34th Annual International Conference of the IEEE EMBS
San Diego, California USA, 28 August - 1 September, 2012

Sleep EMG Analysis using Sparse Signal Representation and Classification

Mehrnaz Shokrollahi and Sridhar Krishnan

***Abstract*— The development of automatic sleep based abnormality detection in patient for sleep related problem is a key field in the recent research. However the sleep signals are obtained as long-time recordings and inhibit complex characteristics, making their analysis computationally challenging. As a result, recognition methods that facilitate efficient dimensionality reduction are developed to suit different applications. In recent years sparse representation schemes provide an effective means for achieving best possible data reduction by comparing the input with pre-formulated dictionaries, especially for huge datasets. Recent research proves the usability of these methods for signal classification. In this paper, a robust technique is provided for sparse representation of small dataset signal types. Here, the signal decomposition is obtained using the ℓ^1-minimization technique, following which a generalization based on the leave-one-out (LOO) is performed. The dependency of the proposed algorithm is analyzed, using a sparsity measure, in order to verify the dependency between the input data and extracted feature space. Performance measures obtained using long-term sleep data shows an average classification accuracy of 80% and further validates the usefulness of the technique for long term biomedical signal analysis.**

I. INTRODUCTION

A significant amount of research and effort has been devoted to the analysis of the effect of the size of both the testing and training samples on the design of the pattern recognition systems specially in biomedical applications [1]. In general, due to the nature of biomedical signals, many variables are likely to be correlated. When the number of variables are high, an increase in the discriminatory power is not necessarily ensured. Thus a subset of these variables can be chosen such that the others may not contain substantial additional information. Many different variable selection methods have been proposed; among those, sparse representation have attracted a great deal of attention in the past few years for applications such as face recognition, text recognition, gene expression array analysis and many more. However, to the best of authors knowledge there has not been any work done on analysis of sparse representation for EMG analysis in sleep.

Sparse representation schemes provide a compact representation for signals using a combination of atoms (or elements) chosen from an over-complete dictionary. However, owing to non-availability of overcomplete dictionaries that suits all data types, the representation tends to span through multiple dictionaries resulting in residual errors during signal approximation. Recent research in sparse approximation involves extension of these techniques towards obtaining optimal signal discrimination, although some of the conventional techniques such as Fourier transform and wavelet transform operate on the principle of sparseness in a broader sense.

Following the introduction of sparse theory, many approximation tools were developed for specific tasks [2]. The main principle behind these techniques is that for a given signal (often a larger dataset), the representation that offers the most compact solution (and corresponding minimal error) is considered for further analysis and/or decision making [3].

Huang *et al*. [4] proposed the use of sparse representation for signal classification application, by incorporating reconstruction properties, discrimination power and sparsity during the approximation process. The coding also uses an empirical set of weight factors to optimize the overall performance such that the error rate is reduced to 0.25 when the occlusion is present. However, their validation is based on a non-linear Fisher information score. Following this, Wright *et al*. [5] considers the problem of automatic face recognition. They propose a general algorithm for image-based object recognition and claimed that if sparsity is properly harnessed, the choice of features is no longer critical i.e. number of features does not have a direct correlation with the performance as long as the sparseness criterion is satisfied. Sparse techniques have also been successfully used for cancer diagnosis applications using gene expression data [6], proving their usability for real-life signals. The authors performed a rigorous validation approach whereby the performance of the algorithm was compared with a variety of support vector machines. The original goal of these works was reconstruction using sparse representation in face database, where the limitation of the sample size is not an issue.

Irrespective of the availability of these techniques for classification applications, there still exist certain open problems such as [5]:

1) Incompleteness of sparse theory
2) Non-availability of a robust technique (or generalization) that can accommodate small dataset
3) Dimension reduction (Feature extraction)of long term signals (such as sleep) is important
4) Ease of Validation that could accommodate a variety of synthetic and real signals.

In addition to the above, for a given classification problem, lack of adequate samples tends to bias the training and testing results, limiting their performance consistency for higher-

*This work has been funded by Canada Research Chair (CRC) program.
M. Shokrollahi and S. Krishnan are with the Department of Electrical and Computer Engineering, Ryerson University, 350 Victoria St, Toronto, Ontario, Canada, M5B 2K3. mshokrol at ee.ryerson.ca & krishnan at ee.ryerson.ca

order problems. To reduce this inconsistency, Leave-One-Out (LOO) method is considered an optimal validation approach, with a least biased estimate [7]. This research work attempts to address the issue associated with lack of robust generalization techniques for sparse approximation when using small dataset. We propose generalization algorithms, based on the linear programming problem, using the LOO approach. The resulting signal approximates are validated based on the degree of sparsity measure and classification accuracy. Further, a preliminary investigation about the dependency between the input and feature vectors sets is provided.

A long-term sleep signal collected during Rapid eye movement (REM) stage is used for performance assessment. In Section II we present a brief overview about the sparse approximation theory. The methodology for sparse signal classification is elaborated in Section III and the proposed generalization algorithm is explained in Section III-B. Section IV discusses the performance evaluation of the algorithm for sleep signals. Finally, in Section V we conclude the paper by highlighting the main contributions of this work.

II. Sparse Representation by ℓ^1 minimization

Let $A_i = [a_1^i, a_2^i, ..., a_n^i] \in \Re^{dXn_i}$ be a matrix and each column of which is the training samples from the i^{th} class, where $i = 1, ..., k$, and $\mathbf{y} \in \Re^{dX1}$ be a new test sample. If a_{lj}^i and y_j be the j^{th} entry of a_l^i and $\mathbf{y}$ respectively, matrix A will be the entire training set in which it's elements are set by concatenating the training sample of all k classes:

$$A = [A_1, A_2, ...A_k] = [A_1^1, A_2^1, ..., A_{n_k}^k] \in \Re^{dXn} \quad (1)$$

then a linear representation of $\mathbf{y}$ can be rewritten in terms of all training samples as

$$\mathbf{y} = A\mathbf{x}_0 \in \Re^d \quad (2)$$

where ideally

$$\mathbf{x}_0 = [0, ...0, \alpha_{i,1}, \alpha_{i,2}, ..., \alpha_{i,n_i}, 0, ...0]^T \in \Re^n \quad (3)$$

is a coefficient matrix whose entries are zero except those associated with the i^{th} class. In other words, the nonzero entries in the estimate $\mathbf{x}_0$ will all be associated with the columns of A from a single object class i, so we can easily assign the test sample $\mathbf{y}$ to the class.

From (3) it is apparent that the representation of $\mathbf{y}$ is naturally sparse if the number of object classes k is reasonably large. So the problem can be converted into finding a column vector $\mathbf{x}$ such that $\mathbf{y} = A\mathbf{x}$ and $\|\mathbf{x}\|_0$ is minimized, where $\|\mathbf{x}\|_0$ is ℓ_0 -norm, and it is equivalent to the number of nonzero components in the vector $\mathbf{x}$, which is the so-called sparse representation. This can be expressed as the following optimization problem [8]:

$$\hat{\mathbf{x}}_0 = argmin\|\mathbf{x}\|_0 \quad \text{subject to: } A\mathbf{x} = \mathbf{y} \quad (4)$$

However, finding the solution to sparse representation problem is NP-hard due to its nature of combinational optimization [5][9]. Nevertheless, recent developments in the theory of sparse representation and compressed sensing [4][10]have shown that if the solution x_0 is sparse enough, the solution of ℓ_0 minimization is equivalent to the following ℓ_1 minimization problem

$$\hat{\mathbf{x}}_1 = argmin\|\mathbf{x}\|_1 \quad \text{subject to:} A\mathbf{x} = \mathbf{y} \quad (5)$$

This problem can be solved in polynomial time by standard linear programming (LP) methods i.e. [11].

III. Methodology

A. Classification Based on Sparse Representation

Variable selection methods such as Principal Component analysis (PCA) seeks for directions to project the data such that the projected data explain most of the variability of the original setting. In this way one obtains a low dimensional representation of the data without losing much information. PCA in general targets for the problems that are mainly unsupervised which may not be suitable for many classification problems (PCA has been used as a preprocessing step for a classifier). Sparse representation on the other hand establishes a more rigorous mathematical framework for studying long term biomedical signals, consequently attracting a great deal of attention in the past few years. However, to the best of authors knowledge there has not been any work done on analysis of sparse representation for long term EMG analysis in sleep and this is the main motivation behind this work. In the following we will explain how our algorithm works:

Given a new test sample $\mathbf{y}$, the sparse representation $\hat{\mathbf{x}}_1$ is computed using (5). In an ideal case, the nonzero entries in the estimate $\hat{\mathbf{x}}_1$ will all be associated with the columns of A from a single object class i, and based on the global sparse representation $\mathbf{y}$ can simply be assigned to the object class with the single largest entry in $\hat{\mathbf{x}}_1$. Yet, noise and modeling error may lead to small nonzero entries associated with multiple object classes [5].

For each class i, δ_i: $\Re^n \longrightarrow \Re^n$ is the characteristic function which selects the coefficients associated with the i^{th} class. In other words, for $\mathbf{x} \in \Re^n$, $\delta_i(\mathbf{x}) \in \Re^n$ is a new vector whose only nonzero entries are the entries in $\mathbf{x}$ that are associated with class i. The approximate of the given test sample $\mathbf{y}$ is $\hat{\mathbf{y}}_i = A\delta_i(\hat{\mathbf{x}}_1)$, which is estimated using only the coefficients associated with the i^{th} class. In this work, a generalized sparse representation is achieved using the training samples by assigning $\mathbf{y}$ to the object class that minimizes the residual between $\mathbf{y}$ and $\hat{\mathbf{y}}_i$:

The algorithm given below summarizes the complete classification procedure. Our implementation minimized the ℓ^1 -norm using LP algorithm based on [11]. Our algorithm is different from previous attempts because we focus to attain a method using LOO and later to generalize it to Leave-M-Out (LMO) for classification of the sparse coefficients.

Algorithm : Generalized algorithm based on Sparse Representation using Leave-M-Out Cross-Validation

1) **Input**: a matrix of feature set, M number of samples to be out for testing

2) Use cross validation to randomly split the data into a distinct matrix of training samples $A = [A_1, A_2, ...A_k] \in \Re^{dX(n-m)}$ and test sample $\mathbf{y} \in \Re^{dXm}$
3) Solve the ℓ^1 minimization problem
$\hat{\mathbf{x}}_1 = argmin\|\mathbf{x}\|_1 \;\; subject \; to \; A\mathbf{x} = \mathbf{y}$
4) Compute the degree of sparsity of $\hat{\mathbf{x}}_1$
5) Compute the residuals: $r_i(\mathbf{y}) = \|\mathbf{y} - A\delta_i(\hat{\mathbf{x}}_1)\|_2$
6) Identify($\mathbf{y}$) $= argmin_i \;\; r_i(\mathbf{y})$
7) Leave-M-Out Cross-validate to check the correctness of the results (special case LOO)
8) **Output**: The average of correctly identified.

The choice LOO in this algorithm refers to cases where deficient number of samples is accessible. In these cases the unbiased approximation and the variability of the estimate when the entire training set is varied is approximately zero, i.e. the difference between the expected LOO error and the true error is approximately zero, where the expectation is over random training sets of samples from the same underlying distribution.

Prior to proceeding on with the validation of the proposed algorithm for EMG in REM sleep, the sparsity that the method has to offer for signal representations is evaluated. This is done by computing the sparsity index, as defined by Hurly *et al* [12] and is explained in the following subsection.

B. Sparsity Measure

There is a great deal of study suggesting that random features perform well for sparse signal recovery [13]. For example for any signal x that are sparse or sparse in some known orthobasis, the properties of sparse recovery from random measurements are well understood. If A does not have good properties for sparse recovery using ℓ^1 minimization, then signal could not be recovered [14][15]. On the other hand the measure of sparsity has not been fully analyzed for signal classification. The motivation is that in this work we are looking for markers where the characteristics of the representation is more visible by attempting to represent them in a sparser domain. These markers can be used in integration with a defined feature space for obtaining more robust classification.

In this paper, we exploit the discriminative nature of sparse representation since the residual measures how well the representation approximates the test sample; and the degree of sparsity measures how good the representation itself is. Several sparseness measure have been proposed and used in literature. One of the sparsity functions (S_p) that has been used in this paper for each sample x is as follows [12]:

$$S_p(x) = (\frac{1}{C_p})(\frac{(\frac{1}{n-1})\sum_{i=1}^{n}|x_i - m|^p}{\frac{1}{n}\sum_{i=1}^{n}|x_i|^p})^{\frac{1}{p}} \quad (6)$$

where

$$m = (\frac{1}{n})\sum_{i=1}^{n} x_i \quad (7)$$

and

$$C_p = (\frac{(n-1)^{p-1}+1}{n^{p-1}})^{\frac{1}{p}} \quad (8)$$

The default value for p is chosen to be one and n is the length of the feature set . The sparsity is one if and only if a vector contains a single non-zero components, and is zero if and only if all the components are equal.

IV. Experimental Results

We test the sparse representation-based LMO classification algorithm using Autoregressive (AR) coefficients from the Electromyogram (EMG) signals in rapid eye movement (REM) stage of sleep [16]. In sleep laboratories sensors gather huge amount of data in a computer, and then sleep experts study the signals to extract the information from the data. As they gather a huge amount of data, looking for an event which could be only a small part of the data requires highly trained technicians, and fatigue of the technician due to long work hours could affect the accuracy of diagnosis. Thus, automation could play a role in development of faster, cheaper and reliable mobile health care. In other words if we apply the appropriate tools to automatically extract accurate information and provide valuable inputs to health experts, they could spend less time on extracting information and more time on helping cure the diseases.

A. Database

The dataset consists of signal segments from 8 chin EMG signals (4 with normal behavior and 4 with RBD) undergone the sleep test. A traditional scoring system for sleep has been established [17], with the electrophysiological parameters of EEG, EOG and EMG. The system used for recording chin EMG signals during sleep includes 3 relatively midline electrodes, one above the jaw line, one below the jaw line and one back-up electrode. The two electrodes are typically subtracted from another to eliminate artifacts shared by both electrodes. The EMG signal is freely triggered and band-pass filtered at $10 - 100$ Hz. The impedance of each electrode is less than $10k\Omega$ with a minimum digital resolution of 12 bits per sample. The sampling rate is 256 Hz. Similar electrodes are used to record EEG and EOG amongst other physiological parameters. In this study, a subject is defined as historically normal if there is no history of any violent behavior during the night sleep; otherwise it is considered as abnormal.

B. Evaluation

We computed the classification rates with the feature space dimensions of 956×26. This feature space is calculated using adaptive signal processing to adaptively segment the signal into stationary segments and then use each segment to calculate the AR coefficients of model order 26 [16]. In Table I, we present the classification accuracy of the AR coefficients using sparse representation for different M, as well as the degree of sparsity of the sparse approximation coefficients. The degree of sparsity for all sparse approximation of this feature set is calculated using Equation 6. From this Table, it

TABLE I: The classification accuracy of AR coefficients feeded into a sparse representation algorithm for different M of Leave-M-out as well as the degree of sparsity of sparse coefficients

M	**Classification Accuracy** (%)	**Sparsity of the Sparse Coefficients**
1	80.5	0∼1 (avg = 0.79)
10	79.8	0.18∼0.78 (avg = 0.67)
100	77.1	0.37∼0.67(avg = 0.55)
200	66.4	0.4∼0.56 (avg = 0.5)
400	59.7	0.39∼0.58 (avg = 0.5)
478	59.5	0.41∼0.56 (avg = 0.49)

is apparent that, the generalization scheme in practical application strongly depends on the value of M. It is also visible that sparsity measures reduces as we increase the value of M. The higher the value of M, the lower is the accuracy. Although LOO cross-validation method generally incurs a high computational cost, it is the least biased estimate since rigorous validation is performed compared to LMO cross-validation which exhibit a comparatively higher variance [7]. In this case, the algorithm reaches its maximum learning capacity (stability) by LOO method, since the perturbation induced by LOO is small and therefore the classifier is stable. As we increase the size of the perturbation, stability is less likely to hold and the sparse approximation is less sparse, thus the accuracy reduces by about 20%. To prove the generality of the algorithm, we analyzed the algorithm using different number of segments. This is shown in Figure 1, where we used 10%, 20%,...100% of the segments and applied LMO for both $M = 1$ and $M = 50\%$ of the original feature space. The overall performance of all the cases when $M = 1$ is about 80.5%, however this value reduces by about 20% to 60% as we increase M. The results compared to the previous work [16] shows that sparse representation for classification using LOO generalization increases the overall performance accuracy by about 10% in which the overall implementation of the algorithm itself takes only few seconds on a typical 3 GHz PC.

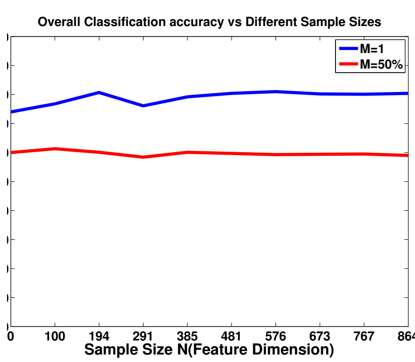

Fig. 1: **Overall classification accuracy for $M = 1$ and $M = 50\%$ of the feature space for different segment length of the AR coefficients**

V. Conclusion

In this paper, we have contended both theoretically and experimentally that exploiting sparsity is critical for performance evaluation of EMG in REM sleep. We proposed a novel generalization algorithms, based on the linear programming problem, using the LOO and LMO approaches. The resulting signal approximates are validated based on the degree of sparsity measure and classification accuracy. With sparsity properly harnessed, the inconsistency of lack of generalization reduces when LOO method is considered. Performance measures obtained using long term sleep data has validated the usefulness of the technique for real time data processing of long term biomedical signals.

References

[1] S.J. Raudys and A.K. Jain, "Small sample size effects in statistical pattern recognition: Recommendations for practitioners," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 13, no. 3, pp. 252–264, 1991.

[2] J. Rissanen, "Modeling by shortest data description," *Automatica*, vol. 14, no. 5, pp. 465–471, 1978.

[3] M.H. Hansen and B. Yu, "Model selection and the principle of minimum description length," *Journal of the American Statistical Association*, vol. 96, no. 454, pp. 746–774, 2001.

[4] K. Huang and S. Aviyente, "Sparse representation for signal classification," *Advances in neural information processing systems*, vol. 19, pp. 609, 2007.

[5] J. Wright, A.Y. Yang, A. Ganesh, S.S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 210–227, 2008.

[6] H. Xiyi and W. Fang-Xiang, "Sparse representation for classification of tumors using gene expression data," *Journal of Biomedicine and Biotechnology*, vol. 2009, 2009.

[7] R. Kohavi, "A study of cross-validation and bootstrap for accuracy estimation and model selection," in *International joint Conference on artificial intelligence*. Citeseer, 1995, vol. 14, pp. 1137–1145.

[8] D.L. Donoho and M. Elad, "Optimally sparse representation in general (nonorthogonal) dictionaries via l1 minimization," *Proceedings of the National Academy of Sciences*, vol. 100, no. 5, pp. 2197, 2003.

[9] M. Xu, C. Zheng, L. Zhang, and N.T.Y. Vincent, "Tumor classification via sparse representation based on metasample," in *Knowledge Acquisition and Modeling, 2009. KAM'09. Second International Symposium on*. IEEE, 2009, vol. 1, pp. 31–34.

[10] A. Banerjee, R. Juang, J. Broadwater, and P. Burlina, "Sparse feature extraction for support vector data description applications," in *Geoscience and Remote Sensing Symposium (IGARSS), 2010 IEEE International*. IEEE, pp. 4236–4239.

[11] E. Candes and J. Romberg, "l1-magic: Recovery of sparse signals via convex programming," *URL: www. acm. caltech. edu/l1magic/downloads/l1magic. pdf*, 2005.

[12] N. Hurley and S. Rickard, "Comparing measures of sparsity," *Information Theory, IEEE Transactions on*, vol. 55, no. 10, pp. 4723–4741, 2009.

[13] J. Wright and Y. Ma, "Dense error correction via l1-minimization," in *Acoustics, Speech and Signal Processing, 2009. ICASSP 2009. IEEE International Conference on*. IEEE, 2009, pp. 3033–3036.

[14] R. Gribonval, H. Rauhut, K. Schnass, and P. Vandergheynst, "Atoms of all channels, unite! average case analysis of multi-channel sparse recovery using greedy algorithms," *Journal of Fourier analysis and Applications*, vol. 14, no. 5, pp. 655–687, 2008.

[15] D.L. Donoho and J. Tanner, "Precise undersampling theorems," *Proceedings of the IEEE*, vol. 98, no. 6, pp. 913–924, 2010.

[16] M. Shokrollahi, S. Krishnan, D. Jewell, and B. Murray, "Autoregressive and cepstral analysis of electromyogram in rapid movement sleep," in *World Congress on Medical Physics and Biomedical Engineering, September 7-12, 2009, Munich, Germany*. Springer, 2009, pp. 1580–1583.

[17] A. Rechetschaffen and A. Kale, "Manual of standardized terminology techniques and scoring system for stages of human subjects 3rd. los angels: Brain information service," *Brain Research Institute of Uni. versity of California*, 1973.