33rd Annual International Conference of the IEEE EMBS
Boston, Massachusetts USA, August 30 - September 3, 2011

A Low Complexity Seizure Prediction Algorithm

Michael J. Brown[1], Theoden Netoff[2], and Keshab K. Parhi[1]

University of Minnesota, Minneapolis, MN 55455, USA
Department of Electrical and Computer Engineering[1], Department of Biomedical Engineering[2]

ABSTRACT

A new low complexity seizure prediction algorithm is proposed. The algorithm achieves high sensitivity and low false positive rates in 10 out of 18 epileptic patients from the Freiburg database. Its primary achievement is two orders of magnitude computational complexity reduction. The reduced complexity makes an implantable medical device application realizable. In the subset of ten highly predictable patients average sensitivity is 96%, average specificity is 0.25 false positives per hour, and 13.5% of time is spent in false alarms. For all eighteen patients tested, the average sensitivity is 83%, the average specificity is 0.38 false positives per hour, and the amount of time spent in false alarms is 21.1%. This result may be compared with sensitivity of 97.5%, specificity of 0.27 false positives per hour, and 13% of time is spent in false alarms of prior results without complexity reduction.

Index Terms — *Epilepsy, Seizure Prediction, Implantable device, Support Vector Machine (SVM), Feature Selection*

1. INTRODUCTION

Epilepsy causes approximately 1% of the entire world's population to experience sporadic, debilitating seizures. The current preferred treatment for preventing epileptic seizures is drug therapy, which works well in a large number of patients. Of the remaining patients, many are good candidates for surgical treatment. However, approximately one quarter of epileptics do not benefit sufficiently from any treatments. An implantable medical device for predicting seizures will drastically improve these patients' quality of life, since an alert would allow them to take precautions before a seizure occurs. An implantable prediction device will also make closed loop therapy feasible – automatically releasing fast-acting medication or other therapies to prevent the seizure entirely. In addition, continuous medication for treating epilepsy can be debilitating, and brain surgery is inherently dangerous and expensive with no guarantee of success. Therefore, viable alternative treatment would benefit all epileptics.

2. BACKGROUND

There has been significant research in the field of seizure prediction and detection. Machine learning approaches such as support vector machines (SVMs) have been shown to be very promising for classification of pre-ictal (before seizure) and inter-ictal (between seizure) electro-encephalogram (EEG) data [1]. However, in the past these methods have been far too computationally complex for implementation in an implantable medical device. Medical device industry professionals have used ~50 μW as a guideline for power consumption of a realizable seizure detection or prediction implantable device [2]. To reach this goal, the machine learning algorithm must use very few features. Other factors such as filtering, SVM kernel type, and sampling rates also have a dramatic effect on power consumption.

All of these factors must be addressed, while still retaining the highest prediction accuracy possible, for a successful seizure prediction algorithm. To accomplish this, the proposed algorithm makes use of extensive off-line training to choose a limited number of optimal features. The algorithm begins by combining features to produce new, more efficient, and easier to classify features. From this optimized feature set, the algorithm then selects only a few features to allow correct classification of data. An exhaustive search of feature combinations would require weeks on a supercomputer. Instead, a greedy approach has been designed. Finally, it is very important for the algorithm to choose positive pre-ictal samples and negative inter-ictal samples carefully. If a large portion of positive training samples look identical to negative training samples, the model produced by the SVM is certain to be poor. Therefore, the selection of the pre-ictal positive data is tuned very carefully within the algorithm.

3. DIVERGENCE MEASUREMENT

The basis for much of the complexity reduction achieved in the proposed algorithm is through the use of the divergence measurement method proposed by Henze and Penrose [3]. The best possible features used for classification have high mean difference between positive and negative samples and low variance. Accurately measuring mean or variance can be difficult and imprecise. Also, a formula would be necessary to determine tradeoffs between mean difference improvement and variance degradation, or vice versa.

Instead, the Henze-Penrose divergence (HPD) measurement can be used to determine if a feature will allow easier classification of positive and negative samples. Single dimensional HPD begins with a list of $N/2$ positive and $N/2$ negative samples sorted by amplitude. In easily classifiable data, all of the positive samples would be at one end of the list, and all the negative samples would be at the other. In

data which is difficult to classify, the two groups are heavily mixed. HPD counts the number of adjacent points ***R*** of differing classification, and applies the following formula to produce a divergence measurement.

$$HPD = 1 - R / N$$

An HPD value of ~1 represents excellent divergence, and an HPD value of ~0.5 means the data is effectively random. Multi-dimensional HPD uses a minimum spanning tree (MST) instead of a sorted list, where ***R*** in the above formula is the number of edges connecting points of differing classification.

3. TEST METHODS

The proposed prediction algorithm has been tested on the Freiburg epilepsy database [4]. This database has been used extensively for the development and testing of new seizure prediction and detection algorithms. It provides 6 channels of intracranial EEG recordings sampled at 256 Hz from 21 epileptic patients who were candidates for surgical intervention. More than 24 hours of data is available for each patient. Since the database is publicly available, researchers can compare and contrast different algorithms on the same data set. To accurately measure the effectiveness of the seizure predictor, leave-one-out cross-validation is used. For each patient, one seizure and a large block of inter-ictal data are completely isolated from the off-line training methods to be used as a test set. Once a predictor has been designed based upon the training seizures, it is then applied to the test set in order to measure prediction accuracy and false positive rate. Each seizure is used once as a test set; no data is excluded from testing. Many prior works exclude near seizure inter-ictal data based upon suggestions in a paper by F. Mormann [5]. However, false positive rates excluding this near seizure inter-ictal data are also provided to facilitate comparison to past works.

4. OFF-LINE TRAINING METHODS

The final predictor is patient specific, so the entire training process shown in Fig. 1 is repeated for each patient. Arbitrarily chosen model parameters are avoided. Instead, the algorithm is designed in such a way that these parameters (such as pre-ictal length) are deduced.

4.1. Preprocessing – Filtering

Before extracting the power spectral density of the EEG channels, the signals are first pre-whitened to remove the predictable auto-regressive components of the EEG. The 20-tap whitening filter for each channel is trained using inter-ictal data from that channel only. This makes differences between inter-ictal and pre-ictal EEGs easier to identify by the divergence measurements performed in off-line processing.

4.2 Feature Extraction

Power spectral density features are extracted from each of the six channels using Welch's method [6]. This produces a total of 720 features. Twenty second windows with half overlap are used. This effectively generates a sample every ten seconds. The twenty second window and 1Hz resolution were chosen through empirical testing.

4.3 Feature Consolidation

Next, single dimensional Henze-Penrose divergence (HPD) measurement is used to test linear combinations for beneficial feature consolidation. The following algorithm defines the feature consolidation process:

For each feature X_i
  $Y_i = X_i + X_{i+1}$
  If $HPD(Y_i)$ >= both $HPD(X_i)$ and $HPD(X_{i+1})$
    Replace X_i with Y_i
    Eliminate feature X_{i+1}
Repeat until no Y_i results in HPD improvement

Frequency bands are only allowed to combine with adjacent frequency bands. For the implantable device the consolidations will be incorporated into the power spectrum estimation at no cost. This method typically reduces the number of features to less than half of the original number.

4.4 Feature Selection

After feature consolidation has maximized the information contained in each feature, feature selection is used to drastically reduce the number of features by selecting only those with excellent separation between positive and negative samples. Selecting several features with good separation is insufficient. Each added feature must improve the classification of additional points that weren't already classified correctly. Multi-dimensional Henze-Penrose divergence is used to determine the four features which best complement each other using the following method:

F_1 = feature with max($HPD(X_i)$)
F_2 = feature with max($HPD(F_1, X_i)$)
F_3 = feature with max($HPD(F_1, F_2, X_i)$)
F_4 = feature with max($HPD(F_1, F_2, F_3, X_i)$)

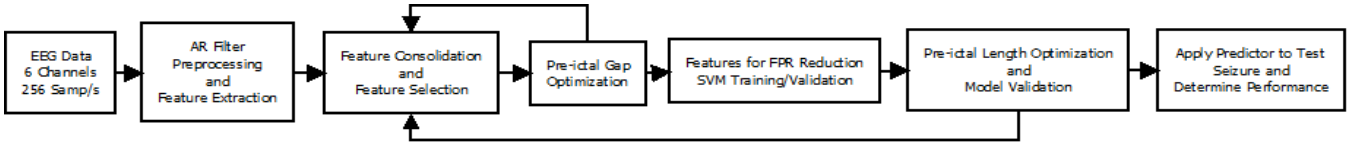

Figure 1 – Off-line Training Algorithm Flowchart

An exhaustive search of all possible feature combinations is intractable. Instead, the feature with the highest single-dimensional divergence is chosen first. The second feature is selected to maximize divergence when used in combination with the first. The third and fourth features are selected using the same greedy approach. Computation time for feature consolidation and selection on a single patient is only a few minutes on a standard PC and less than a minute on a high performance machine. This allows consolidation and selection to be performed hundreds of times during the optimization process.

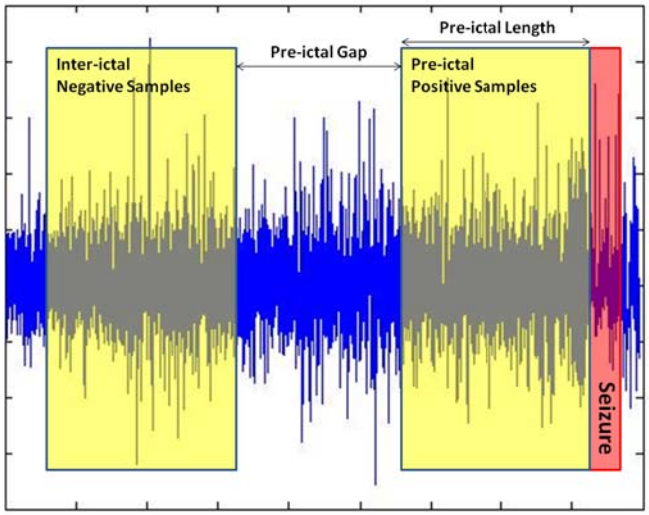

Figure 2 – EEG Illustration Defining Pre-ictal Length and Gap

4.5 Pre-ictal Gap Optimization

All machine learning algorithms require training on positive pre-ictal samples and negative inter-ictal samples. In prior work pre-ictal length, from which all positive samples are extracted, was arbitrarily assigned. Negative samples are safely chosen from a long period without seizures. However, in order to *predict* seizures, it is necessary to find a change in EEG patterns immediately preceding the seizure. By performing HPD on two groups of EEG samples, the maximum divergence achieved by feature consolidation and selection indicates whether a change has occurred between the recordings. The algorithm tests divergence for inter-ictal samples chosen from farther away from the pre-ictal samples. This finds negative samples preceding the assumed change in EEG. We define the distance between negative inter-ictal samples and positive pre-ictal samples as the pre-ictal gap, as illustrated in Fig. 2. To determine the optimal pre-ictal gap a grid search followed by a pattern search is performed on training seizures. The grid search changes the gap for all seizures simultaneously to find maximum divergence. The pattern search then attempts to change the gap for each specific seizure to further improve divergence.

4.6 Features for False Positive Reduction and Support Vector Machine Parameter Optimization

To reduce the false positive rate in long inter-ictal periods, four additional features are added using several small sections of the long inter-ictal periods provided by the Freiburg data set. A linear kernel for the SVM is used since it offers much lower computational complexity than radial-basis function and polynomial kernels. The SVM misclassification cost parameter is set through out-of-sample cross-validation on 20% of training data using a gradient search based on classification accuracy. Since the number of available inter-ictal training samples is an order of magnitude higher than the number of pre-ictal training samples, the SVM cost ratio parameter is used to balance the misclassification error. For example, if 100 pre-ictal and 1000 inter-ictal samples are available for training, then pre-ictal samples are given 10 times the misclassification cost so that sensitivity and specificity are given equal importance. This could be adjusted to reduce false positives at the expense of accuracy, or to increase accuracy at the expense of a higher false positive rate.

4.7 Pre-ictal Length and Predictor Validation

After the SVM has been trained, it is validated a second time with the data used to train it. This is a quality assurance check ensuring the predictor is able to predict the seizures it was trained upon. In previous seizure prediction efforts, pre-ictal length was defined arbitrarily to be anywhere from 1 minute to 2 hours preceding the seizure. In the proposed algorithm, the entire training process is repeated for pre-ictal lengths varying from 6 minutes to 30 minutes in 2 minute intervals. The model with the best prediction accuracy on training data is chosen as the final model applied to the test seizure. If models trained with different pre-ictal lengths have the same accuracy, then the model with the lowest false positive rate on training data is used.

4.8 Post-processing

The support vector machine output must be filtered to reduce sporadic false positives. The SVM output for this algorithm is filtered by a 9-tap median filter to eliminate outliers caused by artifacts and an 8-tap mean filter for further smoothing. The hardware cost for this type of filtering is quite small.

5. RESULTS AND DISCUSSION

Figure 3 is the block diagram for the on-line prediction algorithm. In the best 10 patients shown in table 1, average sensitivity is 96%, average specificity is 0.25 false positives per hour, with 13.5% of time spent in false alarm state. Including all patients, average sensitivity is 83%, average

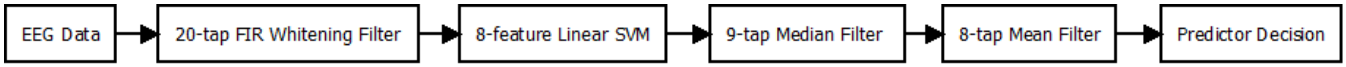

Figure 3 – On-line Predictor Block Diagram

Table 1 - Prediction Results Table

1. Patients 2, 8, and 13 are excluded due to insufficient seizure and/or inter-ictal data.

Patient	Sensitivity Predicted/Available : %	Advance Prediction Time Shortest/Longest : Avg.	Overall FP per hour / %FP	P-value	Without Seizure FP per hour / FP%	Near Seizure FP per hour / FP%
1	4/4 : 100%	7 min. / 127 min. : 64 min.	0.095 / 6%	0	0.078 / 2.2%	0.3 / 24.75%
3	5/5 : 100%	27 min. / 83 min. : 65 min.	0.25 / 7.8%	0	0.256 / 8%	0.11 / 4.6%
4	5/5 : 100%	18 min. / 101 min. : 68 min.	0.04 / 3.44%	0	0.026 / 2%	0.17 / 16%
5	4/5 : 80%	14 min. / 95 min. : 52 min.	0.276 / 20.4%	0.0004	0.282 / 21%	0.2 / 8.4%
6	3/3 : 100%	7 min. / 64 min. : 39 min.	0.34 / 44%	0	0.31 / 44.3%	0.59 / 37.3%
7	3/3 : 100%	12 min. / 97 min. : 41 min.	0.053 / 2.4%	0	0.02 / 0.67%	0.29 / 26%
9	5/5 : 100%	55 min. / 137 min. : 87 min.	0.242 / 6.6%	0	0.264 / 6.1%	0.08 / 9.4%
10	5/5 : 100%	14 min. / 67 min. : 43 min.	0.508 / 23.4%	0	0.479 / 23%	0.78 / 29%
11	2/4 : 50%	63 min. / 140 min. : 102 min.	0.502 / 21.5%	0.033	0.485 / 0.76%	1.63 / 19.5%
12	4/4 : 100%	5 min. / 55 min. : 19 min.	0.08 / 4.25%	0	0.065 / 0.1%	0.53 / 16.75%
14	4/4 : 100%	36 min. / 81 min. : 64 min.	0.8 / 43.5%	0	0.848 / 46.1%	0.21 / 4.75%
15	4/4 : 100%	32 min. / 134 min. : 80 min.	0.61 / 35.7%	0	0.645 / 38%	0.16 / 8.75%
16	2/5 : 100%	12 min. / 84 min. : 59 min.	0.606 / 19.6%	0.055	0.662 / 21.8%	0.06 / 0.94%
17	4/5 : 80%	48 min. / 103 min. : 76 min.	0.347 / 23.7%	0.0007	0.322 / 17.3%	0.64 / 18.7%
18	2/5 : 40%	7 min. / 8 min. : 8 min.	0.273 / 10.5%	0.0098	0.205 / 7%	1.07 / 36.4%
19	3/4 : 75%	24 min. / 45 min. : 38 min.	0.68 / 39%	0.0231	0.68 / 40.4%	0.59 / 22.6%
20	3/5 : 60%	10 min. / 18 min. : 13 min.	0.685 / 42.5%	0.1033	0.717 / 42.4%	0.4 / 43%
21	3/5 : 60%	23 min. / 40 min. : 30 min.	0.577 / 26.9%	0.0205	0.523 / 25.4%	1.1 / 43.2%
Average	65/80 : 83%	53 min.	0.386 / 21.1%	0.01366	0.385 / 19.2%	0.495 / 24%

specificity is 0.386 false positives per hour, with 21% of time spent in false alarm state. Prior work without complexity reduction achieved average sensitivity 97.5%, average specificity is 0.27 false positives per hour, and 13.0% of time is spent in false alarm state [7]. Our algorithm selects eight features, and in more than half of patients it selects high frequency bands from 100-128Hz. This supports conclusions reported in some recent epilepsy research publications [8]. Advance prediction time ranges widely from patient to patient, providing further evidence that pre-ictal length should be tailored to individual patients. Many of the advance prediction times are comparable to those found by other researchers [9].

7. CONCLUSION

The proposed algorithm produces a predictor capable of good accuracy and low false positive rates on ten of the eighteen patients from the Freiburg database. This algorithm produces a patient-specific predictor of low enough complexity for implementation in an implantable seizure prediction device. Future work will focus on improving performance of post-processing and reducing hardware cost of pre-processing. A 20-tap whitening filter running at sampling frequency on every input channel will be very hardware and power intensive. The 9-tap median and 8-tap mean filtering of the SVM output are inexpensive, but the complexity cost for a higher performance filter that significantly reduces false positives may be justified.

8. ACKNOWLEDGMENT

This research was supported in part by a grant from the Institute of Engineering in Medicine (IEM) at the University of Minnesota.

9. REFERENCES

1. T Netoff, Y Park, and K Parhi. “Seizure prediction using cost-sensitive support vector machine,” *31st IEEE EMBS*. Minneapolis, MN. pp. 3322-3325. Sept 2009.
2. A. Shoeb, et al., “A micro support vector machine based seizure detection architecture for embedded medical devices,” *31st IEEE EMBS*: Minneapolis, MN. pp. 4202-4205. Sept 2009.
3. N. Henze and M. Penrose, “On the Multivariate Runs Test,” *Annals of Statistics*, 1999.
4. *Freiburg EEG database*. Available from: https://epilepsy.uni-freiburg.de/freiburg-seizure-prediction-project/eeg-database.
5. F. Mormann, et al., “Seizure prediction: the long and winding road,” *Brain*. **130**(2): p. 314. 2007.
6. P. Stoica and R. Moses, *Spectral Analysis of Signals*. Prentice Hall, 2005.
7. Y. Park, L. Lao, T. Netoff, and K. Parhi, “Seizure Prediction with Spectral Power of EEG Using Cost-Sensitive Support Vector Machines,” *Epilepsia*, To Appear.
8. G. Worrell, et al., “High-frequency oscillations and seizure generation in neocortical epilepsy,” *Brain*. **127**(7): pp. 1496. May 2004.
9. L. Chisci, et al., “Real-time Epileptic Seizure Prediction using AR Models and Support Vector Machines,” *IEEE Transactions on Biomedical Engineering*, pp. 1124-1132 Feb 2010.