Proceedings of the 28th IEEE
EMBS Annual International Conference
New York City, USA, Aug 30-Sept 3, 2006

Reducing the Number of Channels for an Ambulatory Patient-Specific EEG-based Epileptic Seizure Detector by Applying Recursive Feature Elimination

Elena L. Glassman, *Student Member, IEEE,* John V. Guttag
MIT CSAIL and EECS

***Abstract*—We are building an ambulatory version of a patient-specific epileptic seizure detector based on scalp EEG signals. Since patients have to wear the electrodes all the time, it is desirable to use the minimum number of electrodes needed to achieve good performance. In this paper, we describe a method that uses Recursive Feature Elimination (RFE) to design detectors that use small numbers of electrodes. We also present results that indicate that the appropriate number of electrodes varies across patients. It is frequently the case that a surprisingly small number of electrodes, sometimes as few as two, suffices to construct a detector with expected performance comparable to that of detectors that use a full twenty-one-channel montage.**

I. Introduction

Qu and Gotman [1] and Shoeb et al. [2] describe methods for constructing patient-specific algorithms that detect the onset of epileptic seizures using scalp EEG signals. Studies reported in [3] and [4] indicate that Shoeb's method, which uses machine learning and support vector machines [5], produces patient-specific detectors with excellent sensitivity, specificity, and latency for most patients—when used with full 21-channel EEG montages.

In [4], Schachter et al. describe a prototype ambulatory system that detects seizure onset using the method described by Shoeb et al. in [3]. The system is rather cumbersome. It includes a cap with 21 EEG channels, the hardware needed to capture and process those channels, and the battery needed to power the hardware. If the number of channels were significantly reduced the system could be made considerably less cumbersome.

As we show later in this paper, the number of channels varies widely across patients. For some patients a one-channel detector works as well as a 21-channel detector, but for others we looked at 15 channels are needed to attain performance comparable to that of a 21-channel detector.

A brute force approach to determining the number of channels needed is outlined in Figure 1. The underlying idea is to estimate the expected performance of detectors using varying numbers of channels, and then choose the smallest number of channels for which the expected performance is comparable to the expected performance of a 21-channel detector. Unfortunately, this approach is computationally intractable since it would involve training and testing on approximately 2^{21} different combinations of channels.

This work was supported by the National Library of Medicine, Acer Inc., Delta Electronics Inc., HP Corp., NTT Inc., Nokia Research Center, and Philips Research under the MIT Project Oxygen Partnership, and CIMIT, the Center for the Integration of Medicine and Innovative Technology.

In this paper, we describe a method that uses recursive feature elimination (RFE) [6] to design support vector machine (SVM) based detectors that use small numbers of electrodes. We also present results that indicate that it is frequently the case that a surprisingly small number of electrodes (often as few as two) suffices to construct a detector that performs as well as detectors that use a full twenty-one channel montage.

1) Estimate the expected baseline performance of a 21-channel detector

2) For n between 1 and 20, estimate the expected performance of detectors built using all subsets of n channels.

3) Choose the smallest n such that the expected performance of the n-channel detector is at least as good as the expected performance of the 21-channel detector

Figure 1: Brute force method for choosing subset of channels

II. System Overview

A. The EEG Data

The electroencephalogram (EEG) is an electrical record of brain activity that is collected using an array of electrodes uniformly distributed on a subject's scalp. A channel is defined as the difference between a pair of (typically adjacent) electrodes.

B. Epileptic Seizure Detector

The EEG-based, patient-specific epileptic seizure detector is based on the detector described in [2]. It employs wavelet analysis to extract features from 21 channels of scalp EEG and an SVM built using a radial basis function (RBF) kernel. Since an instance of the detector is patient-specific, that instance is trained for a particular patient by training on data from that patient only.

C. Algorithm for Reducing the Number of EEG Channels

Roughly speaking, we replace step 2 in Figure 1 by:

2) For n between 1 and 20, use recursive feature elimination to choose the n best channels. Estimate the performance of the detectors built using those channels.

The process of choosing n is described in more detail in the next section.

The performance of a detector is evaluated in terms of its false positive (FP) rate, false negative (FN) rate, and latency. A false positive occurs when the detector declares a seizure outside of the window of time that the professional, who labeled the dataset, identified as a seizure. A false negative occurs when the detector fails to declare a seizure at any time during the window of time that the professional identified as a seizure. The latency is the number of seconds between when the labeling professional marked a seizure onset and when the detector declared a seizure.

Since there are 2^{21} different possible subsets that can be made from the original 21 channels, it is not practical to perform an exhaustive search to find the subsets with which the detector obtains the best performance for a particular patient. Instead, we use RFE, a greedy algorithm, to choose a subset of each size that seems to provide the detector with sufficient information to perform well on future inputs. We use the version of RFE for non-linear SVM kernels, since the RBF kernel is non-linear.

The key to RFE is the way it uses the SVM machinery to rank the contributions of each channel in the set of channels being used for detection. Once RFE ranks the current set of n channels, the channel ranked as least important in the set is removed. This produces a set of n-1 channels. This rank-and-remove process is repeated on the set of n-1 channels, which produces a set of n-2 channels. The process continues until one channel remains. When RFE is applied to a set of n channels, it produces a total of n-1 subsets. Though there is no guarantee that each subset found is indeed the best subset of that size, there are good reasons to believe that RFE finds one of the better subsets.

III. Determining the Appropriate Subset Size

Leave-one-out cross validation is frequently used to estimate the generalization performance of classifiers built using machine learning.

For the 10 patients analyzed in this paper, the dataset included, on average, 5.5 seizures per patient. Each seizure is embedded in a larger EEG stream that contains non-seizure EEG. For each patient, we use leave-one-seizure-file-out cross validation to evaluate the performance of detectors built using various numbers of channels.

This process is described in Figure 2. The basic idea is to find the smallest number of channels n, such that the average cross validation performance of detectors built using n channels is at least as good (with respect to each of the false negative rate, the false positive rate, and the latency) as the average cross validation for the 21-channel detector. In Figure 2, the function *buildDetector(C, S)* builds an SVM detector using the channels in C and training on the files in S. The function *RFE(n, S)* uses recursive feature elimination to find the n best channels when training on the files in S. The function *update(avePer, d, s)* calculates the performance of the detector d when used on the file s, and updates the measure of average performance *avePer*.

Note that this procedure finds the number of channels to be used, but does not directly compute which channels to use. It does find a set of channels for each <*size, cross validation set*> pair, but RFE may find different channels for different cross validation sets.

Once the number of channels has been determined, we run RFE using all of the files in S to choose a set of channels. We then train a detector on those channels and all of the files in S to get our ambulatory detector. The performance of the resulting detector is estimated by the average FP rate, FN rate and latency measured for all of the n-channel detectors built during leave-one-seizure-file-out cross-validation.

```
"Find average performance for full montage"
init(aveAllPerf)
for s = each seizure in set of seizures S
  C = all 21 channels
  d = buildDetector(C, S – {s})
  update(aveAllPerf, d, s)
end
"Find min. number of channels with average
performance at least as good as aveAllPerf"
numNeeded = 21
for n = 20 to 1
  init(aveSubsetPerf)
  for s = each seizure in set of seizures S
     S' = S – {s}
     C = RFE(n, S') "Find n best channels"
     d = buildDetector(C, S')
     update(aveSubsetPerf, d, s)
   if aveSubsetPerf >= aveAllPerf
     numNeeded = n
return numNeeded
```

Figure 2: Using cross validation to find number of channels for a patient.

IV. Results

For each of the 10 patients in the dataset, Table 1 shows the number of channels chosen for each patient by the procedure outlined above. It also lists the number of files and the number of seizures for each patient.

Since the data in a channel is the difference in scalp potential between two electrodes, the number of channels listed in Table 1 is not the same as the number of electrodes that would be necessary for the ambulatory detection system. Since adjacent channels may share an electrode, if the number of channels listed is q, the number of electrodes may be as small as $q+1$.

Patient	1	2	3	4	5	6	7	8	9	10
Subset Size	12	1	13	19	3	1	1	3	15	3
Files	10	5	4	5	8	5	7	5	5	8
Seizures	9	5	4	5	5	5	7	3	5	7

Table 1: Number of channels chosen for each patient, and number of files and seizures used for training.

Table 2 shows, for each patient, the expected false negative rate, false positive rate, and latency derived for the n-channel detector. Recall that by construction, the n-channel detector performs at least as well as the 21-channel detector. For a few patients, the reduced channel detector performs slightly better in some respects than the 21-channel detector, but the differences are not statistically significant.

Patient	n	FN	FP	Latency
1	12	0.11	0.1	9.75
2	1	0	0	10
3	13	0	2.75	1.25
4	19	0	0.4	11.4
5	3	0	0.25	6
6	1	0	0	12.2
7	1	0	0.1429	14
8	3	0	0	12.7
9	15	0	0	7.4
10	3	0	1.125	10.7

Table 2: Expected false negative rate, false positive rate, and latency for *n*-channel detectors.

Recall that different subsets of the data may lead to different choices of channels for the same patient. Figure 3 shows how often each channel was chosen for each patient. For example, we had four seizure files for patient 2. For three of the leave-one-out tests RFE chose channel 1 (electrodes FP1 and F7), and for one it chose channel 21 (electrodes F7 and T7). Figure 4 contains part of a seizure drawn from the EEG collected for patient 2. The seizure has an abrupt and unmistakable onset during which channels 1 and 21 (the top two channels in the figure) behave similarly. Our algorithm for building an n-channel detector for this patient chose a single channel, channel 1.

In general, for those patients requiring only a small number of channels, the channels cluster in the same region of the head. In contrast, for those patients for whom many channels are needed, e.g., patient 9, the channels are typically widely dispersed. Figure 5 contains part of a seizure drawn from the EEG collected for patient 9. Even though fewer channels seem to be involved than for patient

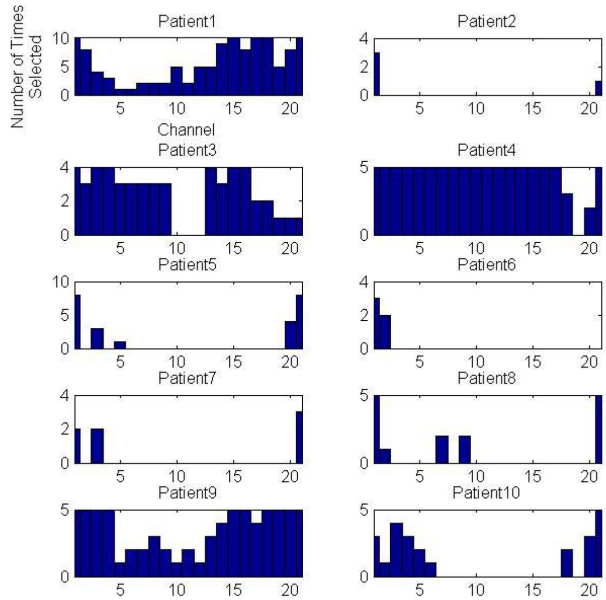

Figure 3: Histograms of channels chosen during selection process.

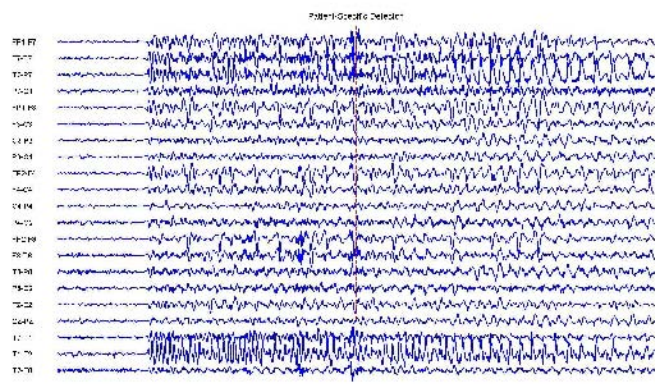

Figure 4: EEG of a seizure for Patient 2.

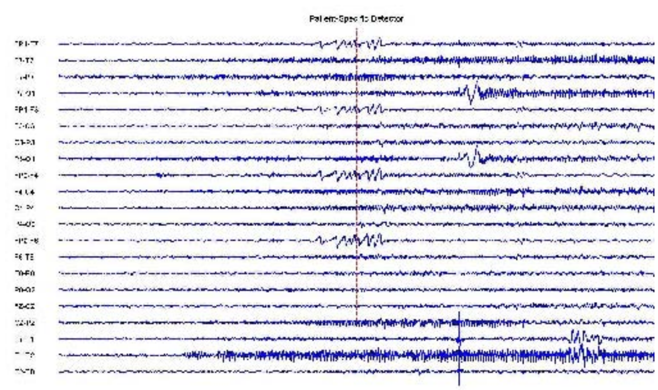

Figure 5: EEG of a seizure for Patient 9. Vertical line indicates the 10-second mark within the window.

2, it takes more channels to reliably detect the seizure. This is because at times subsets of channels show seizure-like activity that does not evolve into a clinical seizure, as seen in Figure 6. Our algorithm for building an n-channel detector for this patient chose 15 channels involving 18 of the 21 electrodes. The only channels not used were 5, 6, 7, 10, 11, and 12. This is consistent with what the histogram in Figure 3 would lead one to expect.

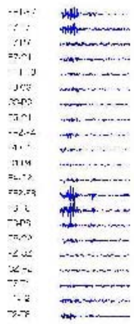

Figure 6: Inter-ictal bursts for Patient 9.

V. Discussion

One should be careful not to draw definitive conclusions from a study of a small amount of EEG for ten patients. However, Table 1 suggests that for some patients it should be possible to perform epileptic seizure onset detection with a small number of channels. For six of ten patients, as few as three channels are probably sufficient for detecting seizures of the types observed during our tests.

The number of channels needed for a patient depends, not surprisingly, on the characteristics of the patient's seizures. Some patients' seizures are focal in origin and consistently originate in a single small region of the brain. For those patients a small number of electrodes placed over the focus can be sufficient. For generalized seizures, in which seizure activity is present on most if not all electrodes, any electrode may be as good as any other, and again a small number of electrodes may be sufficient.

Some patients have different kinds of seizures with different origins. These patients will require more electrodes. Additionally, some channels may be naturally noisier than others or may produce confounding data (e.g., inter-ictal bursts that don't lead to clinical seizures). In such cases, more channels may be necessary to discriminate seizures from other activities.

Acknowledgments

The authors gratefully acknowledge many helpful discussions with Dorothy Curtis, Martin Glassman, Daniel Leeds, Eugene Shih, Ali Shoeb, and Zeeshan Syed.

References

[1] H. Qu and J. Gotman, "A patient-specific algorithm for the detection of seizure onset in long-term EEG monitoring: possible use as a warning device," IEEE Trans Biomed Eng. 1997 Feb;44(2).

[2] A. Shoeb, S. Schachter, B. Bourgeois, S. T. Treves, and J. Guttag "Automated Seizure Onset Detection as a New Method of On-Demand VNS Stimulation: Initial Technical Development and Findings," American Epilepsy Society 59th Annual, Meeting, December 2005.

[3] A. Shoeb, H. Edwards, J. Connolly, B. Bourgeois, S. T. Treves, and J. Guttag, "Patient-Specific Seizure Onset Detection," *Epilepsy and Behavior*, August 2004.

[4] S. Schachter, A. Shoeb, B. Bourgeois, S. T. Treves, and J. Guttag, "Feasibility of Detecting Seizure Onsets in Ambulatory Patients: Initial Implementation," American Epilepsy Society 59th Annual, Meeting, December 2005.

[5] B. E. Boser, I. M. Guyon, and V. N. Vapnik. *A training algorithm for optimal margin classifiers*. In D. Haussler, editor, 5th Annual ACM Workshop on COLT, pages 144-152, Pittsburgh, PA, 1992. ACM Press.

[6] I. Guyon, J. Weston, S. Barnhill, and V. Vapnik, Gene selection for cancer classification using support vector machines. *Machine Learning* 46:389–422, 2002.