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

Sparse Approximation of Long-term Biomedical Signals For Classification Via Dynamic PCA

Shengkun Xie, Feng Jin and Sridhar Krishnan
Department of Electrical and Computer Engineering,
Ryerson University,
Toronto, ON M5B 2K3, CANADA.
The corresponding email: shengkun.xie@ryerson.ca

***Abstract*—Sparse approximation is a novel technique in applications of event detection problems to long-term complex biomedical signals. It involves simplifying the extent of resources required to describe a large set of data sufficiently for classification. In this paper, we propose a multivariate statistical approach using dynamic principal component analysis along with the non-overlapping moving window technique to extract feature information from univariate long-term observational signals. Within the dynamic PCA framework, a few principal components plus the energy measure of signals in principal component subspace are highly promising for applying event detection problems to both stationary and non-stationary signals. The proposed method has been first tested using synthetic databases which contain various representative signals. The effectiveness of the method is then verified with real EEG signals for the purpose of epilepsy diagnosis and epileptic seizure detection. This sparse method produces a 100% classification accuracy for both synthetic data and real single channel EEG data.**

***Index Terms*—Dynamic Principal Component Analysis, Sparse Approximation, Feature Extraction, Signal Classification.**

I. Introduction

EEG as the most effective and explicit method to reveals the characteristic findings in several epilepsy related syndromes [1]. The manual evaluation of EEG has therefore been adopted as the routine process in epilepsy monitoring and seizure detection for clinical diagnosis and treatment plan making. Unfortunately, due to the absence of objective definition for the characteristic waveforms related to epilepsy, the diagnosis involves a combination of the medical history of the patient and EEG interpretation by expert neurologists which is hampered by poor inter-observer reliability [2]. In automatic diagnosis and seizure detection, classification of EEG signal become an important task of this process. In particular, the long-term observational non-stationary time series of EEG data is difficult to analyze by many existing classification techniques including support vector machine or statistical processing tools such as conventional principal component analysis (PCA) because of the complex characteristics, few number of data samples and extremely high dimension of the signal.

In high-dimensional data classification, sparse representation of the underlying signal is a key to classification that aims for a simple classifier because of the data dimensional reduction property [3]. The sparse representation for signal classification looks for a low dimensional feature vector that contains the most important information. Feature extraction therefore involves simplifying the extent of resources required to describe a large set of data accurately. Some feature extraction approaches have been investigated for the study of epilepsy and seizure detection, including a measure of signal similarity in the correlation dimension ([4], [5]), a measure of energy variation [6] and a measure of accumulated energy [7]. However, these single scale measures or their combinations are not suitable for a long-term observational signal that appears to be multi-scale in nature. In this paper, we propose a multivariate statistical approach using dynamic principal component analysis (DPCA) to extract feature information from a univariate long-term observational signal. The use of DPCA aims first to introduce additional dimensions for the univariate signal to improve the representation of data similarity, and then to represent the signal via PCA-based low rank approximation to extract features that explain the major data variance. With the DPCA framework, a method of combining the first few PCs with the partial energy measures of the signal in PC space is proposed to deal with spiky non-stationary signals. In order to demonstrate its applications in epilepsy and seizure detection problems, synthetic databases that contain representative signals are first tested, followed by an investigation using a set of real EEG data.

II. Methodologies

A. Principal Components Extraction by Dynamic PCA

Many study of time series data in event detection problems suggests that a feature extraction method should be applied to understand the complex structures of the data and differences among them. The dependence of measurements suggests that additional time-dependent variables should be introduced to data analysis. Extraction of time-dependent variables was originally accomplished by introducing dynamic principal component analysis [8]. This method applied to one dimensional time series data considers a collection of observations, $\{y(1), y(2), \ldots, y(N)\} \subseteq R^p$ from a biomedical signal $y(k)$, where N is the number of observations. The data matrix $\mathbf{Y}$ for further PCA is arranged as follows:

$$\mathbf{Y} = [y(k-l+1), y(k-l+2), \cdots, y(k)], \quad (1)$$

where l is a time lag. This implies that using DPCA, one first needs to select l in order to further analyze the data. This procedure is often conducted by time series regression techniques that involve parametric modeling such as fitting the data to auto-regressive models. For feature extraction, the determination of l is through the cross-validation procedure, which selects the l that corresponds to the optimal classification accuracy. When l is chosen, the number of underlying variable of stochastic process is increased from 1 to l. PCA is then applied to the covariance matrix of the data matrix $\mathbf{Y}$ to evaluate the dynamics of stochastic processes $y(k)$ by analyzing the eigenvalues of the covariance matrix. In order to potentially improve the performance of DPCA in event detection problems, we propose a method of applying a non-overlapping moving window technique. This method decreases correlation of each extracted time-dependent variable in the window when the width of the moving window is large. Using the non-overlapping moving window technique, the data matrix constructed from the N observations of a biomedical signal $y(k)$, denoted by $\mathbf{D}^y$, is organized as follows:

$$\mathbf{D}^y = \begin{pmatrix} y(1) & y(2) & \dots & y(l) \\ y(1+l) & y(2+l) & \dots & y(2l) \\ \vdots & \vdots & \vdots & \vdots \\ y(ml-l+1) & y(ml-l+2) & \dots & y(ml) \end{pmatrix},$$

where $N = ml$, m is the total number of the moving windows of $y(k)$ each with length l. In this case, the observations of each variable in $\mathbf{D}^y$ are less autocorrelated if l is a larger value than the first significant time lag.

Application of the DPCA approach to each univariate time series allows extraction of additional variables in order to extend the number of variables from 1 to l, where l is the length of non-overlapping moving windows. The benefit of applying DPCA to univariate time series data is that, for a large value of l, the sequence of data, $y(k)$, $y(k+l)$, ..., $y(k+ml-l)$, for $k = 1, \dots l$ (i.e., the data of each column of the matrix $\mathbf{D}^y$), becomes approximately uncorrelated. In the training step of signal classification, this approach is applied to each long-term signal of the training set and the obtained data matrix from each signal is then formed together to become the training data matrix for further principal components extraction. Suppose that there are g groups of signal and there is only r signal for each group. The data matrix constructed from these r signals becomes $\mathbf{D} = [\mathbf{D}^{y_1\top}, \mathbf{D}^{y_2\top}, \dots, \mathbf{D}^{y_{gr}\top}]^\top$, with the size $mgr \times l$. Thus, after organizing these univariate time series data into the data matrix $\mathbf{D}$, PCA is then applied to map the matrix $\mathbf{D}$ into a new feature space. This possibly reduces the number of extended variables from the time domain if most of the data variance is explained by the first few PCs. In PCA, the principal component score matrix $\mathbf{L}$ and the principal component loading matrix $\mathbf{V} = (V_1, \dots, V_l)$ are obtained by decomposing the $mr \times l$ observation data matrix $\mathbf{D}$, into $\mathbf{D}$=$\mathbf{LV}$. Sparse approximation via PCA is then obtained by approximating $\mathbf{D}$ by using a linear combination of first few components, that is $\mathbf{D} \approx \hat{\mathbf{L}}\hat{\mathbf{V}}$, where $\hat{\mathbf{L}}$ and $\hat{\mathbf{V}}$ are low rank matrix and $\hat{\mathbf{V}}$ consists of only first few PCs. We call this method a first few PCs (FFPC) sparse approximation method. This sparse approximation method are particularly useful for classifying the stationary highly correlated signals.

Retaining only a few PCs for classification may cause insufficient dimensions of separating features when these first few PCs are only able to explain a small amount of the total data variation. The extracted features are the similarity measures between the observed data and each PCs coordinate, these similarity measures in terms of projection only may not be able to successfully separate the data into different classes. In order to improve the separability of the low dimensional feature vector, we may construct the feature vector that contains the first few PCs, e.g. first two PCs, $\hat{y}_1^s(k)$ and $\hat{y}_2^s(k)$ of the kth window plus the partial energy measures of the kth window in PC space, which is given as $E^{l_1}(k)$=$\sum_{v=1}^{l_1} \hat{y}_v^s(k)^2$, where k is the index of the non-overlapping moving window of the test signal and $k = 1, 2, \dots, m^*$. The l_1 is the number of major PCs selected and m^* is the total number of windows of test signals. This approach makes use of data similarity measure and data energy measure simultaneously. We refer this method to a first few PCs plus energy measure (PCPEM) method. The advantage of this method is that it enable to capture data characteristics in terms of both the data variation and the signal energy measure, in a subspace.

III. Applications

The proposed methods are first applied to synthetic databases that contain all kinds of representative signals including surrogate data with correlations, trends, and nonstationarities. The databases are available in PhysioNet [9], a public service of the Research Resource for Complex Physiologic Signals. The methods are then applied to an publicly available EEG database [10] for the purpose of epilepsy diagnosis and epileptic seizure detection, two important event detection problems in epilepsy study. We focus only on the sparse approximation using a low dimensional feature vector, i.e. a three dimensional feature vector, as an input of data classification, to improve the interpretability of sparse approximation techniques and to promote the proposed method in the application related to data visualization.

TABLE I
THE TABLE SUMMARIZES THE INFORMATION OF SYNTHETIC DATABASES.

	Description of synthetic databases
S_1	Correlated stationary; Signal 1 with α = 0.5, here α is the signal correlation; Signal 2 with α = 0.9, $N = 2^{17}$.
S_2	Surrogate signals with trends; Amplitude of trend A_s = 2; Period T = 128; Signal 1 with α = 0.9; Signal 2 with α = 0.1, $N = 2^{17}$.
S_3	Non-stationary signals with spikes; Amplitude A_{sp} = 1, N = 2^{17}; Signal 1 with spikes prob p = 0.05; Signal 2 with spikes signal only.
S_4	Signals with different local std dev; $N = 2^{18}$;α = 0.1, σ_1 = 1; Signal 1 with σ_2 = 4 (p = 0.05); Signal 2 with σ_2 = 4 (p = 0.95).
S_5	Signals with different local std dev; $N = 2^{18}$;α = 0.9, σ_1 = 1; Signal 1 with σ_2 = 4 (p = 0.05); Signal 2 with σ_2 = 4 (p = 0.95).
S_6	Signals with different local correlations; Width W= 20, $N = 2^{17}$ Signal 1 with α_1 = 0.1 (90%), α_2 = 0.9(10%); Signal 2 with α_2 = 0.9 (10%) only

TABLE II
THE RAND INDEX OF CLASSIFICATION RESULTS USING FIRST 3 PCS AS THE INPUT OF ONE NEAREST NEIGHBOR CLASSIFIER UNDER THE DPCA METHOD WITH DIFFERENT VALUES OF l.

Window Size	S_1	S_2	S_3	S_4	S_5	S_6
$l = 2^6$	0.750	0.819	0.619	0.439	0.956	0.436
$l = 2^7$	0.968	0.980	0.659	0.479	0.998	0.370
$l = 2^8$	1.000	1.000	0.640	0.485	1.000	0.248

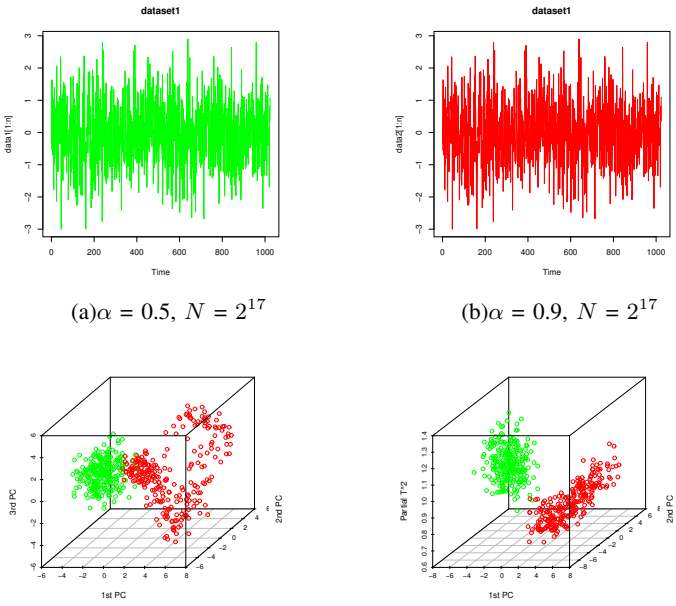
TABLE III
THE RAND INDEX OF CLASSIFICATION RESULTS USING FIRST 2 PCS PLUS THE PARTIAL ENERGY MEASURE AS THE INPUT OF ONE NEAREST NEIGHBOR CLASSIFIER UNDER THE DPCA METHOD WITH DIFFERENT VALUE OF l AND DIFFERENT VALUES OF l_1.

$l = 2^6$	S_1	S_2	S_3	S_4	S_5	S_6
$l_1 = 2^5$	0.760	0.698	0.992	0.990	0.999	0.961
$l_1 = 2^6$	0.777	0.735	1.000	0.999	1.000	0.996
$l = 2^7$						
$l_1 = 2^6$	0.949	0.972	1.000	1.000	1.000	0.972
$l_1 = 2^7$	0.968	0.976	1.000	1.000	1.000	1.000
$l = 2^8$						
$l_1 = 2^7$	1.000	0.984	0.961	0.992	1.000	0.900
$l_1 = 2^8$	1.000	1.000	1.000	1.000	1.000	1.000

A. *Classification of Synthetic Signals*

We consider the synthetic data sets available in [9] because these signals contain the important characteristics that often are observed from complex biomedical signals. The information about the synthetic data sets is summarized in Table I. Parts of the signals (the first 2^{10} data points) are depicted in Figures 1(a), 1(b), 2(a) and 2(b). Our study shows that classification based on the first few PCs is possible only for stationary signals. In Figures 1(c) and 1(d), the extracted few dimensional features are linearly separable, but the results shown in Figures 2(c) and 2(d) suggest that retaining only first few PC may fail the data classification that uses a linear classifier. However, the feature extraction using the first two PCs plus the partial energy measure performs better than the one with PCs only as the extracted features are more linearly separable.

The Rand index (RI) [11] is calculated to determine class membership agreement for evaluating the performance of signal classification. Table II shows the results of the RI for 6 different synthetic data sets with different choices of window size l. With a small value of l, the classification accuracy (specified by Rand index) is small, but with the increase of the value of l, the classification accuracy is dramatically improved, particularly for S_1, S_2 and S_5. The results shown in Figures 2(c) show that the FFPC method with the first 3 PCs only is not able to provide a separable features for classification using a simple classifier when the signals appear to be non-stationary, e.g. for data sets S_3, S_4, S_6. In Table III, with the increase of both l and l_1, the classification accuracy of using the PCPEM method that combines the first 2 PCs with the partial energy measure in the feature vector is increased. Also, this combined method is very promising in dealing with both stationary and non-stationary data, in particular, for the spiky signal classification. The results shown in Figures 1(d) and 2(d) report this issue for data set S_1, S_3 and S_4, where S_1 contains stationary signals and both S_3 and S_4 consist of non-stationary, spiky signals.



(a)α = 0.5, $N = 2^{17}$ (b)α = 0.9, $N = 2^{17}$

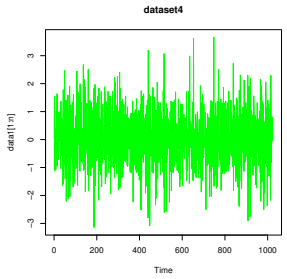
(c)FFPC with first 3 PCs (d)PCPEM with first 2 PCs plus energy measure

Fig. 1. The time series plot of the first 2^{10} time points of the signals and the scatter plots of extracted three dimensional features for data set S_1.

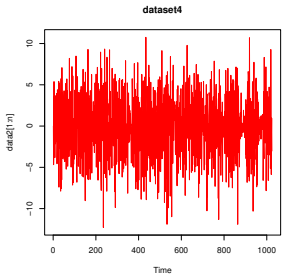
B. *Epilepsy Diagnosis and Epileptic Seizure Detection*

The biomedical applications of the sparse approximation in event detection problems are also demonstrated by using a set of EEG signals in both of epilepsy diagnosis and epileptic seizure detection problems. We consider the database in [10] that consists of a set of EEG signals coming from healthy volunteers, from patients during seizure-free intervals and from patients at an onset of epileptic seizure. The normal EEG signals (i.e., Sets A and B in [10]) behavior similarly to the synthetic data set S_1 and the epileptic EEG signals (Set C, D and E in [10])are similar to the types of synthetic data sets S_3 and S_4.

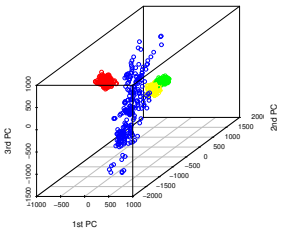
In order to diagnose epilepsy and detect epileptic seizures, the non-overlapping moving window technique is used to extract additional variables from an original univariate EEG training signal. This method partitions each EEG signal of Sets A, B, C, D and E using the window size l=256. We compare the performance of data classification based on the FFPC method and the PCPEM method, in both epilepsy diagnosis and epileptic seizure detection problems. From the scatter plots of three features extracted shown in Figures 3(a) and 3(b) one can see that both the FFPC method and the PCPEM method may perform similarly in the diagnosis of epilepsy. But the scatter plots of three features extracted shown in Figures 3(c) and 3(d) suggest that classification based on



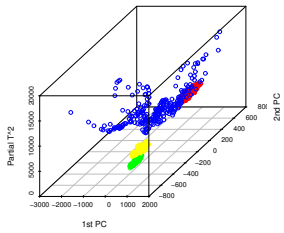
(a)σ_1 = 1, σ_2 = 4 (probability p =0.05)



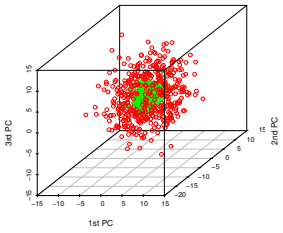
(b)σ_1 = 1, σ_2 = 4 (probability p =0.95)



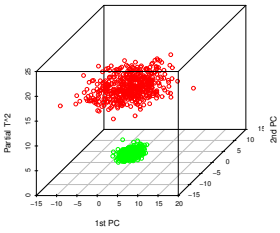
(a)FFPC for epilepsy diagnosis



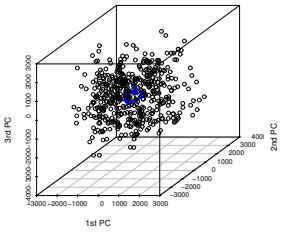
(b)PCPEM for epilepsy diagnosis



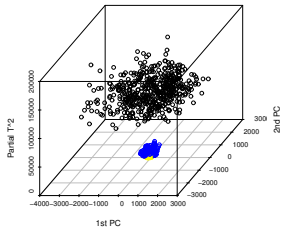
(c)FFPC with first 3 PCs



(d)PCPEM with first 2 PCs plus energy measure



(a)FFPC for seizure detection



(b)PCPEM for seizure detection

Fig. 2. The time series plot of the first 2^{10} time points of the signals and the scatter plots of extracted three dimensional features for data set S_1.

Fig. 3. Three dimensional features scatter plot obtained from the DPCA method for EEG signals in Sets A, B, C, D and E for the FFPC method and for PCPEM method in both epilepsy diagnosis and seizure detection. The scatter plots in green, red, yellow, blue and black colors stands for signals in Set A, B, C and D and E, respectively.

the features from the PCPEM method may lead to a higher classification accuracy than from the FFPC method in epileptic seizure detection problems. In epilepsy diagnosis with l=256, the classification leads to an accuracy rate of 95.9% for FFPC method and a rate of 97.8% for PCPEM method. When l=512, the PCPME method reach the 100% accuracy. In seizure detection, both of methods reach the 100% accuracy rate when l=512 is used.

IV. Conclusions

Using the DPCA framework, a small number of principal components (sparse approximation) are successful in the application of event detection in highly correlated signals, in particular, the stationary signals. Within the DPCA framework, a few principal components plus the energy measure of signals in PC subspace are highly promising in the application of event detection in both stationary and non-stationary signals. This methodology contributes to event detection problems in biomedical signal and is applicable to both stationary and non-stationary ones. As several characteristic EEG patterns are associated with well-defined epilepsy syndromes, it would be more clinically significant to classify EEG into more classes according to its corresponding epilepsy syndromes, which is important for selection of therapy and assessment of prognosis of the epilepsy. Differentiating between ictal and interictal EEG findings are clinically important but arbitrary, unlike the conventional epileptic seizure detection methods, the presented method can be expanded easily to accommodate multi-class classification of various epileptic EEGs.

References

[1] S. Noachtar and J. Rémi, The rold of EEG in epilepsy: A critical review, *Epilepsy & Behavior*, 15, **2009**: 22-33.

[2] G. W. Williams, H. O. Luders, A. Brickner, *et al*, Interobserver variability in EEG interpretation. *Neurology* 1985 **35**: 1714-1719.

[3] H. A. Chipman and H. Gu, Interpreable Dimension Reduction, *Journal of Applied Statistics*, Vol **32**, No.9, 969-987, November 2005.

[4] M. J. Quyen, Martinerie, M. Baulac and F. Varela, Anticipating epileptic seizures in real time by non-linear analysis of similarity between EEG recordings. *NeuroReport*, vol. 10, no.10, 2149-2155, 1999.

[5] S. Chandaka, A. Chatterjee, and S. Munshi. Crosscorrelation aided support vector machine classifier for classification of EEG signals. *Expert Systems With Applications*, **36**(2P1):1329–1336, 2009.

[6] R. Esteller, J. Echauz, M. D'Alessandro, G. Worrell, S. Cranstoun, G. Vachtsevanos and B. Litt. Continuous energy variation during the seizure cycle: towards an online accumulated energy. *Clin. Neurophy.*; 116:517-526, 2005.

[7] S. Gigolab, F. Ortiza, C. E. DAttellis, W. Silvac and S. Kochenc, Prediction of epileptic seizures using accumulated energy in a multiresolution framework. *Journal of Neuroscience Methods*; **138**:107–111, 2004.

[8] W. Ku, R. H. Storer, and C. Georgakis, Disturbance Detection and isolation by dynamic Principal Component analysis, *Chemometrics and Intelligent Laboratory Systems*, **30**, 179-196, 1995.

[9] *http://www.physionet.org/physiobank/database/♯synthetic*

[10] *http://epileptologie-bonn.de/cms/front_content.php?idcat=193*

[11] M. Rand, Objective Criteria for the Evaluation of Clustering Methods, *Journal of the American Statistical Association*, Vol. **66**, No. 336, pp. 846–850, 1971.