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

Optimal Feature Selection for Seizure Detection: A Subspace Based Approach

Tolga E. Özkurt, *Student Member, IEEE*, Mingui Sun, *Senior Member, IEEE*, Tayfun Akgül, *Senior Member, IEEE* and Robert J. Sclabassi, *Senior Member, IEEE*

***Abstract*—An epileptic seizure detector's performance definitely depends on features extraction and selection. In this study, we present the short-time average magnitude difference function (sAMDF) as a computationally efficient feature to distinguish seizures from EEG and it is compared with the frequently used curve length. We also suggest using a subspace based approach for feature selection that exploits divergence measure as the dissimilarity criterion. In this approach, basically features are linearly transformed into another reduced space for optimality while decreasing the computational burden. Seizure discrimination performances of transformed features and original features are compared. The obtained results demonstrate that the feature selection with a divergence-based subspace approach is quite useful to discriminate the seizure parts of the signal from the nonseizure ones.**

I. INTRODUCTION

FEATURE extraction and selection steps are significant for an epileptic seizure detector's performance. For epilepsy detection, the main goal is to extract features that can differentiate the seizure stage from the nonseizure one. Since small delay time for the seizure detection is also required; the features' computational efficiency is a crucial factor.

The principal of feature selection is to transform the extracted feature space into another space; such that discrimination between the classes is increased while the dimension is reduced. In [1], the authors use an optimal feature selection method called mutual information selection and a neural network classifier. Depending on this method, 9 wavelet transform based features are selected out of 26 features. A small decrease in the detection rate performance is observed, when they use the selected features instead of all features. In [2], D'Alessandro et al. use a genetic algorithm based feature and channel selection, where the decision criterion depends on Fisher's discrimination ratio and their classifier's performance. In all of these methods, an attempt is made to select an optimal subset of features from a full feature set. In this study, we suggest a feature selection method that changes the full feature space to a reduced optimal space by a linear transformation. This linear transformation is based on a divergence measure that is mostly used in speech signal processing. We also suggest a generalized version of curve length called short-time average magnitude difference function (sAMDF), which may be more useful for detection of epileptic seizures

Manuscript received April 16, 2006. This research is supported in part by NIH grants EB002309 and NS/MH38494 and Computational Diagnostics, Inc.

T. E. Özkurt, M. Sun and R. J. Sclabassi are with the Laboratory for Computational Neuroscience, Department of Neurological Surgery, University of Pittsburgh, Pittsburgh, PA, USA (e-mail: tolga@neuronet.pitt.edu).

T. Akgül is with the Department of Electronics and Communication Engineering, Istanbul Technical University, Istanbul, Turkey

II. METHODOLOGY

A. Data Properties

94 channels EEG data from a five-year old male patient were recorded at the University of Pittsburgh. Sampling frequency is 250 Hz. There are a total of 13 sets, each of which contains 1 minute nonseizure, 1 minute preseizure and 1 minute seizure data. 5 of these sets are used for the training and the remaining 8 sets are employed for the test. Hence, in this study, 2 main classes (nonseizure and seizure) are considered for the EEG data.

B. sAMDF instead of Curve Length

Curve length was originally proposed by Olsen [3] and suggested by Esteller et al [4] for seizure detection as a computationally efficient feature instead of fractal dimension. This feature was also used in a recent work [2] as a feature of epileptic seizure prediction. In [2], considering seizure prediction, it is stated that for half of the patients the curve length was the most promising feature among the other linear features selected by a simple genetic algorithm. This feature is conceived as the amplitude and frequency demodulator which can detect the probable changes in amplitude and frequency of a signal. Curve length is characterized by the simple formula below

$$CL = \sum_{i=1}^{N} \left| x(i-1) - x(i) \right| \tag{1}$$

We propose its more generalized form which is so-called sAMDF [5]

$$CLD(\tau) = \sum_{i=1}^{N} \left| x(i-\tau) - x(i) \right| \tag{2}$$

where τ is the delay or difference to compute the length. While curve length (notice that it is special case of sAMDF for $\tau = 1$) can only differentiate some frequency change in the signal; sAMDF may detect more critical frequency changes for epileptic behavior of EEG signal. In order to determine which delay τ can be more useful; we tested a set of EEG signals with nonseizure, preseizure and seizure parts respectively. Using a running window of length *N=1024* and shifted by $M=256$; all sAMDF values for a set are computed for each $\tau = [1,30]$. Measures computed per window are linearly scaled to an interval [0,255] for each of the 94 channels. Since this feature drops for seizure data; when it is below a threshold value, it is suspected to be a candidate for seizure data. Note that the best performance is obtained when the threshold is set to 40. The performance of the feature is determined by averaging the number of seizure suspected channels of segments (120-125 for this set) that show seizure-like activity. In Fig 1. performance of the feature is plotted against delays τ. It is observed that the performance of sAMDF oscillates for different τ values and the best result is obtained around $\tau = 5$. In order to validate this result, we compared the performances of the curve length and sAMDF($\tau = 5$) for a different nonseizure, preseizure and seizure set. Performances are given in Fig 2. as the sum of decided seizure onset channels per segment for both features. As it can be clearly seen from the figure, with $\tau = 5$ not only the seizure activity is discriminated much earlier; but also many more channels capture this activity. Since sAMDF for $\tau = 5$ outperforms the classical curve length for the discrimination of seizure-like activity; we prefer to use it as one of the features in our experiments.

C. Other Linear Features

The other features are 11 computationally efficient linear features that are used in epilepsy detection literature: statistical moments (variance, skewness, kurtosis) and some spectral parameters used in [6] (Hjorth's mobility & complexity parameters, spectral band powers, spectral edge frequency). Hence we used a total of 12 features with sAMDF to recognize the performance of the proposed preprocessing scheme on seizure detection.

D. Preprocessing with Divergence

In this section, we explain details of the divergence-based subspace approach as preprocessing for seizure detection. Our aim is to move the feature space to a more discriminated space while also reducing dimension.

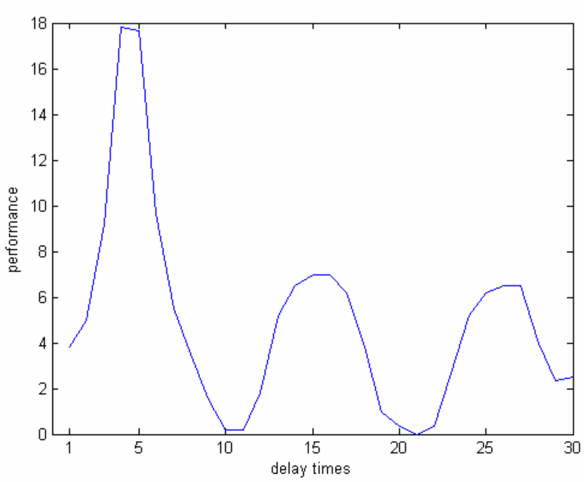

Fig. 1. Performance of sAMDF versus delay time τ

Divergence is a dissimilarity measure between 2 classes like c_1 and c_2 based on information theory [7]. Let the likelihood for an observed n dimensional feature vector x be defined as follows:

$$p_1(x) = p(x \mid c_1),\ \ p_2(x) = p(x \mid c_2) \qquad (3)$$

General divergence measure for all probability density functions is defined as [6]

$$J = \int_x \left[p_1(x) - p_2(x) \right] \ln \frac{p_1(x)}{p_2(x)} dx \qquad (4)$$

Then our main aim becomes to find a $(n\ X\ m)$ dimensional

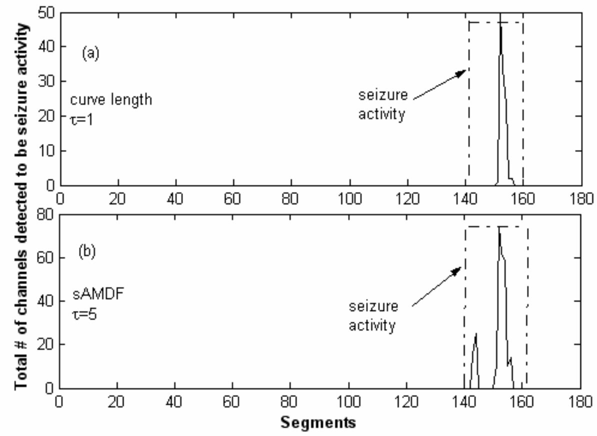

Fig. 2. Comparison of curve length and sAMDF for a set

matrix $\mathbf{A}$ such that the divergence measure between these two classes is maximized

$$y = \mathbf{A}^T x \qquad (5)$$

where y is a m ($m < n$) dimensional vector. If we assume that the features are normally distributed as $p_1(x) = N(\boldsymbol{\mu}_1, \Sigma_1)$ and $p_2(x) = N(\boldsymbol{\mu}_2, \Sigma_2)$; then the divergence measure in transformed y space becomes [8]

$$J = tr\left[(\mathbf{A}^T MW \mathbf{A})^{-1} \mathbf{A}^T MB \mathbf{A} \right] \qquad (6)$$

where MW is a within matrix,

$$MW = 0.5\left(\Sigma_1 + \Sigma_2 \right) \qquad (7)$$

and MB is a between matrix for these two classes,

$$MB = (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^T \qquad (8)$$

and it is shown in [9], that the transformation matrix that maximizes divergence J can be found as the first m eigenvectors of $MW^{-1}MB$ that correspond to the largest m eigenvalues. Notice that in this scheme, not the data itself but the features are assumed to be normally distributed.

TABLE I

UNITS DETECTED SEIZURE SEGMENT NUMBERS, MEAN VALUES AND VARIANCES OF SCALED LIKELIHOOD VALUES FOR ORIGINAL FEATURES (X) AND TRANSFORMED FEATURES (Y)

Set	C292243		C292249		C292342		C300113		C292330		C291247		C291441		C290636	
	x	y	x	y	x	y	x	y	x	y	x	y	x	y	x	y
Detected Seizure Segment	141	137	145	145	168	167	160	160	138	138	120	120	143	143	139	139
Mean of noseizure	0.9179	0.9385	0.8823	0.9210	0.9045	0.9332	0.8747	0.9417	0.9421	0.9630	0.9594	0.9738	0.9587	0.9699	0.9505	0.9613
Var of noseizure	0.0035	0.0028	0.0042	0.0046	0.0069	0.0051	0.0188	0.0048	0.0031	0.0017	0.0019	0.0004	0.0011	0.0012	0.0012	0.0011
False Alarms	No	No	No	No	No	No	Yes 3 FA	No	No	No	No	No	No	No	No	No

III. EXPERIMENTAL RESULTS

Data in all channels are divided into segments of length 1024 and shifted by 256 points. Features are extracted for all segments and normalized by dividing some constant value. Let c_1 be nonseizure and c_2 be seizure class. After mean vectors $(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2)$ and covariance matrices (Σ_1, Σ_2) are computed for each class; the linear transformation matrix **A** is obtained for $m=5$. Then multiplying the features by $\mathbf{A}^T$, mean vectors $(\boldsymbol{\mu}_{y1}, \boldsymbol{\mu}_{y2})$ and covariance matrices $(\Sigma_{y1}, \Sigma_{y2})$ are both computed in reduced and transformed y space. Totally mean vectors $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_{y1}$; covariance matrices Σ_1 and Σ_{y1}; and the linear transformation matrix **A** are saved for the test step.

Test is achieved for all 8 sets separately. After all features are extracted for one test set, likelihood of data is computed only for *nonseizure class*; that is both for original feature x space as $p_1(x \mid \boldsymbol{\mu}_1, \Sigma_1)$ and transformed y space as $p_1(y \mid \boldsymbol{\mu}_{y1}, \Sigma_{y1})$. Since the decision of seizure activity is based on only nonseizure data likelihood; a likelihood under some threshold value yields a seizure onset . Likelihood values for both spaces are linearly scaled into [0,1] in order to compare them and threshold is chosen as 0.5 to decide whether a seizure occurred.

Discrimination of seizure and nonseizure classes can be understood from the mean and variance values of nonseizure data in sets. Ideally likelihood values for nonseizure should be near unity and seizure activity should be near zero. Comparison of results for x space and y space is given in Table I. As observed from the table, nonseizure likelihood mean values for transformed space are greater and variance values are generally less than the values in the original space. This means discrimination between seizure and nonseizure activity is emphasized more for the transformed space y. Besides this, transformed space detects seizure earlier for two sets (for one set 1 second and for the other one 0.25 seconds earlier) and also while there are no false alarms for y space; 3 false alarms are raised for x space which verifies the discriminative ability of divergence-based linear transformation.

In Fig 3., likelihood values for both spaces are supplied for the set C292243. Notice that while the likelihood values are mostly higher for nonseizure activity; seizure activity is also detected earlier for y space (in 137^{th} segment) than for x space (in 140^{th} segment).

IV. CONCLUSION

Curve length, which is actually a special case of sAMDF for $\tau = 1$, has been previously used in epileptic seizure detection/prediction literature. In this study, we show that for our data, performance of sAMDF oscillates for different τ values. It should be declared that different τ values yield changes in different frequency regions. Hence it is empirically concluded that around τ =5, sAMDF reflects the more critical frequency change of seizure onset for our data.

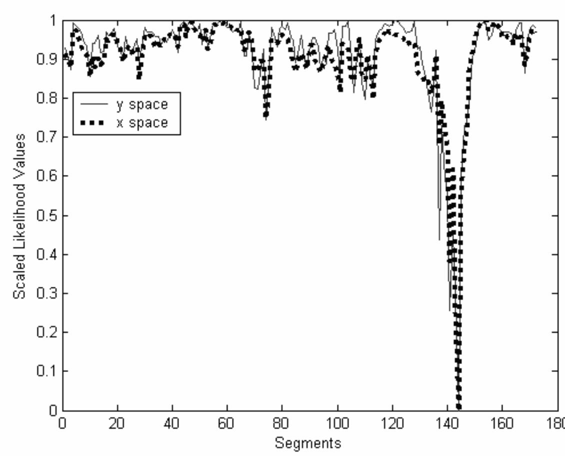

Fig. 3. Scaled likelihood values for the transformed y feature space (solid) and original x feature space (point) for set C292243

Another component of this study is to realize the effect of divergence-based feature selection as a preprocessing for seizure detection. In this approach, a class separation criterion (divergence measure) is used to have a linear transformation matrix that maximizes the distance between nonseizure and seizure data. 2 main conclusions are drawn:

- Dimension of feature space is reduced (from $n=13$ to $m=5$) significantly. Since our main objective is to understand the advantages and/or disadvantages of divergence for pre-processing of epileptic data; we use a very simple hypothesis testing to decide seizure onset. For a more complicated detector of seizures (like neural networks); this reduction will be much more valuable.

- Features between two classes are moved to a more discriminated space. Results show that the reduced transformed space is much better to differentiate seizure class from nonseizure class. Moreover for some data sets; seizures are detected earlier with the transformed feature space.

In this work, we chose to reduce the dimension space to an arbitrary *m* value of 5. Further study is required to compare performances of feature spaces for different *m* values. Additionally, it may be important to observe the performance of this feature selection method for various seizure detectors.

REFERENCES

[1] P. Zarjam, M. Mesbah and B. Boashash, "An optimal feature set for seizure detection systems for newborn EEG signals," *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, pp. 33-35, 2003.

[2] M. D'Alessandro et al., "Epileptic seizure prediction using hybrid feature selection over multiple intracarnial EEG electrode contacts: a report of four patients," *IEEE Trans. On Biomedical Engineering*, vol. 50, pp. 603-615, May 2003.

[3] P. Olsen, et al., "Automatic detection of seizures using electroencephalographic signals," *U.S. patent No*. 5.311.876, November 1992.

[4] R. Esteller, J. Echauz, T. Tcheng, B. Litt, and B. Pless, "Line length: An efficient feature for seizure onset detection," *Proc. of the 23rd Annual EMBS International Conf.*, October 2001.

[5] L. R. Rabiner and R. W. Schafer, *Digital Processing of Speech Signals*, Englewood Cliffs, NJ: Prentice-Hall, 1978.

[6] F. Mormann, "On the predictability of epileptic seizures," *Clinical Neurophysiology*, vol. 116, pp. 569-587, March 2005.

[7] J. Campbell, "Speaker recognition: A tutorial," *Proc. of the IEEE*, vol. 85, pp. 1437-1462, September 1997.

[8] P. C. Loizou and A. S. Spanias, "Improved speech recognition using a subspace projection approach," *IEEE Trans. On Speech and Audio Processing*, vol. 7, pp. 343-345, May 1999.

[9] K. Fukunaga, *Introduction to Statistical Pattern Recognition*, NY:Academic, 1972.