CS587 - MidTerm

Epileptic Seizure Prediction with a Convolutional Neural Network

Ronald Broberg
University of Colorado at Colorado Springs

***Abstract*—A five layer convolutional neural is used to classify EEG segments as interictal or preictal.**

I. Introduction

A spedific five layer Convolutional Neural Network (CNN) was described by LeCun in 1998[1]. Applied to a variety of sources. Mirowski used a similar CNN to predict the onset of epileptic seizures by detecting preictal clips[2]. EEG data recorded from epileptic patients are often divided into four categories: interictal, preictal, ictal, and postictal[3]. Reliable detection of preictal states has the potential to allow for medical intervention which may prevent the seizure. A CNN similar to Mirowski is applied to a more complex data set described below.

II. Data

Data was prepared by the University of Pennsylvania from a variety of sources. Data from both canine and human subjects were included. Canine subjects were generally recorded with 16 channels at 400 Hz. One canine subject was included with only 15 channels. Recordings from two human patients were also included. The first patient's EEG was recorded with 15 channels at 5000 Hz. The second patient's EEG was recorded with 24 channels also at 5000 Hz. Preictal series were created by selecting one hour of EEG data prior to the 315 seconds before the onset of seizure. Ictal series were created by selecting one hour segments with at least four seperating it from a seizure. Both preictal and interictal series were divided into six segments each of 10 seconds duration. Segments were then divided into labeled training data and unlabeled test data. Segments were left in sequential order in the training data but were randomly shuffled in the test set.

Case	Freq	Chan	Preictal	Interictal	Test	Sz	Trnd
Dog 1	400	16	24	480	502	4	3
Dog 2	400	16	42	500	1000	8	5
Dog 3	400	16	72	1440	907	12	8
Dog 4	400	16	97	804	990	16	11
Dog 5	400	15	30	450	191	5	3
Patient 1	5000	15	18	50	195	3	2
Patient 2	5000	24	18	42	150	3	2

III. Feature Extraction

Numerous combinations of univariate (mean, variance, skew, kurtosis, entropy, autocorrelation) and multivariate (cross correlation, Lyapunov exponents, wavelet features) features have been used in detection and prediction of epileptic seizures in EEG recordings. This study uses a 0-time lag cross correlation measure in a time series with approximately five second windows.

The EEGs were recorded at differing frequencies and with a differing number of channels across the set of cases. In addition to the differing phenomenologies of the seizures presented in the heterogenous cases, this variability in dimensions in the data encouraged the development of a unique model for each case.

Data from all sources were downsized from their original frequencies to 200 Hz. Each five minute segment was then divided into a number of time windows equal to the number of unique pairs of channels, excluding identity pairs. For example, a 16 channel data set would have $\frac{(16^2-16)}{2} = 120$ pairs of channels. Thus, a 16 channel time series would be broken into 120 time windows which, for a ten minute sample at 200 Hz equals a time window of $\frac{600s}{120} = 5$ seconds. Similarly, a 105 channel data set would have 5.7 second windows and the 246 channel data set has 2.4 second windows. While chosen to keep the feature array square, this time window is similar to the 5 second window used in Miroswki 2008.

For each 5 second window, the down sampled EEG time series for each channel was correlated against each of the other channels without time lag. The values of these cross correlations across the entire time range of each training segment was then stored as a vector. These vectors are the input data into the LeNet5 model described below.

A correlation matrix is presented as an image for one interictal and preictal sample from the recordings for Dog 1.

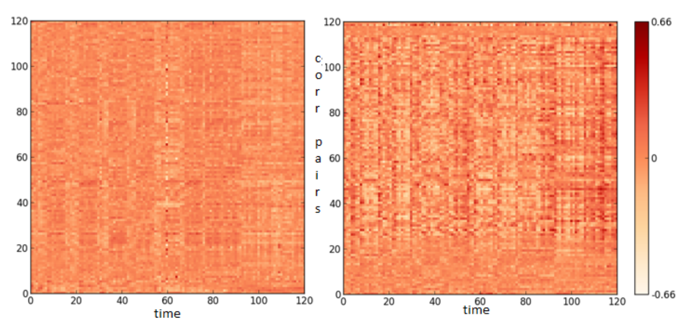

Figure 1: A sample of the time variant cross correlation matrix for an inter ictal segment (left) and a preictal segment (right).

IV. Five Layer LeNet Model

Convolutional neural networks can be trained to extract high dimensional patterns from complex, non-linear input and classifly them [1] [2]. As such, they are often used in

R. Broberg is with the Department of Mathematics, University of Colorado at Colorado Springs, Colorado Springs, CO, 80909.

classification problems involving images or multivariant time series.

The CNN architecture used here is similar to that of Mirowski 2008 [2]. It is a five layer neural network with three convolutional layers and two pooling layers. The model is necessarily configured differently for inputs of differing dimensions. For 16 channel recordings with 120 cross-correlation pairs, the 1^{st}, 3^{rd}, and 5^{th} layers have convolution kernels of size 1x25:1x9:1x1. The subsampling on the 2^{nd} and 4^{th} layers have dimensions of 1x4 and 1x2. For 15 channel recordings with 105 cross-correlation pairs, 1^{st}, 3^{rd}, and 5^{th} layers have convolution kernels of size 1x10:1x9:1x1 and the subsampling layers are also 1x4 and 1x2. Finally, for the 24 channel recordings, the 1^{st}, 3^{rd}, and 5^{th} layers have convolution kernels of size 1x65:1x9:1x1. The subsampling on the 2^{nd} and 4^{th} layers have dimensions of 1x8 and 1x2. Figure 2 presents the model as configured for a 16 channel input.Table 2 presents the configuration parameters for each input type. The parameters were selected to reduce dimensionality most in the lower layers, keeping similar dimensions in the higher layers of each model.

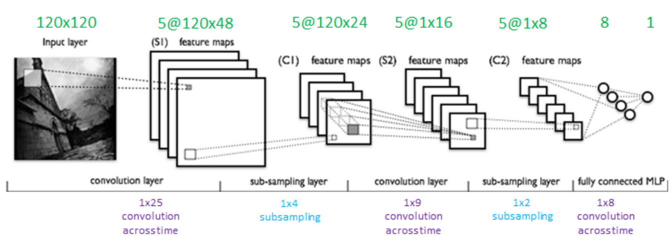

Figure 2: Lenet 5 Convolutional Neural Network

Channels	Layer 1	Layer 2	Layer 3	Layer 4	Layer 5
16	25	4	9	2	1
15	10	4	9	2	1
24	65	8	9	2	1

The model was coded in the Python Theano library which is optimized for GPU matrix operations and closely followed the LeNet 5 Deep Learning tutorial presented by the Theano development team [4]. The model was run on an Amazon EC2 AWS g2.2xlarge instance. Code was archived on Github. Data was accessed by sshfs from a server owned by the author.

V. Cross-Validation

Dividing the training data into training and validation sets had to be handled with care as the preictal segments were derived in series from a small number of seizure events. Recall that each one hour preictal period was divided into six ten minute segments. These segment have a high degree similarity and including segments from the same series in both the training and validation sets is essentially sharing information between the two leading to overly confident CV results. Instead, a preictal series of a six segments that was to be withheld for validation had to be withheld in it's entirity.

In addition, the training data set is heavily weighted towards interictal segments. For Dog 1, the preictal:interictal ratio is 1:20. To create a more balanced data set, preictal segments were repeated in the training segment with the addition of gaussian noise until the preictal:interictal ratio was near 1:1.

VI. Results

The test data was kept seperate from the authors and its composition is unknown. Test classifications are submitted via an automated web interface and an AUC score is returned. Purely random submissions generate scores near .500. Perfect submissions would generate scores near 1.000. At this time, the best results using this approach score .640 but that is for a submission with models for just two dogs.

VII. Future Work

There are several directions that can be investigated for model improvement.

A. Feature Selection

While the intitial model uses cross-correlation for the bivariate variable, other bivartiates could be investigated. Mirowski also looked at Lyapunov exponents and a wavelet based meaure of synchronicity. The cross-correlation variable used here may be improved by looking for the maximal correlation as in Mirowski and not just the 0-time lag correlation.

Univariate variables such as variance, skew, kurtosis and entropy could be dimensioned in a fashion allowing them to be concatenated with the multivariate variables to provide modeling on more than one variable at a time.

B. Complete Model for All Cases

Initial results were presented using models for just cases. Test results calculated with models for all cases is an obvious priority.

C. Best Model Forward

As presently configured, the model runs for a predetermined number of epochs and uses the final weights for predictions. A better design would save the weights from the best CV scores during training and use that model for prediction.

D. Model Ensemble

Currently, the preictal states left out of training are selected at random. An alternate approach would be to train a set of models with different precital states withheld and then average the results across all of them to predict preictal states. For instance, Dog 1 has four series of preictals segments. Four models could be trained, each with a different precital series withheld. A particular test segment would then be presented to each model for classification and the results averaged. This would create a probablistic classifier.

REFERENCES

[1] Y. LeCun,L. Bottou,Y. Bengio,P. Haffner, *Gradient-Based Learning Applied to Document Recognition*, Proceedings of the IEEE, 86(11):2278-2324, November 1998

[2] P.W.Mirowski, et al. *Comparing SVM and convolutional networks for epileptic seizure prediction from intracranial EEG." Machine Learning for Signal Processing*, 2008. MLSP 2008. IEEE Workshop on. IEEE, 2008.

[3] B. Vazquez, O.Devinsky, *Epilepsy and anxiety.* Epilepsy & Behavior 4 (2003): 20-25.

[4] J. Bergstra, O. Breuleux, F. Bastien, P. Lamblin, R. Pascanu, G. Desjardins, J. Turian, D. Warde-Farley and Y. Bengio. *Theano: A CPU and GPU Math Expression Compiler*. Proceedings of the Python for Scientific Computing Conference (SciPy) 2010. June 30 - July 3, Austin, TX