Deep learning with raw data from Daya Bay

To cite this article: S Kohn et al 2017 J. Phys.: Conf. Ser. 898 072050

View the article online for updates and enhancements.

Related content

- Latest results from Daya Bay
  Vit Vorobel and Daya Bay Collaboration

- Calibration and Reconstruction of the Daya Bay Antineutrino Detector
  Bei-Zhen Hu and Daya Bay Collaboration

- Latest progress from the Daya Bay reactor neutrino experiment
  Zhe Wang and Daya Bay Collaboration
Deep learning with raw data from Daya Bay

S Kohn\textsuperscript{1,2†}, E Racah\textsuperscript{2}, C Tull\textsuperscript{2†}, D Dwyer\textsuperscript{2†}, Prabhat\textsuperscript{2}, and W Bhimji\textsuperscript{2}

\textsuperscript{1} Department of Physics, University of California, Berkeley, CA 94720
\textsuperscript{2} Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, CA 94720
\textsuperscript{†} On behalf of the Daya Bay Collaboration

Abstract. The Daya Bay experiment uses reactor antineutrino disappearance to measure the $\theta_{13}$ neutrino oscillation parameter. In this proceeding, the convolutional autoencoder machine learning technique is tested against a well-understood uncorrelated accidental background. The eventual goal for this technique is to reduce the background with the largest contribution to the rate uncertainty in the antineutrino data set, $\beta$-n decay of $^9$Li produced by cosmic-ray muons.

The Daya Bay Reactor Neutrino Experiment is a precision neutrino oscillation experiment that measures the disappearance of $\bar{\nu}_e$ at the far site relative to the measurements obtained at the near sites [1]. The value of the $\theta_{13}$ neutrino mixing angle can be extracted from the difference in rates and energy spectra between the different detector sites.

The IBD reaction, $\bar{\nu}_e + p \rightarrow n + e^+$, is used to identify when antineutrinos interact in the antineutrino detectors (ADs), which are filled with Gd-doped liquid scintillator. The positron loses energy through scintillation and eventually annihilates within a nanosecond. The combination of these processes causes a prompt burst of scintillation light. The neutron thermalizes and captures on a Gd nucleus after approximately 25 μs, releasing a delayed burst of scintillation light. The characteristic prompt-delayed double coincidence of the IBD reaction is the primary discriminator used for rejecting background.

Although the double coincidence is a useful discriminator, there are other physical processes in the ADs which also cause a double coincidence, including accidental coincidences and the $\beta$-n decay of $^9$Li. Accidental coincidences are a result of the natural decay of radiocontaminants in and near the ADs, which causes a steady background of single flashes. Occasionally, two uncorrelated flashes accidentally occur within the coincidence time window and are accepted as an IBD event.

This background rate is straightforward to predict since the singles rate is well-measured in each AD and the process is Poissonian. Approximately 1%-2% of the IBD sample is accidental background, and that percentage is known to approximately 1%, for a total contribution to the rate uncertainty of $\lesssim 0.01\%$. As a consequence, accidentals do not contribute a significant uncertainty to the $\theta_{13}$ result and are an ideal trial data set for the convolutional autoencoder technique.

A further advantage of accidentals is that since they are formed from two uncorrelated events, it is simple to create an “artificial” set of accidental background out of real data, even if the true physical origin of each event is unknown. Modulo the energy selection, all that is necessary is for random singles events—displaced significantly in time—to be paired up and decreed to be accidentals. Such an artificial data set is by construction 100% pure and statistically similar to the true accidentals data set.
On the other hand, the more difficult background to identify is $^9$Li decay, which results in an electron and neutron being released. Since the electron signal in liquid scintillator is similar to a positron signal—and the electron energy in this decay is within the expected range of positron energies for IBD—this signal is almost indistinguishable from a true IBD signal.

The rate of $^9$Li background can be estimated to moderate precision since it is known that $^9$Li is created when cosmic-ray muons traverse the AD, and decays with a lifetime of $257.2 \text{ ms}$. By examining the change in the IBD rate as a function of time since the last muon event, the number of $^9$Li events contaminating the data set can be estimated. The final rate is $<1\%$ of the IBD rate, although the uncertainty on this amount is large enough that the contribution to the final IBD rate uncertainty is $0.1\%$, approximately 10 times larger than the uncertainty due to any other source of background.

Characterization of these backgrounds is a difficult process, and new techniques from the field of machine learning may be able to improve this characterization. Standard tools such as boosted decision trees and supervised neural networks are not well-suited to all problems. If the number of event classes (types) is unknown, or if the simulated events are known (or suspected) to have a bias in certain features such as timing or electronics noise, creation of a satisfactory training set would be impractical or impossible. In these situations, unsupervised learning can provide insights into important patterns in the data that may have discriminating power between different event classes.

The convolutional autoencoder neural network is such an unsupervised technique which can process images (or other arrays of pixels) and form groups of similar images [2, 3]. Convolutional networks (whether in an autoencoder architecture or not) operate on data where spatial relationships are important, such as photographs or data from an array of physics detectors. The autoencoder architecture can be trained without a labeled set of data, i.e. it allows for unsupervised training. The goal of the training is to create a compressed encoding scheme which preserves “important” features of the input data, and whose encodings can be decoded to a close “reconstructed” approximation of the original input. The importance of a feature is determined by whether knowledge of that feature is useful in creating an accurate reconstruction of the original input image.

![Figure 1: Inside of a cylindrical Daya Bay antineutrino detector module (AD).](image)

The data used for this analysis is collected from photomultiplier tubes (PMTs) in the Daya Bay antineutrino detectors (ADs). As shown in Figure 1, each AD contains a cylindrical array of PMTs in 8 rings with 24 columns per ring, for a total of 192 PMTs per AD. Hence each input “image” for the neural network will be an image with $8 \times 24$ “pixels” containing data from each PMT in the AD. The cylindrical geometry is unrolled into a rectangle, so that rows of pixels correspond to rings of PMTs, and columns of pixels correspond to columns of PMTs. The total antineutrino data set is approximately $2 \times 10^6$ antineutrinos in 4 years.
For a preliminary study, the convolutional autoencoder neural network is asked to learn about the differences between IBD signal events and accidental background events so that the autoencoder’s performance can be evaluated. In particular, the convolutional autoencoder’s ability to determine that there are different types of events in the training set will be examined.

Figure 2: Example prompt (top) and delayed (bottom) flashes for an IBD event. The color scale corresponds to the charge on the PMT.

The training set was formed out of 20,000 events taken from a single AD during normal operation of the Daya Bay experiment. The training set was split evenly between events tagged as IBD and the artificial accidents described earlier. The value for each pixel was taken to be the charge on the corresponding PMT, scaled so that the smallest charge of the whole data set is $-1$ and the largest charge is $+1$ (in arbitrary units).

One way to convert the prompt and delayed event images into an input to the autoencoder is to concatenate them, as shown in Figure 2. However, the autoencoder would not be aware that there is a correlation between the top half of the pixels and the bottom half in this image: namely, each pixel in the top half has a pixel in the bottom half which corresponds to the same physical PMT, just at a later time. To encourage the network to pay attention to these correlations, the concept of image channels is employed. An image with multiple channels can be thought of as assigning a vector value rather than a scalar value to each pixel. These channels are often utilized in the machine learning community to process red, green, and blue information in digital images. For this study, two channels are used: prompt charge and delayed charge. Thus there is an explicit connection in the autoencoder between the prompt and delayed charges for a given PMT.

The architecture used for the neural network is detailed in Figure 3. It was chosen as an adaptation of a supervised convolutional network studied previously [4]. In the diagram, the bottleneck layer is where the encoding is extracted. The encoding size is 16 pixels, hence the information stored in the original input (384 pixels) must be summarized in 16 pixels.

If the convolutional autoencoder is trained well, then it should be able to reconstruct the most important parts of the images given just the encodings. An example reconstructed image is shown in Figure 4. It is clear that the location of the bright spot is preserved, indicating that it is considered an important feature by the autoencoder.

By examining the 16-pixel encodings, it should be possible to distinguish between different classes of 384-pixel images. Visualization of these encodings is a useful way to quickly determine if there are any patterns. The technique used for visualization is known as t-distributed
Figure 3: Architecture of the convolutional autoencoder used in this study.

Figure 4: Example input (left) and reconstructed (right) images for a prompt light flash.

Stochastic Neighbor Embedding, or t-SNE [5]. This technique converts a set of points in a high-dimensional space (such as 16 dimensions) into a set of points in a low-dimensional space (such as 2) in such a way that the Euclidean distances between nearby points are preserved as well as possible. This technique allows for the easy identification of clusters of points.

The t-SNE visualization for a random subset of the training set is shown in Figure 5. Each point in the plot represents one event pair, and the position of each point represents the t-SNE representation of the 16-pixel encoding, also known as “semantic space.” Each point is colored according to whether the event it represents came from the IBD data set or the artificial accidental data set.

From the figure, it is clear that the convolutional autoencoder assigned different encodings to images representing IBD events compared to images representing accidentals. This is evident in the split between red points and blue points. Since the true identity of each event was not provided to the convolutional autoencoder at any point, it can be concluded that the autoencoder identified features which help determine whether an event is an IBD or an accidental.

Clearly, the discriminating power of the autoencoder is not exemplary, given the contamination of blue points in the red region and vice versa. However, the autoencoder learned on its own that some events had a different set of features compared to other events.

Further study is needed to understand which specific features were identified by the
Figure 5: t-SNE embedding of 5120 events. Each point represents one IBD or accidental event. The colors represent the data set the event was drawn from—IBDs or artificial accidentals. (This origin information was not provided to the autoencoder.) Points which are close to each other were assigned similar encodings by the autoencoder.

autoencoder. This analysis can be done both by examining images in different regions of the t-SNE plot and by examining the learned parameters of the autoencoder, for example by using guided backpropagation [6].

The information gained from understanding the discriminating features identified by the autoencoder could be used to construct a traditional analysis which can tag accidentals and IBDs and has well-understood systematics and biases (unlike an unsupervised neural network). Such an analysis is the subject of a further study. The same analysis pipeline could be used to improve identification of $^9$Li if convolutional autoencoders are successful at distinguishing $^9$Li events from IBDs.

This technique could be improved by using more sophisticated neural network architectures, including adversarial autoencoders [7], and by incorporating time information into the input data so the autoencoder can make decisions based on the timing of the light flashes.

This work was supported by the U.S. Department of Energy, Office of Science, Office of High Energy Physics.

References
[1] An F P et al. (Daya Bay Collaboration) 2015 Phys. Rev. Lett. 115(11) 111802
[2] LeCun Y, Bottou L, Bengio Y and Haffner P 1998 Proc. of the IEEE 86
[3] Fergus R, Zieher M D, Taylor G W and Krishnan D 2010 IEEE Conf. on Computer Vision and Pattern Recognition 2528–35
[4] Racah E, Ko S, Sadowski P, Bhimji W, Tull C, Oh S Y, Baldi P and Prabhat 2016 Revealing fundamental physics from the daya bay neutrino experiment using deep neural networks (Preprint stat.ML/1601.07621)
[5] van der Maaten L and Hinton G 2008 Journal of Machine Learning Research 9 2579–605
[6] Springenberg J T, Dosovitskiy A, Brox T and Riedmiller M A 2014 Striving for simplicity: The all convolutional net preprint (Preprint cs/1412.6806)
[7] Makhzani A, Shlens J, Jaitly N and Goodfellow I J 2015 Adversarial autoencoders (Preprint cs.LG/1511.05644)