Seizure On-set Prediction

Christopher Lesch, *Member, IEEE*

***Abstract*—Epilepsy is a frequently studied debilitating neurological disorder that affects approximately one percent of the population. In this paper we investigate the effectiveness of various seizure on-set prediction models. We attempt to design a model capable of differentiating between preictal and interictal seizure states from small segments of intercranial EEG. We begin with variance and correlation methods, progress to the usage of maximal cross correlation, and finally incorporate a measure of the convergence of short term Lyapunov exponent maxima into our prediction scheme. The results show that our method manages to capture some measurement of the preictal brain, however, there is still much that our model cannot accurately account for.**

***Keywords*—seizure, prediction, maximum Lyapunov exponent, SVM, Kaggle.**

I. Introduction

Epileptic seizures affect approximately one percent of the worlds population [1]. Seizures pose a significant health risk to the millions of people they affect. There is the obvious physical toll they place on the body, however, there is also a severe emotional toll. Research suggests that for many seizure patients the emotional fear of impending seizures is often one of the biggest concerns, for some even more so than the actual seizures [2]. Finally, seizure patients also suffer from dangerous situational risk. For example, operating a motor vehicle and suddenly having a seizure is far more dangerous than suddenly having a seizure while sedentary. Accurately predicting the on-set of seizures therefore has the ability to greatly improve the quality of life for patients suffering from frequent epileptic seizures. Warning the patient of an impending seizure would allow them to remove themselves from dangerous situations and emotionally prepare themselves for a possible event. In the ideal case preventative treatment, such as fast acting medicine or electrical stimulation, may even be capable of avoiding the seizure altogether.

Despite a significant amount of research no methods of seizure prediction have thus far achieved clinical applicability [1]. All current methods suffer from the trade-off between sensitivity and specificity, adjusting parameters to improve the detection rate inevitably results in the generation of more false positives. Too many false positives and a prediction scheme becomes useless, the patient will simply begin to ignore warnings of seizures, even when correct. We address the seizure prediction problem through the analysis of bivariate feature measures. Our methods succeed in distinguishing between interictal (between seizure) and preictal (preceding seizure) segments with some degree of success. However, the algorithms developed here within still face many shortcomings and leave room for the development of future work on-top of our model.

The organization of this paper is as follows. Section II contains a description of the dataset we used in our analysis, section III contains an overview of our methods, section IV contains a discussion and analysis of our results, and section V addresses our conclusions and directions of future work.

Christopher Lesch is with the Department of Electrical Engineering and Computer Science at the University of Michigan, Ann Arbor, MI, 48105 USA e-mail: chrlesc@umich.edu

Manuscript created December 12, 2014

II. Dataset

The dataset we used was provided by the American Epilepsy Society and is hosted through Kaggle. The full dataset can be downloaded from here. The dataset consists of continuous time intercranial EEG recordings from seven different subjects, five dogs and two humans. The recordings are broken down into ten minute clips labeled as either "preictal" for pre-seizure segments or "interictal" for non-seizure segments. However, sets of six sequential segments from the training data correspond to adjacent time points, resulting in one hour groupings of continuous time recordings. For the scope of this project, preictal is defined to cover the hour prior to seizure on-set with a five minute event horizon. The horizon is enforced to ensure that the resulting algorithm can predict seizures with enough warning to allow for seizure counter measures to be put in place. Additionally, the horizon ensures that any potentially missed ictal recording is not included in the pre-ictal segments. Interictal data segments were chosen randomly from the full amount of data, with the enforced restriction that interictal segments be as far away from any seizure as possible. This resulted in a separation from any seizure activity of one week in the canine recordings and of four hours in the human recordings. The test data consisted of unlabeled ten minute segments of preictal and interictal recording.

III. Methods

In this section we discuss the three different approaches that we took to the problem of predicting seizure-onset. Section III-A discusses variance and correlation based methods, section III-B introduces maximum cross correlation, and section III-C outlines the development and incorporation of the short term Lyapunov exponent maxima.

A. Variance and Correlation

For each ten minute segment of labeled training data we computed the inter-channel variance for each of the n channels separately. We then computed the cross channel correlation matrix, from which we extracted the $\frac{n(n-1)}{2}$ meaningful values. This left us with $n + \frac{n(n-1)}{2}$ values to use as features. We created a support vector machine (SVM) trained on these simple features and scaled the model parameters to fit the posterior

probability distribution of the training data. The result was a prediction model that generated scaled predictions from zero to one instead of the binary classification that SVM's normally produce. Such a scaled prediction was deemed necessary to improve our score on the aucROC performance metric the results would be judged upon.

B. Maximum Cross Correlation

Here we attempt to improve our performance by incorporating a model that more accurately accounts for the physical situation. The intercranial electrodes are spatially separated around the subjects brain. The propagation of electric currents through the cortex dictates that there will be some degree of variation in the time delay of signal acquisition of a single originator. As such, our classification algorithm should rely on the cross channel correlation that occurs at the proper intra-channel delay. This can be found by calculating the maximum cross correlation within a small time window around the actual time of signal detection. Our algorithm checks for the maximal cross correlation within ± 0.5 seconds at a resolution of 0.1 seconds. This procedure generates $\frac{n(n-1)}{2}$ features that replace the correlation features from section III-A, but the rest of the pipeline remains unchanged.

C. Short Term Lyapunov Exponent Maximum (STL_{max})

The brain is a complex non-linear dynamical system, and as such any attempt at accurately predicting its behavior must attempt to model some form of this non-linearity. In [3] the authors demonstrate that the intercranial EEG signals recorded from different cortical sites progressively converge as the brain transitions from the interictal state to the ictal state. In [4] the authors leverage this convergence by measuring the value of STL_{max} as a characteristic of the chaotic behavior of an individual EEG channel.

In our procedure we compute the value of STL_{max} by first separating our data into ten second epochs. Then, for each epoch we calculate the value of L_{max} using the following iterative algorithm described in [5]:

First we create the time-delayed embedded time series $\boldsymbol{X}$:

$$\boldsymbol{X} = [\boldsymbol{X_1 X_2} \dots \boldsymbol{X_M}]^T \tag{1}$$

$$\boldsymbol{X_i} = [x_i x_{i+J} \dots x_{i+(m-1)J}] \tag{2}$$

Where in these equations J is the reconstruction delay, or the lag, and m is the embedding dimension of the state space. Both of which can be estimated efficiently from the input data. m is estimated in accordance with Taken's theorem and satisfies $m > 2n$. J can be found where the autocorrelation drops to $1 - \frac{1}{e}$ of its initial value. Finally, $M = n - (m-1)J$. The next step involves calculating the nearest neighbor of the initial point.

$$d_j(0) = \min_{\boldsymbol{X_{j'}}} ||\boldsymbol{X_j} - \boldsymbol{X_{j'}}|| \tag{3}$$

$$|j - j'| > \text{mean period} \tag{4}$$

We impose the constraint in (4) such that the temporal separation between nearest neighbors is greater than the mean period. This constraint forces j and j' to lie on divergent trajectories. The mean period can be found as the reciprocal of the mean frequency of the epochs power spectrum. Under these conditions the maximum Lyapunov exponent can be found using a least-squares-fit to the linear section of the average over all $d_j(i)$'s:

$$y(i) = \frac{1}{\Delta t} \langle \ln d_j(i) \rangle \tag{5}$$

Once we compute L_{max} for each epoch, we have short term Lyapunov exponent maxima (STL_{max}). Figure 3 shows the values of STL_{max} over an hour of continuous time EEG. If we then measure the difference between STL_{max} values of two or more channels we obtain a measure of the convergence of chaotic behavior within the cortex [1].

In [4] after the determination of STL_{max} values for all of the observed channels the authors compute a T-score index. The T-score index is then used to predict the probability that some critical grouping of channels will converge at some future point in time. When the computed value of the T-score index drops below a threshold value an impending seizure warning is generated. In this regard, our usage of STL_{max} differs from the original algorithm. Instead of calculating a single continuous time T-score index for an entire grouping of channels we compute an averaged T-score index for each possible pairing of channels during a ten minute segment of data. Instead of using the T-score index as the single predictor of preictal or interictal we include each pairwise T-score as an additional feature in our feature vector from section III-B. We then use the same SVM strategy as section III-A to create our classifier.

IV. Results

Using just variance and cross channel correlation as features, the SVM performed surprisingly well. The method achieved 64% rocAUC on Kaggle and achieved a local prediction accuracy of 89.8% with a well calibrated threshold. All local prediction accuracy results are reported using 20 fold cross-validation with a random 20% section of the data withheld for testing purposes in each fold. As expected the addition of maximal cross correlation improved the specificity and overall performance of our classification. This improvement is visualized in Figure 1. The figure shows that maximal cross correlation rises faster and remains higher than that of the simpler variance and covariance dependent methods. These results suggest that accounting for the temporal effects of the spatial distribution of electrodes captures additional identifying information. The resulting rocAUC score on Kaggle was found to be 68% for this method and the local prediction accuracy was found to be 92.6% with a well calibrated threshold. The results on Kaggle were significantly lower than any of my cross-validation scores, this is likely due to the fact that the Kaggle challenge withheld 40% of the data for testing purposes. Furthermore, the testing data consisted of 50%

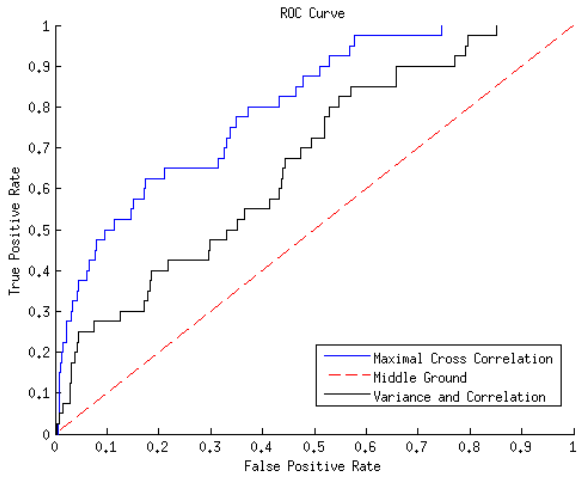


Fig. 1. A graph of the ROC curves for correlation and variance compared against that of maximal cross correlation

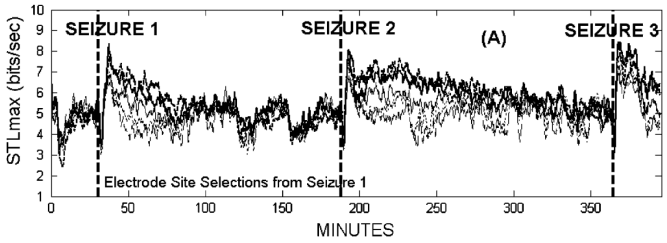
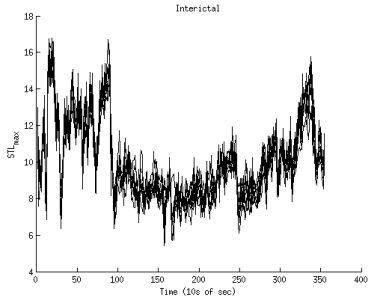


Fig. 2. Plot of the STL_{max} values as calculated by [4] for a series of three seizures

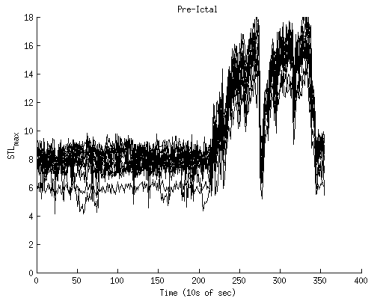
preictal segments a much higher percentage than what the training data contained.

When looking at the resulting STL_{max} graphs for an interictal and preictal segment, in Figure 3, there is an immediate disparity between the two. This suggests that there may be some value to this feature that will confer some ability to distinguish the two classes. However, neither of the graphs seems to really resemble the pattern shown in Figure 2 from the work of [4]. The values during the preictal segment are not strictly lower than those during the interictal segment. Furthermore, the values of the preictal segment do not approximate a strictly decreasing function, in fact quite the opposite. One hypothesis to explain part of this difference is that the interictal and preictal segments in our work are not taken from closely spaced time points. In fact, they must be at a minimum a week apart and are likely even much farther apart compared to [4] where they seizure events are only about 150 minutes apart. The non-stationarity of EEG over time could be responsible for some of this difference.

Despite the differences in appearance of STL_{max}, our algorithm accuracy improved to 72% aucROC on Kaggle when incorporating the additional features. Due to time constraints, and cluster allocation limits being reach, no graphics are currently available for the STL_{max} algorithm.



(a) Interictal



(b) Preictal

Fig. 3. STLmax For two different 1hr segments of EEG

V. Conclusions and Future Work

One of the biggest limitations to the accuracy of our model was the limited amount of available preictal data. There were twenty times more interictal training segments than their were preictal segments, and there were only about 24 preictal segments per subject. This limited amount of data increases the likelihood that our SVM model will treat the preictal segments as noise and that greatly impacts our prediction accuracy. Another substantial limitation was that imposed by available compute power. The dataset for the Kaggle challenge was around 60 GB worth of data. Computation of the STL_{max} algorithm on the full dataset took about ten days of compute time on an octacore compute node with 20GB of available RAM. Having such long delays before knowing the results of the slightest change is quite preventative of fast progress.

Overall, I was able to achieve an aucROC score of 72% in the Kaggle competition. The competition came to a close on November 17th and the winner of the competition achieved an aucROC score of 84%. While these scores certainly indicate performance greater than random guessing, there is still a great deal of improvement that can be had. One of the greatest difficulties in accurately predicting seizures is the lack of obvious feature choices. To date, there has not been a particular feature that has been found to be a majority principal component of the prediction accuracy. In future work I am interested in pursuing the effectiveness of convolutional network models and other unsupervised learning techniques. Unsupervised learning techniques may hold the key to discovering relationships within the EEG that cannot be simply formulated with our knowledge today.

ACKNOWLEDGMENTS

The authors would like to thank Dr. Ali Shoeb and Professor Zeeshan Syed for their help in directing the focus of our work and their invaluable adivising.

REFERENCES

[1] P. Mirowski, Y. LeCun, D. Madhavan, and R. Kuzniecky, "Comparing svm and convolutional networks for epileptic seizure prediction from intracranial eeg," in *Machine Learning for Signal Processing, 2008. MLSP 2008. IEEE Workshop on*, Oct 2008, pp. 244–249.

[2] R. S. Fisher, B. G. Vickrey, P. Gibson, B. Hermann, P. Penovich, A. Scherer, and S. Walker, "The impact of epilepsy from the patients perspective i. descriptions and subjective perceptions," *Epilepsy research*, vol. 41, no. 1, pp. 39–51, 2000.

[3] L. D. Iasemidis, P. Pardalos, J. C. Sackellares, and D.-S. Shiau, "Quadratic binary programming and dynamical system approach to determine the predictability of epileptic seizures," *Journal of Combinatorial Optimization*, vol. 5, no. 1, pp. 9–26, 2001.

[4] L. Iasemidis, D.-S. Shiau, P. Pardalos, W. Chaovalitwongse, K. Narayanan, A. Prasad, K. Tsakalis, P. Carney, and J. Sackellares, "Long-term prospective on-line real-time seizure prediction," *Clinical Neurophysiology*, vol. 116, no. 3, pp. 532 – 544, 2005.

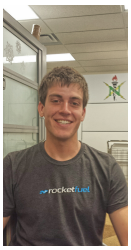
[5] M. T. Rosenstein, J. J. Collins, and C. J. D. Luca, "A practical method for calculating largest lyapunov exponents from small data sets," *Physica D: Nonlinear Phenomena*, vol. 65, no. 12, pp. 117 – 134, 1993.

[6] T. Maiwald, M. Winterhalder, R. Aschenbrenner-Scheibe, H. U. Voss, A. Schulze-Bonhage, and J. Timmer, "Comparison of three nonlinear seizure prediction methods by means of the seizure prediction characteristic," *Physica D: Nonlinear Phenomena*, vol. 194, no. 34, pp. 357 – 368, 2004.

[7] D. Wulsin, J. Gupta, R. Mani, J. Blanco, and B. Litt, "Modeling electroencephalography waveforms with semi-supervised deep belief nets: fast classification and anomaly measurement," *Journal of neural engineering*, vol. 8, no. 3, p. 036015, 2011.

[8] L. D. Iasemidis and J. C. Sackellares, "The evolution with time of the spatial distribution of the largest lyapunov exponent on the human epileptic cortex," *Measuring chaos in the human brain*, pp. 49–82, 1991.

[9] H. Kantz, "A robust method to estimate the maximal lyapunov exponent of a time series," *Physics Letters A*, vol. 185, no. 1, pp. 77 – 87, 1994.



Christopher Lesch graduated with a bachelors in Computer Engineering from the University of Michigan, Ann Arbor in 2012. He is currently pursuing a masters degree from the University of Michigan, Ann Arbor in robotics and machine learning. His primary focus is on machine learning and its applications to real time operating systems.