

Seizure Prediction from Intracranial EEG Recordings

Alex Fu, Spencer Gibbs, and Yuqi Liu



1 INTRODUCTION

SEIZURE forecasting systems hold promise for improving the quality of life for patients with epilepsy. One proposed forecasting method relies on continuous intracranial electroencephalography (iEEG) recording to look for telltale feature sets in EEG data that suggest imminent seizure threats. In order for EEG-based seizure forecasting systems to work effectively, computational algorithms must reliably identify periods of increased probability of seizure occurrence. If a reasonably well-performing algorithm could be implemented to identify those periods, then it is possible to warn patients only before seizures, enabling them to live a close to normal life.

Recent research has shown that EEG signals can be classified in four unique categories: interictal (between seizures), preictal (immediately prior to seizure), ictal (during seizure) and postictal (after seizure). In this project, we applied machine learning algorithms to identify potential seizure occurrence using training datasets collected from both dog and human subjects. Specifically, we aimed to differentiate between data collected during preictal and interictal states, and thereby classify periods of increased probability of seizure occurrence in both dog and human subjects with naturally occurring epilepsy.

2 DATA OVERVIEW

THE data used in this project is provided by the American Epilepsy Seizure Prediction Contest, which is sponsored by the In-

ternational Epilepsy Electrophysiology Portal (www.ieeg.org) and hosted by Kaggle.com.

The competition provides iEEG datasets for five dogs and two human subjects, each divided into preictal training examples, interictal training examples, and test examples. Each example contains data collected by 15 electrodes over a span of 10 minutes, with a sampling frequency of 400Hz; therefore, each example has a $15 \times 240,000$ matrix with voltage recordings of the 15 electrodes. Essentially, we have 3.6 million predictor variables for each data set, with 480 data sets available for just one of the subjects. The total amount of available data, when uncompressed, is 99.3 Gb. With such high dimensional data, our challenge is to choose reasonable features to make as accurate as possible predictions of periods with higher seizure probability.

3 METHODS

WE first applied feature selection techniques and chose 3 summary statistics as features, then applied logistic regression and SVM for sample classification.

Based on the results of the initial learning algorithms used, we decided to run additional algorithms on the most promising features, including dimension reduction using PCA analysis and support vector machines on the pre-processed data.

3.1 Feature Extraction

We start with visually analyzing the raw data to gain insights of the EEG signal characteristics. A few initial features we looked at including mean, variance, extreme values, period,

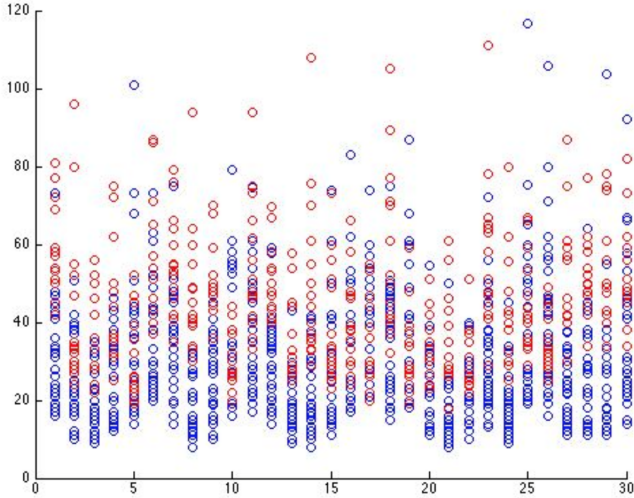


Fig. 1. Extreme value-related feature. Shows 30 positive (i.e. preictal) (blue) and 30 negative (red) training examples. The x-axis denotes the number of examples. For every example, we compute the number of EEG values that fall within the 10th to 90th percentile range of all 15 electrodes. From the diagram we can see the blue labeled positive training examples in general have lower deviations of extreme values from the mean.

and spectral energy distribution derived from Fourier transformations. For example, figure 1 shows an extreme value related feature and how value of that feature differs among positive and negative training examples.

Based on these first observations and previous research in literature, we decided to use a two-step approach for feature extraction. We first extracted first-level features from the raw data, then extracted second-level features from first-level features, and finally trained the model and based predictions using the second-level features.

3.1.1 First-level features

We implemented several first-level features, including curve length, energy, spectral entropy, and energy of wavelet packets. The result is that the signal energy best differentiates between positive and negative samples while maintaining the information of the original raw data. A detailed implementation is as following.

Let the sequence $x(n)$ be a preprocessed and fused input signal, then the instantaneous energy is given by $x(n)^2$. Considering that a sliding window is used, the energy of the signal becomes the average power over the window mathematically defined as

$$E[n] = \frac{1}{N} \sum_{i=(n-1)N+1}^{nN} x(i)^2$$

where N is the size of the sliding window expressed in number of points, range from 1 to 15; n is the discrete time index, range from 1 to 239766 in our data samples.

3.1.2 Second-level features

To generate prediction indicators we extracted 20 different second-level features from the first-level features for each sample. These included sample minimum, maximum, median, mean, variance, standard deviation, skewness, kurtosis, slope, geometric mean, trapez, sum and derivative. Some of these features did not necessarily give good results initially, but we keep them in the list for developing the objective feature vector.

3.2 Feature Selection

We started with the filters and wrappers method for feature selection, and finalized with the minimum Redundancy Maximum Relevance (mRMR) algorithm (Ding and Peng, 2005). The algorithm ranks a set of features minimizing the redundancy among the subset of features while maximizing the relevance of the features. It first runs an F-test as a relevance measure, and then a computation of the Pearson's correlation as a redundancy measure. After selecting the first feature, it iterates through the rest of the features and ranks them based on mRMR score.

$$mRMRscore = \max_{i \in \Omega S} \{F(i, s) - \left(\frac{1}{|S|} \sum_{j \in S} |c(i, j)| \right)\} \quad (1)$$

Using this method, we select 4 features: norm of variance, variance of variance, kurtosis, and skewness.

3.3 PCA Variance Analysis

Because of the success with features based on variance, we decided to do additional analysis on related features. Specifically, we ran a dimension-reduction PCA analysis on each training example available from one of the dog subjects (dog 5).

For each of the 450 interictal and 30 preictal 10-minute segments available, we first normalized the data and calculated its covariance matrix to get its eigenvector basis. The data was then projected onto each basis element, and the variance across the vector was calculated. The resulting feature vector contains the variance along each of the 15 principle components.

Next we ran various SVM algorithms using different kernels over different combinations of the feature set, including comparing the top two principle component variances, top three, and all 15. The kernels used included the standard linear kernel, and the gaussian (radial basis function) kernel. The benefit of using the rbf kernel is that it projects the data into an infinite-dimensional space without requiring much additional computation cost. Generally it is easier to get better separation of data in higher dimensions.

$$K(x, z) = e^{-\frac{\|x-z\|^2}{2\sigma^2}} \quad (2)$$

Because of the large quantity of negative (interictal) data and relatively small amount of preictal data, we added L1 regularization and weighted the preictal data regularization terms heavily compared to the interictal values. Then we adjusted the decision boundary so that the error rate of predicting (rare) preictal events was within ten percent of the error rate of predicting (common) interictal events.

In order to test training error vs testing error, we cross validated using a 70%/30% training data to test data ratio, leaving 315 interictal and 21 preictal data sets for training, and 135 interictal and 9 preictal data sets for testing.

4 RESULTS

THE results shown in table 1 are obtained from learning on 114 labeled training samples (30 positive) on 2 dog subjects. The learned

parameters are then used to predict labels of 78 additional samples for the same 2 dogs. It appears that a RBF (Gaussian) kernel SVM with cost adjustment towards avoiding false negatives gives the best results.

TABLE 1
Training and Testing Error Rates of
Second-Level Features

Model	False neg (Train)	False pos (Train)	False neg (Test)	False pos (Test)
Log. Reg.	13.3%	17.2%	16.7%	27.8%
Lin. SVM	10.0%	29.8%	11.1%	48.1%
RBF SVM	13.3%	10.7%	14.8%	16.7%
RBF SVM+CA*	6.7%	16.7%	9.3%	24.1%

4.1 PCA Variance Results

Please see table 2 and the following figures for PCA results. The data is taken from dog subject 5, using 315 interictal and 21 preictal samples for training and 135 interictal and 9 preictal samples for testing.

TABLE 2
Training and Testing Error Rates of PCA
Variance

Model	False neg (Train)	False pos (Train)	False neg (Test)	False pos (Test)
Lin. SVM kernel-2d	22.2%	17.1%	5.6%	82.7%
RBF SVM kernel-2d $\sigma = 1$	19.2%	16.0%	0.7%	82.0%
RBF SVM kernel-2d $\sigma = 10$	25.0%	21.1%	0.7%	85.7%
RBF SVM kernel-3d $\sigma = 0.8$	16.0%	13.0%	0.7%	80.4%
RBF SVM kernel-15d $\sigma = 1$	46.2%	0.6%	5.6%	10.0%

5 DISCUSSION

One interesting observation is how good variance is in differentiating preictal and interictal samples; and contrary to our initial guess

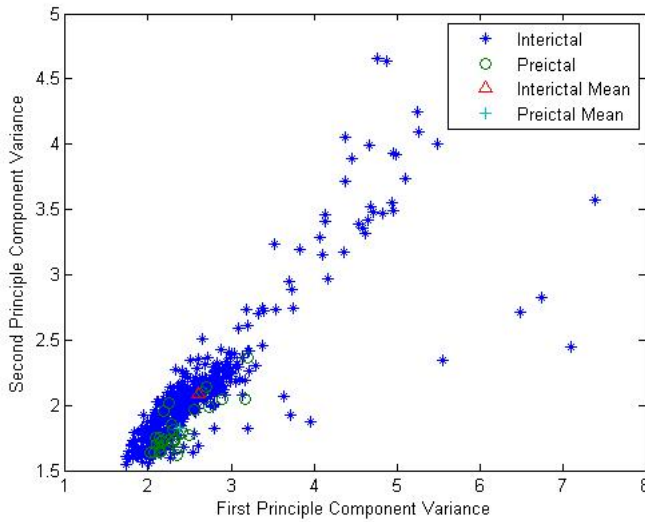


Fig. 2. Normalized Variance On First Two Principle Components

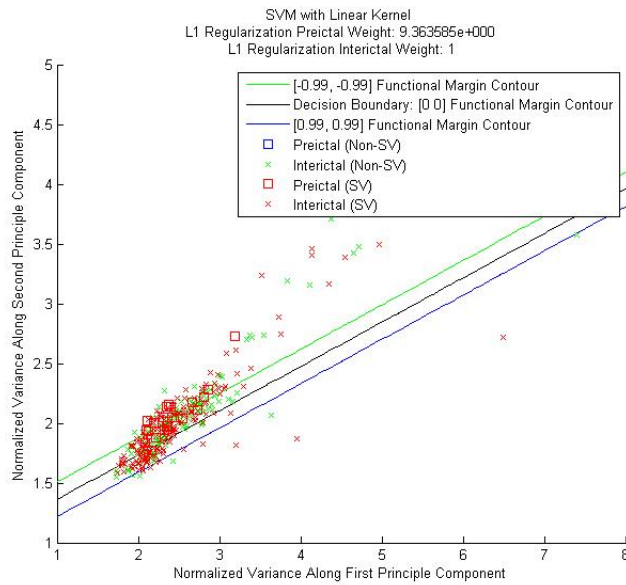


Fig. 3. Normalized Variance On First Two Principle Components, with Linear Kernel Support Boundaries

that iEEG recordings should fluctuate more right before seizures strike, variance actually tends to actually attenuate. In the implementation process, we also found that simpler features, like nth moments of the data, actually served as better features than those obtained through more complex transformations like Fourier Transform.

The learning results are promising, with only

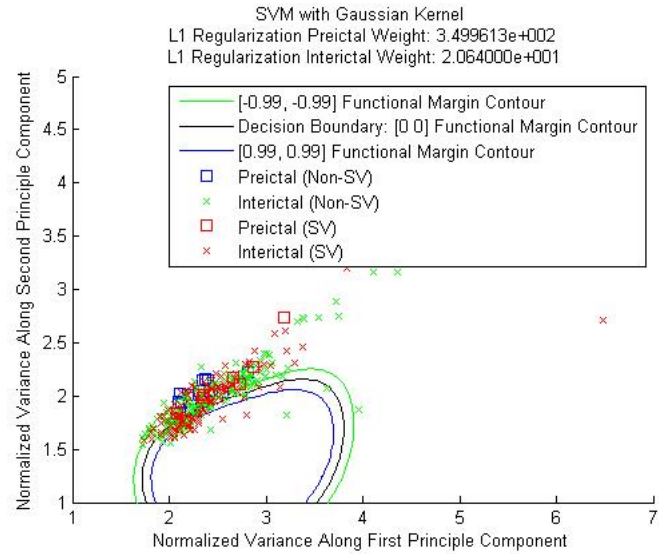


Fig. 4. Normalized Variance On First Two Principle Components, with Gaussian Kernel Support Boundaries, $\sigma = 1$

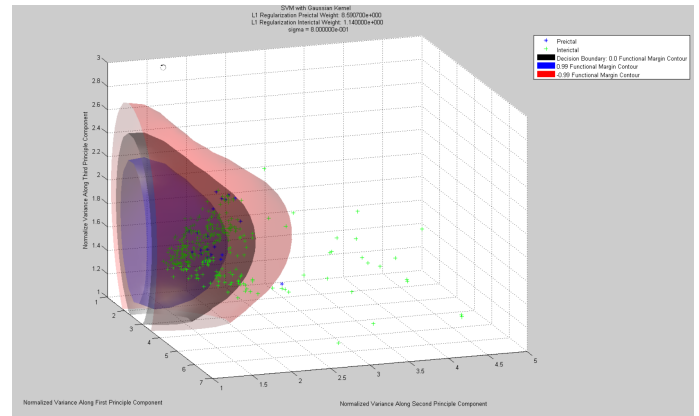


Fig. 5. Normalized Variance On First Three Principle Components with Gaussian Kernel Support Boundaries, $\sigma = 0.8$

9.3% false negative and 24.1% false positive after cost adjustment. But in practical use, those error rates are still way too high. This is partly due to the limitation on computing power - given the massive size of each sample (3.6 million recordings), it takes a significant amount of time to process more than 30 data samples at a time; another limitation is our somewhat superficial understanding of the iEEG data structure - with better understanding of the data, we could probably develop features better reflecting the deep structure of the recordings, such as ones on potential shape changes when entering

preictal states.

5.1 PCA Variance

The results of the PCA variance analysis support the theory of reduced magnitudes of variance during preictal periods. Figure 2 shows the normalized variance along the first two principle components. There does appear to be some significant structure to the data but too much mixing to provide good predictability. Noticibly most of the preictal data is in a small low magnitude clump, confirming the previous results of comparatively low variance in preictal data. Figure 3 shows the same data with a linear SVM applied. Figure 4 shows the same data with a gaussian (rbf) kernel on a reduced data set (250 interictal, 30 preictal). By applying the gaussian kernel we see that the support boundaries can flow around the data more easily. However this may make it more susceptible to overfitting.

Figure ?? shows the data comparing the first three principle component variances on 250 interictal and 30 preictal data sets. Figure 5 shows the same results with a gaussian kernel applied. Again as in the 2d case the separation is not very good, but also not without information.

Table 2 shows the results of training and testing error rates. Clearly the rates are far too high for practical use since high positive error rates would mean a patient would simply ignore warnings after a while, and high negative error rates mean we are missing dangerous seizure events. It is interesting to note that the covariance does show some degree of structure suggesting the electrode signals are not independent.

6 CONCLUSION

WE have shown a statistically significant correlation between iEEG signal variance and whether the subject is in an interictal or preictal state. As a seizure approaches it seems that iEEG signals begin to die down slightly into a muted state, before erupting during the seizure. We found that variance, variance of variance, skewedness and kurtosis

were the best features based on a mRMR selection algorithm, and using a gaussian kernel with a support vector machine generated the best prediction error rates. Further analysis of variance with PCA and gaussian kernel SVMs confirmed the reduced preictal variance result and showed some structure in the signal covariance but did not generate any significantly better prediction error rates.

7 FUTURE WORK

IN addition to trying a wider range of features, some other directions include additional dimensionality reduction to select only relevant electrodes, and analysis of the 60 min preictal states as an integral period. Another interesting direction would be to preprocess the data after running ICA analysis on each sample set. Suppose there were some dominant source generators in the brain such as some lobes or neural bundles whose activation might signal seizure imminence. Additionally suppose electrical signals add linearly so that the signal received by each electrode is a linear mapping from the signal generating neurons. Then in running ICA we could extract the source signal itself which then might be processed further with Fourier analysis or other techniques to look for something source-signal specific. Another area we started to explore but did not get results yet from is analyzing the spectral composition of the signals. We would like to look at the Fourier Transform of the data and see how energy is distributed across the power spectral density in interictal versus preictal signals.

REFERENCES

- [1] *American Epilepsy Society Seizure Prediction Challenge*. [Online]. Available: <http://www.kaggle.com>.
- [2] Bruno Direito et al., *Feature selection in high dimensional EEG features spaces for epileptic seizure prediction*, presented at the 18th IFAC World Congress, Milano, Italy, 2011.
- [3] Maryann D'Alessandro et al. *Epileptic Seizure Prediction Using Hybrid Feature Selection Over Multiple Intracranial EEG Electrode Contacts: A Report of Four Patients*, in *IEEE Transactions on Biomedical Engineering*, Vol. 50, No. 5, 2003, pp. 603-615.
- [4] Hanchuan Peng, Fuhui Long, and Chris Ding. *Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy*, in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 27, No. 8, pp.1226-1238, 2005.