Adversarial autoencoder analysis on human $\mu$ECoG dataset

Yilin Song $^1$, Yao Wang $^1$, and Jonathan Viventi $^2$

$^1$Department of Electrical and Computer Engineering, New York University, NY, USA
$^2$Department of Biomedical Engineering, Duke University, Durham, NC, USA

July 2, 2018

1 Background

Building scalable generative models to capture rich distributions in human $\mu$ECoG dataset could potentially lead to discovering neural patterns and hence better understanding of how human brain function. In recent years, generative models have been developed that may be trained via direct back-propagation and avoid the difficulties that come with MCMC training, which used by Deep Belief Network and Deep Boltzman Machines. For example, variational autoencoder [1] use a encoder network to predict the posterior distribution over latent variables, generative adversarial network [2] use an adversarial training procedure to directly shape the output distribution. Adversarial autoencoder (AAE[3]) combines this two and turn an autoencoder into a generative model. In their approach, an autoencoder is trained with dual objectives – a traditional reconstruction error criterion, and an adversarial training criterion. The result of the training is that the encoder learns to convert the data distribution to the prior distribution, while the decoder learns a deep generative model that maps the imposed prior to the data distribution. Fig. 1 shows the general structure of AAE.

2 Reconstruction performance and pattern analysis

In our first approach, we assume the "style" of human $\mu$ECoG dataset follows multivariate gaussian distribution and we use structure shown in Fig. 1a to perform spatial-temporal motion pattern analysis on human dataset. Each input sample to AAE is a video clip arranged as a 3d tensor. We have in total around 300 thousands training clip and 50 thousands testing clip. Each input clip has the shape of $16 \times 16 \times 64$ ordered in width, height and time. We use 3-D CNN based encoder-decoder structure as basic building module for AAE. Fig. 2 shows the spatial-temporal pattern learnt through a 2-D Gaussian adversarial autoencoder. We find there are continuous pattern found by 2-D Gaussian adversarial autoencoder. But viewing a higher dimensional pattern other than 2-D poses problems on data visualization, hence we fail to show more interesting pattern potentially could be found by higher dimension representation. Next we demonstrate the reconstruction performance of AAE model on human dataset. In Fig. 3, we show 128 randomly sampled testing video clips and their reconstruction performance using 20-D Gaussian adversarial autoencoder. By using a feature vector of size 20, we are able to preserve the pattern for each $16 \times 16 \times 64$ video clip.
(a) Plain AAE with single distribution assumption

(b) Extension of AAE by disentangle style from clustering pattern

Figure 1: Figure 1a: Architecture of an adversarial autoencoder. The top row is a standard autoencoder that reconstructs an image $x$ from a latent code $z$. The bottom row diagrams a second network trained to discriminatively predict whether a sample arises from the hidden code of the autoencoder or from a sampled distribution specified by the user. Figure 1b, clustering AAE: the top adversarial network imposes a Categorical distribution on the label representation and the bottom adversarial network imposes a Gaussian distribution on the style representation.

Figure 2: Video clip generated by uniformly sampling the Gaussian percentiles along each hidden code dimension $z$ in the 2-D Gaussian adversarial autoencoder. Red axis indicates the 2-D Gaussian adversarial autoencoder distribution. The displayed style video clips are down-sampled by 16.
3 Clustering human signals

Next we seek to find out whether there exist disentangled discrete classes other than the continuous latent style we shown in the last section. Here we use structure shown in Fig. 1b, in which AAE predicts both the discrete clustering label $y$ and continuous latent variable $z$. We tried different number of clusters varing from 10-10 thousands but we fail to find cluster of patterns. We speculate that normal human brain activities are too complicated and there might not exist any cluster of pattern. As a comparison we shown a comparison in Fig. 4 between human dataset and cat dataset [4].

4 ACKNOWLEDGEMENT

This work was funded by National Science Foundation award CCF-1422914.

References


Figure 4: Brain signal activity comparison. Human activities in normal brain are much more complex than cat brain signal and doesn’t reveal clear clusters.