

# TWO ALTERNATIVES FOR ATOMS SELECTION APPLIED TO SCREENING FOR SLEEP DISORDERS

Román E. Rolon<sup>b</sup>, Leandro E. Di Persia<sup>b</sup>, Hugo L. Rufiner<sup>b,†</sup> and Rubén D. Spies<sup>‡</sup>

<sup>b</sup>*Instituto de Investigación en Señales, Sistemas e Inteligencia Computacional (sinc(i), UNL-CONICET),  
Facultad de Ingeniería y Ciencias Hídricas, Univ. Nacional del Litoral, Santa Fe, Argentina,  
rrolon@santafe-conicet.gov.ar, ldipersia@fich.unl.edu.ar, lrufiner@gmail.com*

<sup>†</sup>*Laboratorio de Cibernética, Facultad de Ingeniería, Univ. Nacional de Entre Ríos, Argentina.*

<sup>‡</sup>*Instituto de Matemática Aplicada del Litoral (IMAL), CONICET-UNL, Santa Fe, Argentina,  
rspies@santafe-conicet.gov.ar*

**Abstract:** The Obstructive Sleep Apnea-Hypopnea Syndrome is characterized by repetitive episodes of upper airway obstruction that occur while sleeping, usually associated with a reduction in blood oxygen saturation ( $SaO_2$ ). In this article, application of sparse representations of  $SaO_2$  signals over a subcomplete dictionary for classification tasks is discussed. A sparse representation describes an  $SaO_2$  signal in terms of a linear combination of a few columns of a previously learned dictionary. The  $SaO_2$  signals are used in order to predict the occurrence of apnea-hypopnea events. A dictionary is learned by using a statistical method. Then a greedy pursuit algorithm is used in order to find the solution of a linear inverse problem with sparse constraint. The sparse vectors are used as input of a multilayer perceptron neural network. Finally an apnea-hypopnea index is estimated to grade the severity of OSAHS. Different alternatives for exploiting the activations of most discriminative atoms are evaluated.

**Keywords:** *sparse representations, inverse problems, apnea-hypopnea*

2000 AMS Subject Classification: 21A54 - 55P54

## 1 INTRODUCTION

The Obstructive Sleep Apnea-Hypopnea Syndrome (OSAHS) occurs when there are repeated episodes of complete or partial blockage of the upper airway during sleep. The current diagnostic tool for detecting OSAHS is an overnight polysomnography (PSG) in a sleep laboratory. To grade the OSAHS severity, an index called Apnea Hypopnea Index (AHI) is defined. The AHI index is obtained by counting the total number of Apnea-Hypopnea (AH) events per hour while sleeping. A patient with an AHI lower than 5 is considered *normal*, between 5 and 15 *mild*, between 15 and 30 *moderate* and more than 30 *severe* [2].

In the last fifteen years, many different approaches to traditional signal processing problems were taken. Some of these new formulations gave rise to techniques based on non-linear systems and higher-order statistics, including Independent Component Analysis (ICA) [3] and methods to obtain a Sparse Representation (SR) [9] of a signal.

In a previous work [11] a method was used in order to detect AH events by using only the  $SaO_2$  signal. Now a different approach for detecting AH events by using sub-dictionaries is taken.

In this work we start by comparing the performances of two methods for selecting the atoms of a dictionary. The dictionaries are used as generators of an SR of the  $SaO_2$  signal. A subset of columns of a dictionary is selected and used to solve a linear inverse problem with sparse constraint. Finally the sparse vectors are used in order to train a Multilayer Perceptron (MLP) neural network, which detects the respiratory events.

The next section describes the methods used to learn a dictionary and to estimate the coefficients of an SR for a given signal.

## 2 METHODS

### 2.1 SPARSE REPRESENTATIONS

By a dictionary we shall mean a matrix  $\Phi \in \mathbb{R}^{N \times M}$  whose columns  $\phi_j$  are called atoms. A representation of a signal  $\mathbf{s} \in \mathbb{R}^N$  in terms of a fixed dictionary  $\Phi$  can be stated as follow:

$$\mathbf{s} = \sum_{i=1}^M \phi_j a_j = \Phi \mathbf{a}, \quad (1)$$

where  $\mathbf{a} = (a_j) \in \mathbb{R}^M$ . Although inappropriately, the term ‘‘basis’’ instead of ‘‘dictionary’’ is sometimes used. Since the atoms are not required to be linearly independent and quite often more atoms than the space dimension are used, the latter situation is usually preferred.

When there are more columns in the dictionary than the size of  $\mathbf{s}$ , i.e. when  $M > N$  (in which case the dictionary is called overcomplete), or when the columns do not form a basis, then there may be non-unique representations of a given signal. In this situation a suitable criterion is required to select only one of those representations. In this context, sparsity often refers to the criterion of choosing a representation with just a few non-zero coefficients.

## 2.2 DICTIONARY LEARNING

A slightly more general framework is assumed, where equation (1) is modified to include an additive Gaussian noise  $\varepsilon$  as follows:

$$\mathbf{s} = \Phi \mathbf{a} + \varepsilon. \quad (2)$$

Following usual ICA terminology, equation (2) is referred to as the generative model, meaning that one generates the signal  $\mathbf{s}$  from a set of hidden sources  $a_j$ , arranged as the state vector  $\mathbf{a}$ , using the dictionary  $\Phi$ . The sources  $a_j$  are initially assumed to be statistically independent with a joint prior density  $\pi_{\text{prior}}(\mathbf{a}) = \prod_{j=1}^M \pi(a_j)$ . The atoms of  $\Phi$  can be estimated by maximizing the log-likelihood function of the data, given the dictionary [6],  $L(\mathbf{s}, \Phi) \doteq E[\log \pi(\mathbf{s}|\Phi)]$ , i.e. as follows:

$$\hat{\Phi} = \underset{\Phi}{\operatorname{argmax}} L(\mathbf{s}, \Phi). \quad (3)$$

The log-likelihood function can be found by marginalizing the product of the conditional distribution of the data given the dictionary and the prior distribution of the coefficients:  $\pi(\mathbf{s}|\Phi) = \int_{\mathbb{R}^M} \pi(\mathbf{s}|\Phi, \mathbf{a}) \pi_{\text{prior}}(\mathbf{a}) d\mathbf{a}$ . The maximum in equation (3) can also be approximated by using a gradient ascent method with the updating rule  $\Delta \Phi = \eta \Lambda_{\varepsilon}((\mathbf{s} - \Phi \mathbf{a}_{MAP}) \mathbf{a}_{MAP}^T - \Phi H^{-1})$ , where  $\mathbf{a}_{MAP}$  denotes the mode of the posterior distribution of the coefficients, i.e.  $\mathbf{a}_{MAP} = \underset{\mathbf{a} \in \mathbb{R}^M}{\operatorname{argmax}} \pi_{\text{post}}(\mathbf{a}|\Phi, \mathbf{s})$ ,  $H$  is the Hessian of the log-posterior evaluated at  $\mathbf{a}_{MAP}$ ,  $\eta$  is a positive coefficient (called learning rate parameter) and  $\Lambda_{\varepsilon}$  is the inverse of the covariance matrix of the noise [1, 7].

In order to obtain  $\Phi$  and the vector coefficient  $\mathbf{a}$ , the implementation proposed by Lewicki and Olshausen [6] was used at the dictionary training stage.

## 3 EXPERIMENTS

The Sleep Heart Health Study (SHHS) database<sup>1</sup> is used for this work. This database contains exhaustive information about detailed studies which are appropriately designed to investigate the relationships between sleep breathing disorders and cardiovascular diseases. The full dataset contains nearly 1000 complete PSGs, each one of them containing several biomedical signals such as electrocardiogram (ECG), nasal airflow, respiratory effort and  $SaO_2$ , among others. Annotations of sleep stages, arousals and respiratory events (apnea and hypopnea) are also included. Only the  $SaO_2$  signals and the AH events will be of our interest.

First of all, a wavelet processing technique is used for denoising the  $SaO_2$  signal, which was sampled at 1Hz. The denoised  $SaO_2$  signal is then obtained by making zero the approximation coefficients, at level 8, of the discrete Dyadic Wavelet Transform (DWT) with a mother function Daubechies 2 [5]. In the sequel, the  $SaO_2$  signal shall always refer to the denoised one.

For this article, a subset of 84 studies are selected in order to analyze the performance of two screening alternatives for sleep disorders. The dataset is divided into training and test sets consisting of 20 and 64 studies, respectively. The training set contains 4 groups of 5 PSGs each corresponding to AHI values below

<sup>1</sup>Database: <http://physionet.org/physiobank/database/shhpsgdb/>

5, between 5 and 10, between 10 and 15 and above 15. The test set comprises 64 PSGs with different degrees of illness. For both training and test sets, each  $SaO_2$  signal is segmented into vectors of length 128, with an overlapping of 32 elements. The segments are then arranged as column vectors  $\mathbf{s}_j \in \mathbb{R}^{128}$ . These segments are also labeled as belonging to *class 1* or *class 2*, depending on whether they contain AH events or not, respectively.

Next, we construct two class-training matrices  $S_{train}^{c1}$  and  $S_{train}^{c2}$  by stacking side-by-side all vectors  $\mathbf{s}$  labeled as *class 1* and *class 2* in the training set, respectively. The matrix  $S_{train}$  is then defined as  $S_{train} = [S_{train}^{c1} \ S_{train}^{c2}]$ , while the matrix  $S_{test}$  is built by stacking side-by-side all vectors  $\mathbf{s}$  in the test set.

In the next step two complete dictionaries  $\Phi_{c1}$  and  $\Phi_{c2}$  are learned by using matrices  $S_{train}^{c1}$  and  $S_{train}^{c2}$ , respectively. An overcomplete dictionary  $\Phi_1$  is then built as  $\Phi_1 = [\Phi_{c1} \ \Phi_{c2}]$ . Also, a complete dictionary  $\Phi_2$  is learned without taking into account class information, i.e. by using the whole matrix  $S_{train}$ . In all cases the learning process is performed by means of the Noise Overcomplete ICA (NOCICA) [7] method.

Now given  $\mathbf{s} \in \mathbb{R}^M$ , a constant  $k \in \mathbb{N}$  and a dictionary  $\Phi$ , we formulate the following linear inverse problem with sparse constraint:

$$J_{\Phi, \mathbf{s}}(\mathbf{v}) \doteq \|\mathbf{s} - \Phi \mathbf{v}\|_2^2$$

$$\mathbf{v}(\Phi, \mathbf{s}, k) \doteq \underset{\mathbf{v} \in \mathbb{R}^M, \|\mathbf{v}\|_0 \leq k}{\operatorname{argmin}} J_{\Phi, \mathbf{s}}(\mathbf{v}). \quad (4)$$

We denote by  $\mathbf{c}_1 \doteq \mathbf{v}(\Phi_1, \mathbf{s}_j, k)$  and  $\mathbf{c}_2 \doteq \mathbf{v}(\Phi_2, \mathbf{s}_j, k)$  the solutions of (4) for  $\mathbf{s} = \mathbf{s}_j$ , where  $\mathbf{s}_j$  is the  $j^{th}$  column of  $S_{train}$ . In the same way, we denote by  $\mathbf{d}_1 \doteq \mathbf{v}(\Phi_1, \mathbf{s}_j, k)$  and  $\mathbf{d}_2 \doteq \mathbf{v}(\Phi_2, \mathbf{s}_j, k)$  the solutions of (4) for  $\mathbf{s} = \mathbf{s}_j$ , where  $\mathbf{s}_j$  is the  $j^{th}$  column of  $S_{test}$ . Finally, for the classification step, two different alternatives are evaluated in order to obtain an optimal performance of a MLP.

A Most Discriminative Activation Selection (MDAS) method selects the coefficients, related to the columns of a dictionary, that are “most involved” (activated) in the  $SaO_2$  signal recovery. The  $j^{th}$  column of a dictionary has an activation frequency  $\nu_{ci}^j$  given the *class i*, where  $\nu_{ci}^j$  makes reference to the number of times that the  $j^{th}$  column is used for *class i* signal recovery. Then, the candidates to be considered as input of the MLP are those coefficients ( $\mathbf{c}_1$  and  $\mathbf{c}_2$ ) with higher absolute difference  $D = |\nu_{c1}^j - \nu_{c2}^j|$ , between the activation frequencies for each of the classes. A subset of columns with higher values of  $D$  for each one of the classes is selected by the Most Discriminative Column Selection (MDCS) method. Those columns are used in order to construct new sub-dictionaries  $\Phi_1^1$  and  $\Phi_2^1$ . Then we denote by  $\mathbf{e}_1 \doteq \mathbf{v}(\Phi_1^1, \mathbf{s}_j, k)$  and  $\mathbf{e}_2 \doteq \mathbf{v}(\Phi_2^1, \mathbf{s}_j, k)$  the solutions of (4) for  $\mathbf{s} = \mathbf{s}_j$ , where  $\mathbf{s}_j$  is the  $j^{th}$  column of  $S_{train}$ . Finally the whole vectors  $\mathbf{e}_1$  and  $\mathbf{e}_2$  are taken into account as input for training the MLP. Figure 1 shows a block diagram of the system proposed for detecting AH events by applying the MDCS method. Two classification performance

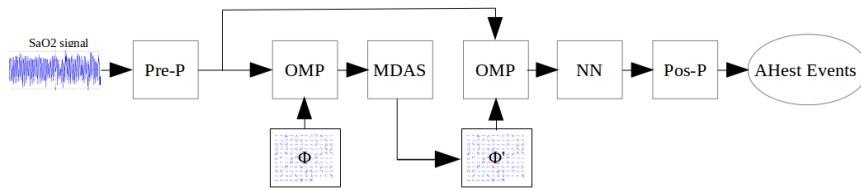


Figure 1: General diagram proposed for detecting AH events by applying MDCS method.

measures are used to compare both methods. The *sensitivity* (SE) is defined as the proportion of segments with AH events for which an event is present, while *specificity* (SP) is defined as the proportion of segments without AH events for which an event is not present. Also, a Receiver Operating Characteristics (ROC) [4] analysis is made and the Area Under the Curve (AUC) measure is reported.

## 4 RESULTS

In the results presented below we used a sparsity level  $k = 16$ . Also, for improving the performance of the classifiers, we took the same number of segments in each one of both classes. For both methods the optimization problem (4) was solved via the OMP algorithm ([8], [10]) due to its high efficiency and effectiveness. The MLP outputs were labeled as “1” or “0” depending on whether an AH event was detected

or not, respectively. Finally the estimated AHI (AHI<sub>est</sub>) was obtained by the total number of “1” divided by the time duration of each study (in hours).

Table 1 shows the sensitivity and specificity measures taking into account the AH events classifications as well as the corresponding AHI<sub>est</sub>-AHI correlation percentages. As seen in Table 1, significantly high correlation percentages for detecting AH events were obtained. No differences between the correlation percentages for MDAS-OAD and MDASI-OAD methods were observed. The correlation percentage and the AUC value obtained by MDCS-CD were significantly higher than those of the MDAS-CD.

Table 1: Number of inputs and neurons in hidden layer used for training the MLP and performance measures obtained.

Method	# I	# NHD	SE	SP	Corr. coef.	AUC
MDAS-OAD	24	14	74,52%	76,73%	90,04%	0,9799
MDCS-OAD	24	14	62,04%	64,73%	90,03%	0,9781
MDAS-CD	30	14	68,86%	67,69%	74,57%	0,8958
MDCS-CD	30	14	63,07%	66,03%	89,84%	0,9736

## 5 CONCLUSIONS

The previous analysis shows that the sparse representation of an  $SaO_2$  signal is a suitable technique for detecting moderate or severe OSAHS. The algorithm NOCICA was found to be a very useful tool for learning dictionaries.

Although similar classification results were obtained by applying MDAS and MDCS methods, it is important to point out that the MDCS method always gives equal or better results in terms of correlation percentages and AUC values. Also, the MDCS method is faster than MDAS in terms of CPU time for testing purposes.

For future work we propose to analyze the impact of the sparsity level  $k$  on the diagnostic performance measures. Also, the number of inputs  $P$  and the threshold of the MLP output will be further studied. The performance of different types of classifiers will be analyzed.

## 6 ACKNOWLEDGEMENTS

This work was supported in part by Consejo Nacional de Investigaciones Científicas y Técnicas, CONICET, through PIP 2014-2016 Nro. 11220130100216-CO and PIP 2012-2014 Nro. 114 20110100284-KA4, by the Air Force Office of Scientific Research, AFOSR / SOARD, through Grant FA9550-14-1-0130 and by Universidad Nacional del Litoral through project CAI+D 501 201101 00519 LI.

## REFERENCES

- [1] ABDALLAH, S.A. *Towards music perception by redundancy reduction and unsupervised learning in probabilistic models*. Ph.D. Thesis, Department of Electronic Engineering, King’s College London, (2002).
- [2] DURÁN-CANTOLLA J. *et al Consenso Nacional sobre el síndrome de apneas-hipopneas del sueño*, Vol. 41, (2005), pp. 1-110.
- [3] HYVÄRINEN A. *Survey on independent component analysis*. Neural Computing Surveys, Vol. 2, (1999), pp. 94-128.
- [4] KUMAR, R. AND INDIRAYAN, A. *Receiver operating characteristic (ROC) curve for medical researchers*. Indian Pediatrics, Vol. 48, (2011), pp. 277-287.
- [5] LESTUSSI F.E., DI PERSIA L.E. AND MILONE D.H. *Comparison of on-line wavelet analysis and reconstruction: with application to ECG*. 5th International Conference on Bioinformatics and Biomedical Engineering, (iCBBE), (2011), pp. 1:4.
- [6] LEWICKI M.S. AND SEJNOWSKI T.J. *A probabilistic framework for the adaptation and comparison of image codes*. Optical Society of America, Vol. 16, (1999), pp. 1587-1601.
- [7] LEWICKI M.S. AND SEJNOWSKI T.J. *Learning overcomplete representations*. Advances in Neural Information Processing 10 (Proceedings NIPS’97), (1998), pp. 556-562.
- [8] MALLAT S.G. AND ZHANG Z. *Matching pursuits with time-frequency dictionaries*. IEEE Transactions on Signal Processing, Vol. 41, (1993), pp. 3397-3415.
- [9] OLSHAUSEN B.A. AND FIELD, D.J. *Vision and the coding of natural images*. American Scientist, Vol. 88, (2000), pp. 238-245.
- [10] PATI Y.C., REZAIIFAR R. AND KRISHNAPRASAD P.S. *Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition*. IEEE Asilomar Conference on Signals, Systems and Computers, Vol. 1, (1993), pp. 40-44.
- [11] ROLON R.E. *et al Most Discriminative Atom Selection for apnea-hypopnea events detection*. Anales del VI Congreso Latinoamericano de Ingeniería Biomédica, (2014), pp. 709:712.