Predicting Seizures from Intracranial EEG Recordings

CSCI-B 490

Authors:

Aaron Schwartz

Ching Yuen Ng

Sebastian Kagemann

Predicting Seizures in Intracranial EEG Recordings

**Objective and Significance**

Our proposal is to create an algorithm that predicts seizures by use of intracranial EEG recordings by analyzing ten-minute segments of 1 hour long EEG data. The goal is to improve patients’ living standards and lead them to live more normal lives. By providing a means of identifying periods of increased probability of seizure occurrence up to an hour before they happen, treatment can be administered to counteract the effects of a seizure.

The creation of a monitoring device capable of identifying the preictal stage from the interictal stage would allow for a patient diagnosed with epilepsy to continue a more normal life.  Activities that would regularly be risky or that require continuous attention could be managed with fast-acting pre-seizure medications.  This device, along with the appropriate medication, can allow for epileptics to maintain everyday activities such as driving or working in a job with potential risk hazard should a seizure occur.  Prediction of seizures also allows for more time-targeted dosage of medication that can reduce expenses and demand for the medicine as well as reduce the side effects from everyday usage.  Portable, personal EEG monitoring devices already exist, and the classifier created from this competition hopes to serve as an algorithm for such a device to better the lives of epileptics.

**Background**

Epilepsy affects approximately 1% of the world’s population. Epilepsy is a brain disorder caused by abnormally excitatory signaling of neurons.  While the exact cause of epilepsy is unknown, it is understood that the disease has a genetic influence, although reparation of nerves from head trauma, especially during developing years, can also cause epilepsy in people with no epileptic family history.  The symptoms of epilepsy include: erratic sensations, emotions, and behavior, as well as convulsions, muscle spasms, loss of consciousness, and occasional seizures.  The frequency of seizures amongst epileptics is highly variable with 80% of epileptics suffering from this symptom and 25% will experience seizures despite available medications and treatments.  Many patients have the option of taking high doses of anticonvulsant medications to counteract seizures but patients often suffer side effects. Current surgeries involving removal of epilepsy-causing brain tissue have been ineffective in preventing spontaneous seizures. This has led to a greater movement in identifying seizures with Electroencephalography (EEG), the recording of electrical activity along the scalp, and predicting them in advance so that we might administer appropriate doses of anticonvulsants to prevent seizures.

EEG is typically recorded along the scalp; however, this dataset uses intracranial recordings which provide more accurate, less noisy data. Brain activity can be classified into 4 states: interictal (between seizures, or baseline), preictal (prior to seizure), ictal (seizure), and post-ictal (after seizures). Fortunately, the progression of symptoms of seizures and auras, or moments of consciousness during seizures, is similar every time.  This consistency may allow for easier prediction of seizures assuming that the similar progressions correspond to similar EEG readings.  Correlation between seizure clusters in several epileptic dogs confirmed the non-randomness of epileptic seizures, and forecasting seizures upon these seizures was more accurate than a trivial classifier.  The focus of this project is to be able to identify the preictal stage from the interictal stage so that a device might have sufficient time to process the EEG data collected and predict an oncoming seizure based on the current data; this why our preictal data spans an hour before the onset of a seizure.  A predictive algorithm would provide general coverage to epileptics, however, personalized devices may need to be developed since each individual's progression of symptoms may be different, and these progressions may result in different EEGs.

**Data**

Given on the website, five EEG sample data from dogs and two from humans.

* EEG was sampled from 16 electrodes at 400 Hz
* Recorded voltages were referenced to the group average
* Long duration recordings, spanning multiple months up to a year
* Recording up to a hundred seizures in some dogs
* Training data is organized into ten minute EEG clips
* “Preictal” pre-seizure data segments
* “Interictal” for non-seizure data segments
* Preictal training and testing data segments are provided covering one hour prior to seizure
* Pre-seizure horizon ensures that seizures could be predicted with enough warning to allow administration of fast-acting medications

**Method and Implementation**

We began our analysis by comparing each segment's minimum value, maximum value, and standard deviation using a random forest of 1,000 trees as our classifier.  Next, we isolated each channel and compared them individually to see if particular channels were more representative of their respective states than others.  While we found minor deviations in some channels, we did not consider them to be significant enough to warrant keeping them as part of our classifier.

Afterwards, due to the data's nature as a time-series, we computed the spectral power for each channel and measured the amount of peaks found after each transformation, which we designated as frequencies that were at or above half the channel's maximum frequency.  We came upon this measure by observing the preictal and interictal spectral powers under median center DC.  When comparing the two states with center DC, we noticed a shift in the commonality of frequencies between the two, with preictal frequencies more commonly exhibiting values near the mean and interictal more often showing values several deviations from the mean.  By our definition of measuring peaks, we believed we'd be able to differentiate between the two states.

The other trials we ran consisted of combinations of the features discussed, along with variance and correlation.  Our strongest classifier was a random forest of 1,000 trees using variance, correlation, and number of peaks as features. We averaged the KNN predictions for several distance measures and averaged those with our Decision Tree predictions. KNN does not learn anything from the training data, which resulted in the algorithm not generalizing well and also not being robust to the noisy data. We would have liked to run additional Fourier transformations, but our inexperience with time-series data made implementing and understanding spectral power a lengthy ordeal.

**Evaluation Strategy**

We evaluated our training data classifier based on Kaggle's own hidden scorer.  Our submissions consisted of our classifier's benchmark scoring against our given test data.  However, when deciding which submissions we thought were acceptable, we compared the features we were testing individually on our training data to confirm worthiness of a submission since we were limited by Kaggle's competition rules. Given that Kaggle evaluated submissions on the area under the ROC curve (AUC), which is what we would have used in cross-validation, we found Kaggle’s submission system to be sufficient in validating our predictions.

**Results**

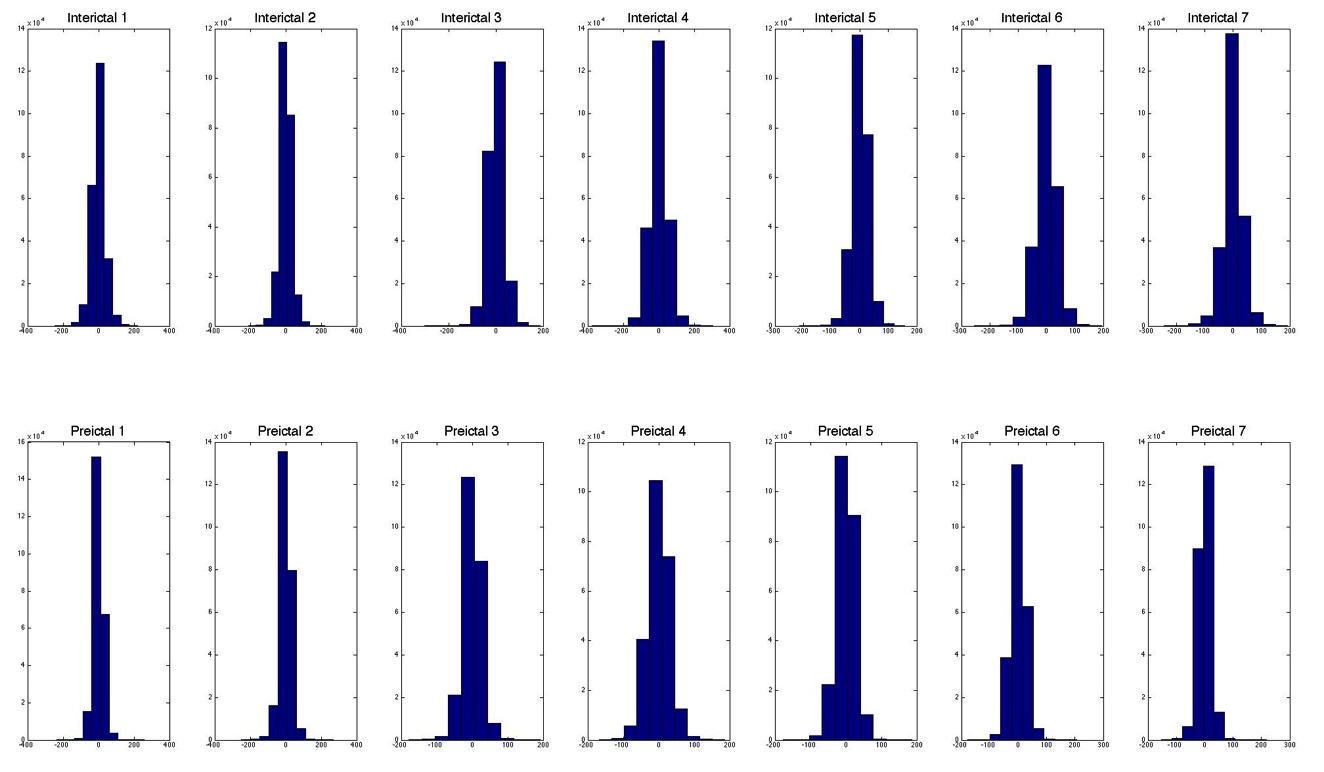
 We explored many different approaches and permutations of feature sets including but not limited to: variance, correlation matrices of signals between electrode channels, minimum values, maximum values, standard deviation, isolating channels that differ substantially between preictal and interictal, average spectral power, # of peaks above ½ max, power/median (Center DC).The process to derive these features involved looking at the data through scripts and plotting histograms to evaluate how features between preictal and interictal states differed across electrode channels.

Figure 1. Histograms of preictal and interictal frequency distributions for electrodes 1-7

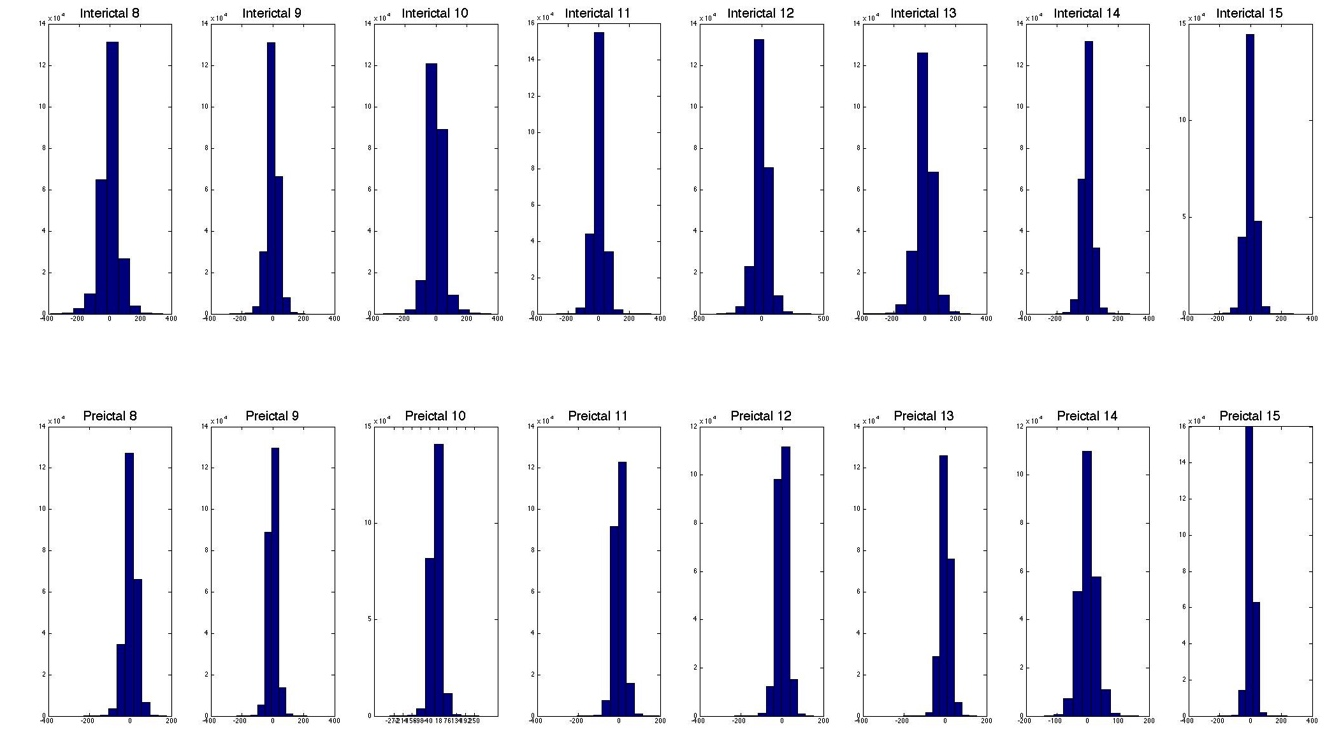
In order to analyze the raw data more efficiently, we created histograms to generalize the frequency distribution in hopes to find a pattern.  We noticed a slight tendency for interictal histograms to have more values offset of the mean than compared to the preictal histograms.  This offset would allow the use of variance to be used as a feature to differentiate between the two states since we expected the interictal data to exhibit more variance than the preictal data.  Reinforcing these differences, computing the correlation between each state would also be more likely to correlate these varying points with each state, allowing correlation matrices to also be used as a feature.

Figure 2. Histograms of preictal and interictal frequency distributions for electrodes 8-16.

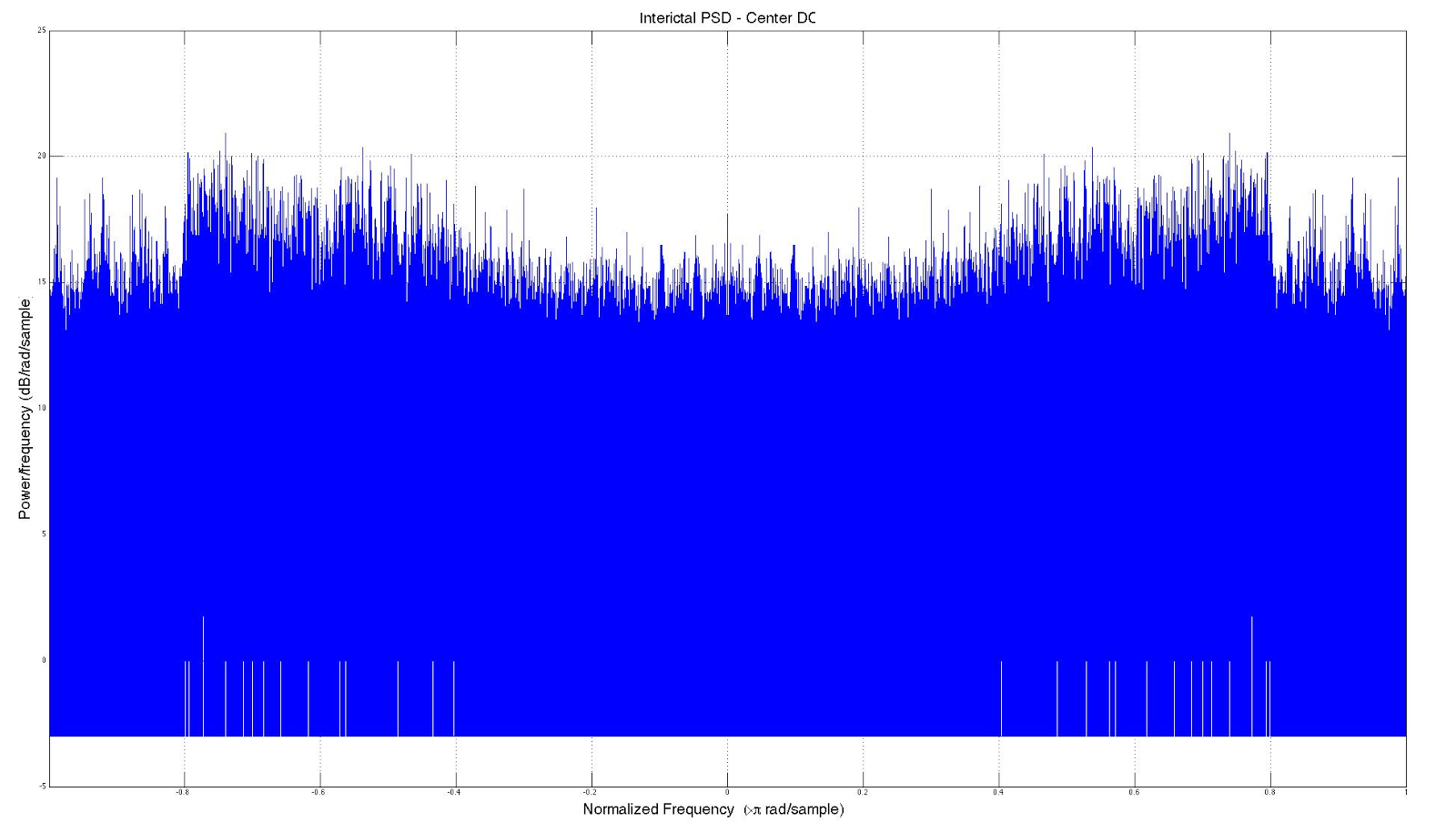
While analyzing the raw data with histograms we decided to try isolating just a few electrodes that exhibited the greatest difference in frequency distribution between preictal and interical states. We determined that electrode channels 1, 3 and 7 had the greatest differences between preictal and interictal after running a script that compared segments with a t-test. To capture qualities of the frequency with less noise than the raw signal we chose a few standard statistical measures: max, min and standard deviation and collected those with our feature engineering script. Following our standard procedure of feeding them into a random forest of 1000 classification trees, we found after validation that isolating these channels was only .78% more effective than extracting min, max and standard deviation from all channels. This informed our decision to develop features around all electrode channels rather than isolating those that showed the greatest polarization between preictal and interictal states.

Figure 3. Interictal power spectral density

During our power spectral density analysis of the data, we plotted spectral powers with the mean value of the waveform (DC component) centered. When comparing the two states with center DC, we discovered preictal frequencies more commonly exhibited values near the mean and interictal more often showing values several deviations from the mean. This led to finding a significant difference across electrodes in the amount of peaks above half max (relative to one electrode’s time series) between preictal and interictal EEG. This discovery drove us to use the number of peaks as a feature that would be used in our top predictor.

The final version of our predictive model produced an area under the ROC curve (AUC) score of 58.424%, ranking us 290th out of 504 teams in the competition. This lower than anticipated performance we attribute to the deadline which was less than a month after we submitted our project proposal. In the conclusion that follows, we touch upon some of the improvements we were about to make to increase our model’s performance. Regretfully we ran out of time to implement these changes and perform more analysis to derive stronger features.

**Conclusions**

Our final classifier used measuring of peaks, variation, and correlation as features in a random forest and achieved an accuracy of .58424.  Extracting features prior to Fourier transformations proved to not be very useful since classifiers using these features had lower accuracy.  This may be due to the time-series nature of EEG or characteristics of the waves themselves.  The given EEGs also may have contained artifacts of other cerebral action other than simply interictal and preictal states.

This was also our first experience in EEG and time-series analysis, making cleaning of the data and extraction of features inorganic compared to our existing knowledge.  Upon successful utilization of calculating and understanding spectral power, our deadline was near.  Spectral power, a general type of Fourier analysis, helped to create one of our strongest features for our classifier, even though the power feature itself wasn't used, but an extracted feature we designed was created from it.  The majority of analyses from other research and the top ranking Kaggle submissions all used spectral power as well. They separated the value into multiple sections of each state by using several 1 minute partitions with overlap along with separating each frequency band.  While we did not think of overlapping our data, this type of partitioning may act as a way to expand the data the classifier is trained on, though the redundancy in each overlap may be prone to creating additional noise.  We believe that additional Fourier analysis would have led to extraction of more useful features and filters with which to clean the data, and further research into EEG preprocessing would have increased our accuracy.

As discussed in the Kaggle forums, Hastie et al. pointed out in *The elements of statistical learning* that random forests suffer if there are too few good variables relative to noisy variables. Given that EEG data is inherently noisy even after preprocessing and operations like Fourier transforms, we believe that linear regression would have been better suited to the problem. Had we had more time, regression would have been incorporated into our model if not focused on specifically while utilizing more robust features derived from Fourier analysis. We found this to be in line with one of the top performing groups, who achieved their results using a weighted average of three separate models: a Generalized Linear Model regression with Lasso or elastic net regularization (via MATLAB's lassoglm function), a Random Forest (via MATLAB's TreeBagger implementation), and a bagged set of linear Support Vector Machines (via Python's scikit-learn toolkit).

**Individual Tasks**

My role in this project included guiding the thought process for how to analyze our data. The trials we ran for each set of feature were mostly guided by me, although in general, everyone in our group had some contribution to every aspect of the project. I read the literature associated with power spectral density in an attempt to communicate its benefits to our analysis, and as a side effect of doing this, developed one of our strongest classifier’s features, the number of peaks from spectral density. In order to facilitate Sebastian’s and Ching’s main roles of being responsible for the coding part of our project, I helped search for MATLAB code that would easily implement spectral power, though luckily MATLAB already had a built-in function for it. Sebastian was responsible for more of the code development and for coming up with our team name, Predrag’s Angels, while Ching was responsible for running the code, ensuring its accuracy, and obtaining our results. Ching also developed the majority of our presentation, though each of us added to the slides that we felt most comfortable presenting. I feel that every member contributed a fair amount to the project and that the project as a whole was representative of all of our efforts combined.

**References**

American Epilepsy Society Seizure Prediction Challenge. (n.d.). Retrieved December 14, 2014, from http://www.kaggle.com/c/seizure-prediction/

Esling, Philippe, and Carlos Agon. "Time-series data mining." *ACM Computing Surveys (CSUR)* 45.1 (2012): 12.

Hastie, Trevor, et al. *The elements of statistical learning*. Vol. 2. No. 1. New York: Springer, 2009.

Howbert JJ, Patterson EE, Stead SM, Brinkmann B, Vasoli V, Crepeau D, Vite CH, Sturges B, Ruedebusch V, Mavoori J, Leyde K, Sheffield WD, Litt B, Worrell GA (2014) Forecasting seizures in dogs with naturally occurring epilepsy. PLoS One 9(1):e81920.

Lotte, F., Congedo, M., Lécuyer, A., Lamarche, F., & Arnaldi, B. (2007). A review of classification algorithms for EEG-based brain–computer interfaces. Journal of neural engineering, 4.

"NINDS Epilepsy Information Page." *National Institute of Neurological Disorders and Stroke*. 27 Aug. 2014. Web. 14 Dec. 2014.

Vlachos, M., A Practical Time Series Tutorial with MATLAB, <http://alumni.cs.ucr.edu/~mvlachos/PKDD05/PKDD05_Handout.pdf> 9th European Confernece of Practices in knowledge and Data Discovery, Portugal, 2005.