that dominate the current systematic uncertainties from the experiment as well as state-of-the-art methods incorporating denoising or variational autoencoders and semi-supervised approaches.

Acknowledgments

The authors gratefully acknowledge the Daya Bay Collaboration for access to experimental data and discussions. We used resources of the National Energy Research Scientific Computing Center, a DOE Office of Science User Facility supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231. S. Ko was supported by National Research Foundation of Korea (NRF) grants (Nos. 2013R1A1A1057949, 2014R1A4A1007895).

References

[1] F. P. An *et al.* [Daya Bay Collaboration]. Improved measurement of electron antineutrino disappearance at Daya Bay. *Chinese Physics C*, 37(1):011001, 2013.

[2] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.

[3] Yoshua Bengio, Pascal Lamblin, Dan Popovici, Hugo Larochelle, et al. Greedy layer-wise training of deep networks. *Advances in neural information processing systems*, 19:153, 2007.

[4] Vincent Dumoulin and Francesco Visin. A guide to convolution arithmetic for deep learning. *arXiv preprint arXiv:1603.07285*, 2016.

[5] Laurens Van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of Machine Learning Research*, 9(2579-2605):85, 2008.

[6] Evan Racah, Seyoon Ko, Peter Sadowski, Wahid Bhimji, Craig Tull, Sang-Yun Oh, Pierre Baldi, and Prabhat. Revealing Fundamental Physics from the Daya Bay Neutrino Experiment using Deep Neural Networks. *arXiv preprint arXiv:1601.07621*, 2016.