A LEARNED EMBEDDING SPACE FOR EEG SIGNAL CLUSTERING

R. Thiyagarajan, C. Curro and S. Keene
The Cooper Union, New York, NY, USA
{krisht, curro, keene}@cooper.edu

***Abstract*—Despite the recent advances in medical data organization and structuring, electronic medical records (EMRs) can often contain unstructured raw data, temporally constrained measurements, multichannel signal data and image data all of which are often difficult to compare and contrast in large quantities due to their sizes and variation. We present a proof of concept system that can alleviate this by mapping EEG data to a relatively compressed n-dimensional space where the Euclidean distance between data points as similarity measure. We optimize a deep neural network mapping by using a triplet-based loss function. A system of this type could be used by medical professionals query and explore EEG data. To verify that this clustering method learns a meaningful representation of the data, we apply a KNN classifier to the output. We achieve a 58.6% classification accuracy operating on the neural network sourced embeddings on the six class TUH EEG Cohorts dataset provided by Temple University.**

I. Introduction

The health care industry commonly stores diverse instrumentation signals such as EEG, EKG, MEG, X-Ray, MRI, CAT in a variety of digital formats commonly referred as Electronic Medical Records (EMR). These records can also contain natural language notes from medical professionals. It is difficult to perform complex information retrieval on these records. Rich information retrieval may open up the ability to compare and contrast patient records en-masse leading to new understandings of diseases pathologies. For example, while the healthcare industry possesses a large amount of data on Alzheimer's Disease, a common chronic neurodegenerative disease, medical professionals are unable to find the underlying cause of this disease and why it worsens over time. If such data can be transformed to into an accessible and patient invariant format, medical professionals may be able to pinpoint the cause of the disease and discover better treatments. Towards this goal, Picone et al. [1] have demonstrated a system that can automatically discover, time-align and annotate EEG events. These annotations can help professionals perform cohort retrieval.

We present a novel system for clustering of raw EEG signal clustering. We seek to learn an embedding space using a deep neural network on one second EEG signals sampled at 250 Hz. We train the network such that the Euclidean distance between any two signals is a measurement of signal similarity. Signals that have the same class and similar features are expected to be clustered close together while those that are different are expected to have larger distances between each other in the embedding space. At inference time, we use the network to map EEG signals onto this embedding space for querying. New samples could be presented to this system at inference time to be mapped into the embedding space. After this mapping, clustering or other distance-based methods could be employed to find similar EEG records as part of a cohort retrieval system.

The signal event detection and classification work in [1] uses hidden Markov models. These models have been shown to work well with sequential data. Although this model achieved 91.4% sensitivity and 8.5% specificity on signal classification, it is not possible to infer similarity of samples from the output of a classifier.

In contrast to the prior work, we optimize a deep neural network according to a triplet loss function. Triplets of anchors, positives, and negatives were chosen such that the anchor and positive were of the same class and the anchor and the negative were of different classes. The network is optimized to minimize the distance between the anchor and positive classes and maximize the distance between the anchor and negative classes. In doing so, the network learns an embedding that places similar signals (anchor and positive) close together in the embedding space, and dissimilar signals (anchor and negative) far apart in the embedding space.

We organize the rest of the paper in the following way. In Section II, we review the literature relevant to triplet loss techniques. In Section III, we describe the data and how we organize it for easy triplet mining. In Section IV, we describe our final models, their quantitative and qualitative characteristics, and our experimental results.

II. Related Work

We choose to an approach similar to [2], which used a deep convolutional neural network (DCNN) trained with a triplet loss function to create an embedding space for facial recognition and similarity search. This model was trained to minimize the distance between an anchor and a positive and maximize the distance between an anchor and a negative where $d_{ap} = f(x_a) - f(x_p)$ is defined as the distance between the anchor and positive and $d_{an} = f(x_a) - f(x_n)$ is defined as the distance between the anchor and the negative. Mathematically,

$$\|d_{ap}\|^2 + \alpha < \|d_{an}\|^2 \tag{1}$$

where $f(x) \in \mathbb{R}^d$ represents the computational graph, a is the anchor, p is the positive which is the same class as the anchor, n is the negative which is not the same class as the anchor and α is a hyperparameter expressing the minimum distance between. Thus, the loss function becomes

$$J = \sum_{i=1}^{N} \left[\|d_{ap}\|_i^2 - \|d_{an}\|_i^2 + \alpha \right] \tag{2}$$

where N is the size of the mini batch the network is being trained on. This approach serves as a first experiment in learning an embedding space for EEG signals. We provide, as a point of comparison, a survey of more sophisticated methods for learning embedding spaces which will be relevant to further experimentation.

Lifted structured embedding [3] is a similar image metric learning scheme, where each positive pair is compared against all the negative pairs weighed by the minimum expected distance between them. The advantage is that the loss function is differentiable and incorporates "online hard negative mining". This scheme considers both the local and global structure of the embedding space. As opposed to triplet approach, this method does not require partitioning data into tuples in any manner. This method achieved state of the art performance on standard datasets such as CUB200-2011, Cars196 and Stanford on-line products. However, this method represents a computational trade-off that may not necessary.

Joint Unsupervised Learning (JULE) [4] is a recurrent method to compute image embeddings. However, recurrent networks are difficult to optimize, suffering from training instability.

III. Data

The data for this study was derived from Temple University Hospital's EEG corpus (TUH EEG corpus). This dataset consists of 22 channel EEG signals with electrodes placed according to the 10-20 system. Each channel is annotated as pertaining to one of six classes as described in Table I with a granularity of 1 second. For details on this dataset see [5].

TABLE I. Dataset classes

Codename	Description
BCKG	Background noise
ARTF	Artifacts
EYBL	Eyeball movement
SPSW	Spikes and sharp waves
PLED	Periodic lateralized epileptiform discharges
GPED	Generalized periodic epileptiform discharges

We apply notch filters at 60 Hz and 120 Hz to remove power line noise, and a bandpass filter (1–70 Hz passband) to remove any high-frequency noise as the bulk of the signal power lies within this band. We apply the Short-Term Fourier Transform (STFT) with a window size of 140 samples and a stride of 2 samples results in a 71×125 dimensional matrix for each one-second window of EEG signal. Additionally, we globally normalize the signal power.

The dataset is highly imbalanced (more than 80% is labeled as background noise). We use stratified sampling to help compensate for this imbalance.

We split the one-second examples into mutually exclusive sets for training and validation. Each set is disjoint in both patient and sample acquisition; that is no single patient appears in both sets, and no two windows from a single acquisition appear in both sets. We follow an 85/15% split for training/validation. Due to the large nature of the training set in this situation and the impossibility of training on every triplet possible, a random selection 15% of the training set was used to train the networks.

IV. Experimental Design & Results

We trained a DCNN trained with a triplet-based loss function and evaluated the results. A DCNN is appropriate in this context because the time domain signal has be transformed using the STFT and can be represented in a matrix like an image. Traditionally, DCNNs tend to have high accuracies on images and therefore, it was deemed appropriate to use a DCNN for this particular problem. To confirm that the system learns a meaningful embedding, we perform a simple KNN classification on the outputs of the system on the validation set. Additionally, we look at a t-SNE plot of the resulting embedding in a 2-dimensional space.

To verify that the performance of this system is reasonable, we also design and train a baseline DCNN classifier of similar complexity, and compare the classification results.

A. *DCNN with Triplet Loss*

Our network architecture is specified in Table II. We set the margin $\alpha = 0.5$, the learning rate to 1×10^{-5} and train for 200 thousand iterations. The embeddings are fed into a k-NN (k = 31). With these hyperparameters, we achieve a validation accuracy of 58.6%. The confusion matrix for this is shown in figure Figure 1. Additionally, we the t-SNE plot of the resulting embedding in Figure 2. From the plot, you can clearly see GPED separating from the other classes in the 2-dimensional space. This agrees with the confusion plot indicating that GPED is correctly classified 94% of the time.

TABLE II. NETWORK ARCHITECTURE FOR DCNN

Layer	Input	Output	Kernel
conv1	$71 \times 125 \times 1$	$71 \times 125 \times 32$	5×5
maxpool1	$71 \times 125 \times 32$	$34 \times 61 \times 32$	5×5
conv2	$34 \times 61 \times 32$	$34 \times 61 \times 64$	3×3
maxpool2	$34 \times 61 \times 64$	$16 \times 30 \times 64$	3×3
conv3	$16 \times 30 \times 64$	$16 \times 30 \times 128$	2×2
maxpool3	$16 \times 30 \times 128$	$8 \times 15 \times 128$	2×2
conv4	$8 \times 15 \times 128$	$8 \times 15 \times 256$	1×1
maxpool4	$8 \times 15 \times 256$	$4 \times 7 \times 256$	2×2
conv5	$4 \times 7 \times 256$	$4 \times 7 \times 1024$	4×4
maxpool5	$4 \times 7 \times 1024$	$1 \times 2 \times 1024$	4×4
flatten	$1 \times 2 \times 1024$	2048	N/A
fc1	2048	1024	N/A
fc2	1024	512	N/A
fc3	512	256	N/A
output	256	64	N/A

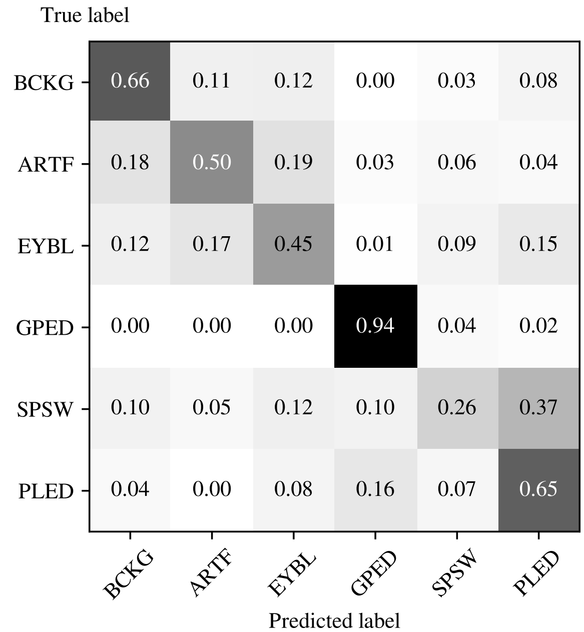

Fig. 1. Confusion matrix for the DCNN clustering network

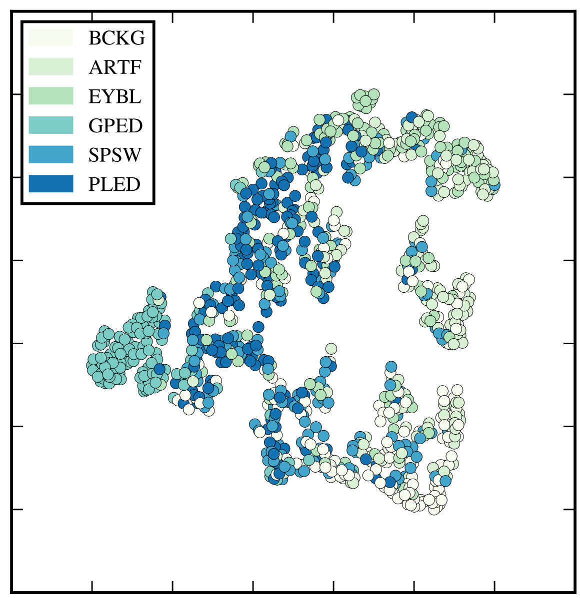

Fig. 2. t-SNE visualization of test examples passed through the clustering network.

B. *Baseline DCNN Classifier with Softmax Loss*

As a baseline we add an extra fully-connected layer to the network in Table II as a classification layer with a softmax activation. As this is now a classification task, the loss function is changed to cross-entropy. We set the learning rate to 1×10^{-5} and train the network for 200 thousand iterations. With these hyperparameters, we achieve a validation accuracy of 50.2%. These results are significant since random guessing would only yield a theoretical accuracy of 16%.

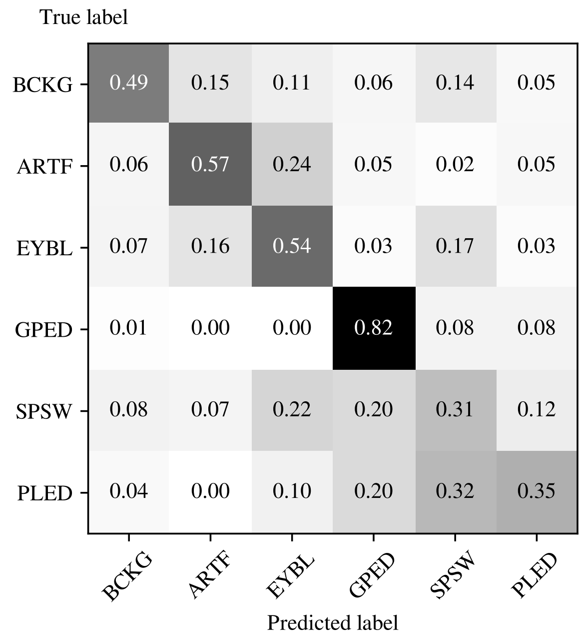

Fig. 3. Confusion matrix for the baseline DCNN classifier

V. Conclusion and Future Work

We demonstrated a method to learn embeddings for EEG signals in an end-to-end fashion. While our system is able to accurately classify the known labels in the TUH dataset, we intend to analyze it for outliers detection and one-shot learning on classes available in the training set. Since our proof-of-concept network is small, it is possible that a more expressive network could obtain better results.

We intend to do an in-depth comparison between the baseline embeddings space (i.e features produced by the penultimate layer) and the experimental one. Each is predicted by functions with identical forms for different parameters. Therefore a comparison between the two embedding spaces speaks directly to the training method for selecting the parameters.

As the TUH corpus also includes physician notes, we would like to investigate ways to incorporate these notes into a cohort retrieval scheme. This could be done by learning a similar embedding for the text data and then performing a clustering on a joint embedded space. Another possible area for investigation would be to leverage an adaptive density discrimination technique [6] to shape clusters in the EEG embedding space using the annotations as side information.

Finally, it is possible that these ideas can be extended to other types of raw medical data such as MRIs and X-Rays.

Acknowledgment

We would like to thank The Cooper Union and Temple University Hospital for providing the resources necessary to perform these experiments.

References

[1] J. Picone, I. Obeid, and S. M. Harabagiu, “Automatic discovery and processing of eeg cohorts from clinical records,” *Big Data to Knowledge All Hands Grantee Meeting (p. 1), Bethesda, Maryland*, 2015.

[2] F. Schroff, D. Kalenichenko, and J. Philbin, “Facenet: A unified embedding for face recognition and clustering,” *CoRR*, vol. abs/1503.03832, 2015. [Online]. Available: http://arxiv.org/abs/1503.03832

[3] H. O. Song, S. Jegelka, V. Rathod, and K. Murphy, “Learnable structured clustering framework for deep metric learning,” *CoRR*, vol. abs/1612.01213, 2016. [Online]. Available: http://arxiv.org/abs/1612.01213

[4] J. Yang, D. Parikh, and D. Batra, “Joint unsupervised learning of deep representations and image clusters,” *CoRR*, vol. abs/1604.03628, 2016. [Online]. Available: http://arxiv.org/abs/1604.03628

[5] I. Obeid and J. Picone, “The temple university hospital eeg data corpus,” *Frontiers in Neuroscience*, vol. 10, p. 196, 2016. [Online]. Available: http://journal.frontiersin.org/article/10.3389/fnins.2016.00196

[6] O. Rippel, M. Paluri, P. Dollar, and L. Bourdev, “Metric learning with adaptive density discrimination,” *arXiv preprint arXiv:1511.05939*, 2015.