Predict Seizures In Long-term Human Intracranial EEG Recordings

Michael Sommer
Information and Computer Sciences Department
University of Hawai'i at Mānoa
Honolulu, HI, U.S.A.
[mpsommer]@hawaii.edu

Julie Schnurr
Department of Geology and Geophysics
University of Hawai'i at Mānoa
Honolulu, HI, U.S.A.
[jschnurr]@hawaii.edu

Tyson Seto-Mook
Information and Computer Sciences Department
University of Hawai'i at Mānoa
Honolulu, HI, U.S.A.
[tmook]@hawaii.edu

I. Introduction

A seizure is a phenomenon in which the abnormal electrical activity in the brain. Seizures can cause unconsciousness or in some cases uncontrollable muscle spasms. Spontaneous recurrence of seizures is a medical disorder called epilepsy. Epilepsy affects around 1% of the human population with varying frequencies and severities. The unpredictability of seizure occurrences also cause anxiety in epilepsy patients. In addition, normal activities such as driving a vehicle or swimming can be unsafe for these patients. Medication can be taken to prevent convulsions, but are not effective for 20-40% of patients with epilepsy. A wearable seizure forecasting system would have a positive impact on the lives of such patients. An early warning system would allow patients to prepare for seizures before they occur by moving to a safe location and stopping any potentially dangerous activities. Even patients who are successfully treated with anticonvulsive medications often experience negative side effects. If an effective seizure forecasting system were developed, anticonvulsive medications could be taken only in the case of an impending seizure, thereby reducing side effects.

II. Problem

Researchers at Melbourne University have been studying epileptic seizures with the goal developing an algorithm for use with a wearable seizure forecasting device. Brain activity can be monitored with an electroencephalogram (EEG) machine. Electrodes are placed on the scalp to record the electrical impulses of brain waves. Intracranial EEG (iEEG) is monitored by placing these electrodes on the cerebral cortex. A public challenge was posted on kaggle.com for the data science community. The challenge is to distinguish between interictal (normal) activity and preictal (preseizure) activity up to an hour before a seizure occurs using a training set of labeled iEEG data from three patients.

III. Approach

We extracted three sets of features from the labeled iEEG data sets. We then trained a generalized linear model with LASSO regularization and evaluated the correctness of our models using the ROC curve and AUC score.

A. Dataset

Three sets of training and test iEEG datasets corresponding to three different patients were provided by Melbourne University. Each training set consists of data labeled interictal or preictal. For the interictal data, it was ensured that the previous seizure occurred a minimum of 4 hours before or after a seizure. Preictal data was taken from the hour before a seizure occurred with a 5 minute buffer. The buffer time ensures that activity from the seizure itself is not included. The recordings were made using 16 iEEG electrodes sampled at 400 Hz which were used to monitor intracranial brain activity over a long period of time (months to years). The electrodes measure a voltage that is referenced from the average of all 16 electrodes at the sample time. Recordings are 1 hour long and broken up into 10 minute chunks which are labeled with their respective position within the hour. However, we did not take the position of the chunks into account in our model. We also only used data from one of the patients (patient 1) due to time constraints.

B. LASSO GLM

Least Absolute Shrinkage and Selection Operator (LASSO), also known as the penalized regression model, is a shrinkage and variable selection method that imposes a constraint on the regression coefficients. The penalization process causes some of the regression coefficients to shrink to zero. The penalization process results in identifying the variables that minimize prediction error. In machine learning, LASSO is a supervised learning method used to automatically select features that have greater prediction accuracy. Generalized Linear Model (GLM) is a generalization of linear regression that allows response variables to have error distribution model.

C. Parameter Optimization

LASSO GLM is parameterized so that its performance can be optimized for a given data set. Finding the correct values for the parameters is treated as a search problem. To optimize our parameters, we used grid search, which exhaustively builds and evaluates a model for each combination of parameters specified in a grid. Through this approach, we determined that LASSO GLM performed best on our data when the L1 term was multiplied by the constant 0.001, the intercept was calculated, the regressors were normalized, and the optimization tolerance was set to 0.01.

IV. Analysis/Results

Our feature sets consisted of various signal energy and statistical metrics applied to both unfiltered as well as band passed data. We iterated through three versions of feature sets as we realized many of the LASSO coefficients were going to zero. For all three feature sets, the measurements were performed on the 16 individual channels and then averaged together. This led to a smaller feature set and quick convergence but could be a problem in the case of more localized changes in brain activity leading up to a seizure. Changes in a few channels could be missed if they average out. In the worst case, large differences between channels could even be strongly correlated with preictal activity. Differences between channels could be measured using correlation coefficients, but performing cross correlations is computationally expensive. Variance in features between channels was explored as a more efficient alternative to correlation. However, including these variances in the feature set led to either non-convergence or decreased performance as measured by the AUC score of the model, so only the average features were included in the final feature sets.

A. Feature Set 1

We computed the energy and power of the signal directly from the numerical integral of the signal squared and the variance respectively. We also measured the power spectral density (PSD) using scipys periodogram function. Not every point in the PSD was used, instead the median was taken every 80 points with a 20 point overlap. This provided a significant reduction in the number of features and widened the frequency bands. Less narrow frequency bands should reduce the likelihood of overfitting, particularly as the number of preseizure examples in the training set was fairly small. Measurements were first made over the entire 10 min duration of the data. We then also observed the linear trend in the metrics measured in 30s windows with 50% overlap. Each measurement was made separately on each of the 16 iEEG channels and then averaged, leading to a total of 104 features. The only features that did not have LASSO coefficients equal to zero were the total energy, variance, and PSD in the 83 Hz. center frequency band (band-width 4Hz.). For the next feature set, less narrow frequency bands were used as we suspected overfitting with regard to PSD in each frequency band. The ROC curve and AUC score for Feature Set 1 are shown in Figure 1.

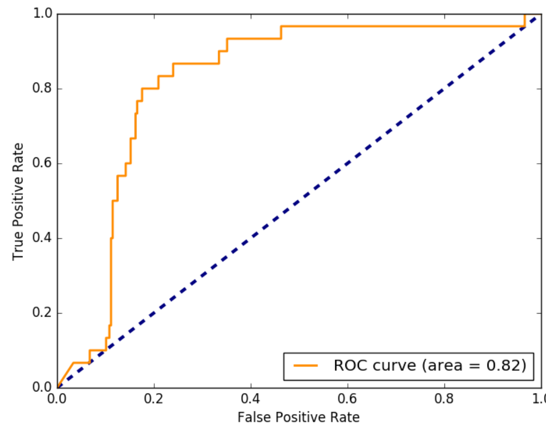

Figure 1: ROC Curve for Feature Set 1
LASSO parameters = Tolerance: 0.01, Alpha 0.001, Fit Intercept: true, Normalize: true

B. Feature Set 2

We band-passed the data using 4th order-butterworth filters corresponding to the bandwidths of the different types of brainwaves: Delta, Theta, Alpha, Beta, Gamma, Mu. We also included an additional 100-200 Hz. band (not typically studied) for completeness. Measurements were again made first over the entire 10 min duration of the data and then we measured the the linear trend in the metrics in 30s windows with 50% overlap. Each measurement was made separately on each of the 16 iEEG channels and then averaged, leading to a total of 48 features shown in Table I below. Features with non-zero LASSO coefficients are highlighted in blue.

As with feature set 1, feature set 2 is designed to distinguish interictal from preictal activity using both the overall energy (integral of the signal squared) and power (variance) of the signal and whether or not the signal is changing over time. The rationale behind the energy and power measurements is that seizures should be preceded by an overall increase in signal in the brain, possibly at particular frequencies corresponding to different types of brain waves. It was theorized that the energy and

Table I: Feature Set 2 - Features with non-zero LASSO coefficients are highlighted in blue.

Unfiltered (All freqs.)	Delta (0-4Hz)	Theta (4-7Hz)	Alpha (7-14Hz)	Beta (15-30Hz)	Gamma (30-100Hz)	Mu (8-13Hz)	High (100-200Hz)
Energy	Energy	Energy	Energy	Energy	Energy	Energy	Energy
Variance	Variance	Variance	Variance	Variance	Variance	Variance	Variance
Energy Trend	Energy Trend	Energy Trend	Energy Trend	Energy Trend	Energy Trend	Energy Trend	Energy Trend
Variance Trend	Variance Trend	Variance Trend	Variance Trend	Variance Trend	Variance Trend	Variance Trend	Variance Trend
Skew Trend	Skew Trend	Skew Trend	Skew Trend	Skew Trend	Skew Trend	Skew Trend	Skew Trend
Kurtosis Trend	Kurtosis Trend	Kurtosis Trend	Kurtosis Trend	Kurtosis Trend	Kurtosis Trend	Kurtosis Trend	Kurtosis Trend

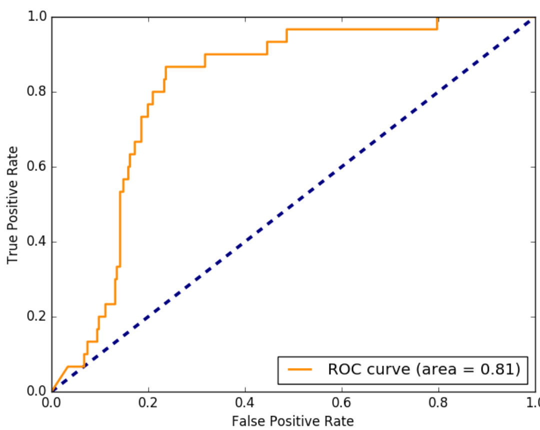

Figure 2: ROC Curve for Feature Set 2
LASSO parameters = Tolerance: 0.1, Alpha 0.001, Fit Intercept: true, Normalize: true

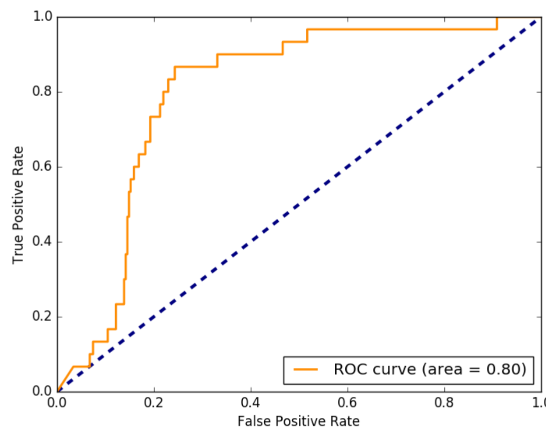

Figure 3: ROC Curve for Feature Set 3
LASSO parameters = Tolerance: 0.01, Alpha 0.001, Fit Intercept: true, Normalize: true

power of the signal should increase over time in the case of preictal activity. The signal was divided into 30 second windows with 50% overlap and energy and power was measured in each window. Skewness and kurtosis was also measured in these windows to provide more ways to check signal variability over time. A linear regression was performed on these measurements in order to get the trend (slope) with respect to time. These measurements were made on the unfiltered signal and repeated for each of the different types of brain waves. The ROC curve and AUC score for Feature Set 2 are shown in Figure 2.

C. Feature Set 3

The energy, variance, skew, and kurtosis trend feature coefficients consistently went to zero. This suggests that the trend in these variables are not good metrics for differentiating between interictal and preictal activity. Therefore the last feature set did not include these. Instead the energy and statistical measures were applied only to the whole 10 minute duration in various frequency bands as shown in the table below. Again, the measurements were made separately on each of the 16 channels and then averaged together this time for a total of 34 features. The features with non-zero LASSO coefficients are in blue on the feature Table II. The ROC curve and AUC score for Feature Set 3 are shown in Figure 3.

V. Conclusion

The three feature sets we applied to the data each produced very similar results in terms of roc curve and auc score. The first feature set we applied produced good results but very few of the parameters were used. The initial good result led us to tweak the feature set by changing the width of frequency bands (feature set 2) and replacing statistical trends with total statistics (feature set 3) rather than coming up with entirely new features to add to the set. Given more time, the next step towards obtaining a better auc score would likely be to completely rethink the feature set, keeping only the features that were consistently used (energy and power).

References

[1] M. J. Cook, T. J. O'Brien, S. F. Berkovic, M. Murphy, A. Morokoff, G. Fabinyi, W. D'Souza, R. Yerra, J. Archer,

Table II: Feature Set 3 - Features with non-zero LASSO coefficients are highlighted in blue.

Unfiltered (All freqs.)	Delta (0-4Hz)	Theta (4-7Hz)	Alpha (7-14Hz)	Beta (15-30Hz)	Gamma (30-100Hz)	Mu (8-13Hz)	High (100-200Hz)
Energy	Energy	Energy	Energy	Energy	Energy	Energy	Energy
Variance	Variance	Variance	Variance	Variance	Variance	Variance	Variance
Skew	Skew	Skew	Skew	Skew	Skew	Skew	Skew
Kurtosis	Kurtosis	Kurtosis	Kurtosis	Kurtosis	Kurtosis	Kurtosis	Kurtosis

L. Litewka *et al.*, "Prediction of seizure likelihood with a long-term, implanted seizure advisory system in patients with drug-resistant epilepsy: a first-in-man study," *The Lancet Neurology*, vol. 12, no. 6, pp. 563–571, 2013.

[2] B. H. Brinkmann, J. Wagenaar, D. Abbot, P. Adkins, S. C. Bosshard, M. Chen, Q. M. Tieng, J. He, F. Muñoz-Almaraz, P. Botella-Rocamora *et al.*, "Crowdsourcing reproducible seizure forecasting in human and canine epilepsy," *Brain*, vol. 139, no. 6, pp. 1713–1722, 2016.

[3] K. Gadhoumi, J.-M. Lina, F. Mormann, and J. Gotman, "Seizure prediction for therapeutic devices: A review," *Journal of neuroscience methods*, vol. 260, pp. 270–282, 2016.

[4] P. J. Karoly, D. R. Freestone, R. Boston, D. B. Grayden, D. Himes, K. Leyde, U. Seneviratne, S. Berkovic, T. OBrien, and M. J. Cook, "Interictal spikes and epileptic seizures: their relationship and underlying rhythmicity," *Brain*, p. aww019, 2016.

[5] R. G. Andrzejak, D. Chicharro, C. E. Elger, and F. Mormann, "Seizure prediction: any better than chance?" *Clinical Neurophysiology*, vol. 120, no. 8, pp. 1465–1478, 2009.

[6] D. E. Snyder, J. Echauz, D. B. Grimes, and B. Litt, "The statistics of a practical seizure warning system," *Journal of neural engineering*, vol. 5, no. 4, p. 392, 2008.

[7] F. Mormann, R. G. Andrzejak, C. E. Elger, and K. Lehnertz, "Seizure prediction: the long and winding road," *Brain*, vol. 130, no. 2, pp. 314–333, 2007.

[8] S. R. Haut, S. Shinnar, S. L. Moshé, C. O'Dell, and A. D. Legatt, "The association between seizure clustering and convulsive status epilepticus in patients with intractable complex partial seizures," *Epilepsia*, vol. 40, no. 12, pp. 1832–1834, 1999.