35th Annual International Conference of the IEEE EMBS
Osaka, Japan, 3 - 7 July, 2013

Non-negative Matrix Factorization and Sparse Representation for Sleep Signal Classification

Mehrnaz Shokrollahi, *Student Member, IEEE*, Sridhar Krishnan, *Senior Member, IEEE*

***Abstract*— Real-life signals such as biomedical signals are non−stationary and random in their pattern, and cannot be characterized by any specific waveform or spectral content. Processing of these natural signals involves consideration of certain significant attributes such as their non-stationary behavior over time, scaling behavior, translation invariance. Due to their random behavior, the existing discriminative methods often fail to provide a reasonable quantification performance, thereby resulting in poor classification rates. In order to address this issue, there exists a need for defining a suitable theoretical framework for biomedical signals. We have proposed, a robust Time−Frequency Nonnegative Matrix Factorization (TF−NMF) framework that uses sparse representation for quantification of sleep signals. This scheme incorporates a novel feature extraction algorithm. For signals that are nonstationary in nature, the degree of sparsity is lower compared to the stationary signals. This results into poor classification accuracy. However our proposed approach has proven that using NMF as input to the sparse representation for classification will improve the discrimination performance. Overall, maximum cross-validation performance of 87.9% was obtained, using the leave−one−out (LOO) approach for sleep abnormality detection using EMG signals. Although the computational complexity of the proposed algorithm might be higher compared to the other similar methods, this TF−NMF based method shows great potential for quantification and localization of time varying signals.**

I. INTRODUCTION

One of the most important problems in many biomedical signal processing applications is to find the best representation of the data. Many type of signals specially real−life signals are non-stationary and contain uncertainty and variability in their pattern, and cannot be characterized by any specific waveform or spectral content. Processing of these natural signals involves consideration of certain significant attributes such as their non-stationary behavior over time, scaling behavior, translation invariance [1]. Due to their random behavior, the existing discriminative methods often fail to provide a reasonable quantification performance, thereby resulting in poor classification rates.

Reflecting the above problem, a number of studies have been proposed for designing a suitable theoretical framework for biomedical signals. One of these methods is to find a correlation between the time and frequency domains of a signal, in order to study both time and frequency aspects simultaneously. The time−frequency (TF) plane provides a 2D signal domain that reveals not only temporal information, such as energy, but also frequency trend over time. Therefore, it is suggested as the most suitable signal plane for analysis of real−world signals which are random over both frequency and time. William *et al.* [2] used TF to represent the event related potential (ERP) activities. Delorme *et al.* [3] used TFD of multi−channel EEG signals for visualization of the temporal dynamics of the brain activities and interactions. Although these researches are beneficial in the area of visualization of the event of interest, their performance depends heavily on the matrix decomposition [4]. In one approach the TFD is interpreted as a matrix decomposition, in which a decomposition is applied to the TF matrix [5]. However, a suitable TF representation provides non-negative TF values, in order to produce meaningful features for representation of time varying signals in biomedical applications. Although this approach successfully analyzes the non-stationarity of the signal, its performance depends heavily on the matrix decomposition. One of these decomposition methods is Non-Negative Matrix (NMF) decomposition. NMF decomposes the TF signal into two components in a way that one contains the spectral structures and the other contains the corresponding temporal location of each spectral structure [4]. These are used as the TF feature vectors. In general NMF is used for local feature extraction. This is very beneficial when dealing with long term signals such as sleep signals, in which the expert can locate the abnormal behavior without analyzing 8 hours long data.

In this paper we have proposed a new method using TF-NMF for selection of the decomposed signal based on the level of sparsity to overcome the problem of high dimensionality of the signal. Using this proposed method along with the sparse representation for classification (SRC), the accuracy of 87.9% has been obtained which is the highest in comparison with other similar methods when dealing with EMG signals in Sleep. More details about the techniques used are presented in the Methodology section. The block diagram of the proposed method is also shown in Fig 1.

II. METHODOLOGY

A. Wavelet Scalogram

Wavelet scalogram is based on wavelet decomposition where orthonormal basis functions with different scales are used to decompose the signal [6]:

$$V_{CWT}(t,s) = \frac{1}{\sqrt{s}} \int x(\tau) g(\frac{\tau - t}{s}) d\tau \quad (1)$$

where $g(\frac{t}{s})$ is the mother wavelet and s is the scaling parameter which corresponds to the size of each basis functions.

This work was supported in part by the Canada Research Chair program.
M. Shokrollahi and S. Krishnan are with the department of Electrical Engineering in Ryerson University, Toronto, ON, M5B 2K3, Canada (Tel: 416-979 5000, ext 2048, email: mshokrol@ee.ryerson.ca, krishnan@ee.ryerson.ca)

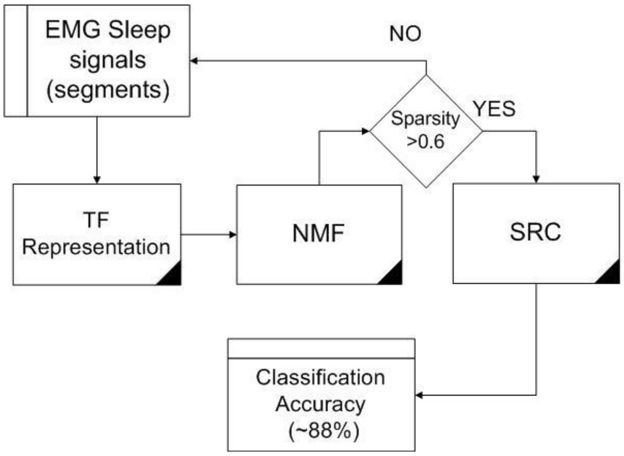

Fig. 1: Block Diagram of the Proposed method

We consider this TF transform, $V(t,s)$, as a matrix $V_{M\times N}$, where N is the number os samples in time and M is the frequency resolution of the constructed TF-transform. In general, wavelet scalogram displays the TF structure obtained from the wavelet transform. In scalogram, each wavelet signal is plotted as a filled rectangle whose its location and size are related to the time interval and the scale range for this wavelet signal. Additionally, scalogram provides a positive and cross-term free TF representation. Although this approach successfully analyzes the non$-$stationarity in the data, its performance heavily depends on the quality of the matrix decomposition (MD) technique. Therefore, we investigate the Nonnegative Matrix Decomposition (NMF) as related to TF quantification.

B. Nonnegative Matrix Factorization (NMF)

There are many matrix decomposition techniques available including Independent Component Analysis (ICA), Principal Component Analysis (PCA) and Non-negative matrix factorization (NMF). These have become an active area of research for extracting features in machine learning, computer vision, and signal processing [7]; however, depending on the application and the data, different results have been reported. Some works demonstrated the advantages of ICA to NMF and PCA. Lee *et al.* compared the result of PCA, NMF and ICA for feature extraction from multiple video frames [8] and showed that ICA-based features result in better video representation than the PCA or NMF-based features. In another study, the authors show that NMF is more stable for larger basis compared to ICA and PCA [9].

In general once the signal is transformed to the TF plane, if the TF representation is of the form $V_{M\times N}$ the decomposed signal is $V_{M\times N} = W_{M\times r}H_{r\times N} = \sum_{i=1}^{r} w_i h_i$, where V is the TF matrix, W is the base and H is the coefficient vectors and r is the rank of the NMF Decomposition. This decomposition is in a way that W represents the signal structure and the H represents the location of the corresponding base vectors in time. With respect to the application used, NMF compared to ICA and PCA has more advantages which as listed below:

- NMF is applied to a nonnegative matrix and constrains the matrix factors W and H that are also nonnegative,
- NMF decomposed factors promise a higher TF representation and localization,
- NMF codes naturally favor sparse, parts-based representations which in the context of recognition can be more robust than non-sparse, global features [7].

The problem with this technique is that the extracted feature vectors have a very high dimension. This is because the length of each feature vector is proportional to the signal's sampling frequency, and as a result they are not very appealing for classification. However, some variation of NMF is subject to additional constraints that allow particularly accurate control over sparseness and, indirectly, over the localization of features. Our proposed approach for this dimensionality reduction is first representing the signal in a linear subspace using sparse representation. Although a part based representation such as NMF is based on the low$-$rank approximation, the role of sparse representation here is simplifying the signal to its most meaningful components [10]. In order to achieve the best separability, one need to modify the sparse approximation methods such that the objective function is enhanced with a discrimination term that represents the separability properties of the signals. In this paper we have analyzed the bases signal, W, from the decomposed matrix using the sparsity measure.

C. Sparsity Measure

In signal representation, the definition for sparsity is given in many different ways. The definition used for sparsity in this work is one in which energy of the signal is concentrated in small number of non-zero coefficients. In general, the measure of sparsity depends on the relative distribution of energy among the coefficients, and should not solely be calculated on the absolute value of each coefficient. As a matter of fact, a good measure of sparsity has to be a weighted sum of coefficients based on the importance of a particular coefficient in the overall sparsity [11].

In most of the current literature, sparsity is measured using the ℓ_0 norm of a vector. If the original signal X has N samples in the spatial domain and has M number of zeros in it, it is said to be K-sparse, where $K = N - M$. When X has no nonzero elements but has M elements which have magnitudes which are small, then the above definition can be extended to "approximate K-sparsity" [11].

One of the measures of sparsity that is analyzed in this paper is as follows [12]:

$$S_p(x) = (\frac{1}{C_p})(\frac{(\frac{1}{n-1})\sum_{i=1}^{n}|x_i - m|^p}{\frac{1}{n}\sum_{i=1}^{n}|x_i|^p})^{\frac{1}{p}} \tag{2}$$

where

$$m = (\frac{1}{n})\sum_{i=1}^{n} x_i \tag{3}$$

and

$$C_p = (\frac{(n-1)^{p-1}+1}{n^{p-1}})^{\frac{1}{p}} \tag{4}$$

The default value for p is chosen to be one and n is the length of the spatial signal. The advantage of this sparsity measure is that is it based on the normalized value compared to the other conventional methods that are based on the convectional norm measures. The value that the $S_p(x)$ takes is between 0 and 1 for any vector. 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. In other words, the sparsity is zero if all the coefficients have equal amount of energy and is one if the concentration of energy is only on one value. This gives us a meaningful measure, exhibiting the sparsity of the distribution.

D. Sparse Representation for Classification

Following the introduction of sparse theory, many different approximation tools have been developed for different tasks. Choosing a representation that offers the most compact solution for further analysis and/or decision making is the main principle behind these techniques [13]. Recent research in sparse approximation involves obtaining optimal signal approximation for signal discrimination or signal reconstruction. Hence, the sparsity of different types of signals have not been fully analyzed for these type of application. The algorithm that was used in this work is briefly as follows [12]: let A be the matrix and each column of which is the training samples from the i^{th} class. The elements of this matrix is a set of feature vectors that have been produced from transforming the signal into TF plane and from there to an NMF algorithm with a rank of $r = 50$. y is the new testing samples, in which a linear representation of y can be rewritten in terms of all training samples as $y = Ax$. If the representation can be converted into finding a column vector x such that $y = Ax$ and most of the values of x is zero then the problem is so called sparse representation. We use cross validation to randomly split the data into distinct matrix for training and testing. Once the system is trained with the training signals we test it using the ℓ_1 minimization problem for sparse representation as in Eq 5

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

The leave$-$M$-$out (LMO) cross validation, validates the correctness of the results (for special case we only use LOO), when only M number of samples are left out and the output is the average of correctly identified.

III. EXPERIMENTAL RESULTS

A. Algorithm

In this paper, we presented a method for classifying nonstationary time series utilizing a feature set based on time frequency representation followed by nonnegative factorization. This was then further analyzed using sparsity measure and sparse representation for classification. The proposed scheme for real biomedical signals analysis and discrimination can be briefly characterized by the following steps and the block diagram of the proposed method is given in Fig1:

Preprocessing: In this study we have utilized the sleep dataset that was provided to us from the SunnyBrook Health Science Center. The sleep dataset consists of signal segments from 8 chin EMG signals (4 with normal behavior and 4 with Rapid eye movement Behavior Disorder (RBD)) undergone the sleep test. A traditional scoring system for sleep has been established [14], 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. 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.

Higher Dimensional Mapping:The preprocessed signal is then mapped to the TF representation , using the wavelet scalogram, and a transformed signal have been obtained. The frequency resolution of this transformed function is 150 (scale is from $1 : 150$) and the time samples are 1000 elements. The advantage of using wavelet scalogram is that it uses an adaptive varying time width defined by the scaling parameter. Another important advantage of scalogram is that it provides positive and cross-term free TF representation.

Feature Extraction: The transformed signal matrix is then decomposed into two matrices, W the bases matrix and H the coefficient matrix. In this work we have used W as the feature vector since it is good for local feature extraction and classification. The rank of the matrix is chosen to be $r = 50$, as the results show that NMF yields the lowest recognition rate for decomposition dimensions of less than 50 ($r < 50$). Because of the high dimensionality of the NMF, we have proposed to use the degree of sparseness. Sparse non-negative matrix factorizations (NMFs) are useful when the degree of sparseness in the non-negative basis matrix or the non-negative coefficient matrix in an NMF needs to be controlled in approximating high-dimensional data in a lower dimensional space.

Classification: Based on the feature space the normal and abnormal RBD classes are defined. The feature vectors along with the class definition together are fed into sparse representation for classification. In this step the influence of both training and testing sample size on the design and performance of pattern recognition systems has been investigated by using sparse representation incorporated with LOO approach. This method is believed to be one of the most optimized validation approaches with least biased estimate [15]. The presentation of the underlying theory has been complemented with examples with real time sleep signals and results on the application of these data are validated based on the degree of sparsity measure and classification accuracy.

For comparison, we have compared our proposed method with the other previous methods [12][16]. This have been shown in Fig. 2. Although our proposed method is unstable

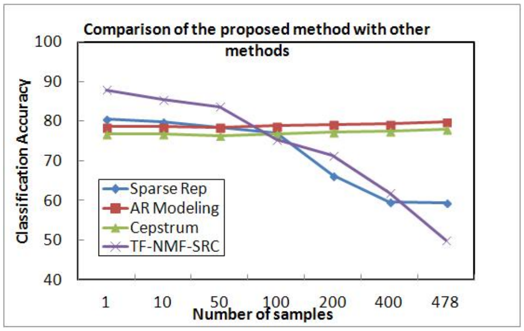

Fig. 2: Block Diagaram of the Proposed method

Signal	$M = 1$	$M = 50\%$	**Sparsity**
Sleep	87.9%	50%	0.73

TABLE I: Result of classification performance of the proposed scheme

when leaving half of the samples out ($M = 50\%$), but it shows a better classification accuracy for LOO.

IV. Discussion

The REM EMG signal was first randomly segmented to smaller segments. Each of these segments were then fed into a TF plane using Daubechies wavelet scalogram. This 2D representation was then used as the input of the NMF decomposition. The bases vector of the NMF, W was then used as feature vectors. The extracted feature vectors were used to discriminate the REM sleep into Normal and RBD classes. However, before using the SRC classification algorithm, the sparsity of the feature vectors were measured. If the sparsity of the feature vector was below some threshold (60%), the respective segment was discarded for classification and the next segment was used. Thereby increasing the classification accuracy. After the respected segments were found they were used as the input to SRC algorithm. The results show that the accuracy decreases drastically as the value of M increases. Therefore, the highest accuracy is achieved when the sparsity is high as well as when we use the LOO approach. Using our proposed scheme a maximum overall accuracy of 87.9% (using LOO) is achieved.

From the results, it can be verified that the TF$-$NMF based analysis offers new insight to suitably represent a nonstationary signal in the joint time-frequency space using sparse representation. Also it is evident that the proposed feature extraction succeeds in discriminating the normal from the abnormal REM sleep signals using the LOO approach. Table I illustrates the details of the results.

V. Conclusion

In this study we have presented a robust TF$-$NMF based sleep quantification scheme using sparse representation for classification that incorporates sparsity measures and novel feature analysis algorithm. For signals that are nonstationary in nature, the degree of sparsity is lower compared to the stationary signals. This result into poor classification accuracy. However our proposed approach has proven that using NMF as input to the sparse representation for classification will improve the discrimination performance. Overall, maximum cross-validation performance of 87.9% was obtained, using the LOO approach. Although the computational complexity of the proposed algorithm might be higher compared to the other similar methods, this TF$-$NMF based method shows great potential for analysis and sparse quantification of the time$-$varying signals. Despite of the fact that this is a preliminary studies that employ TF-NMF analysis using SRC, for sleep signal quantification, a proper TFD would be eventually desirable for accurate feature localization applications.

References

[1] S.S. Mahmoud, Z.M. Hussain, I. Cosic, and Q. Fang, "Time-frequency analysis of normal and abnormal biological signals," *Biomedical Signal Processing and Control*, vol. 1, no. 1, pp. 33–43, 2006.

[2] E.M. Bernat, W.J. Williams, W.J. Gehring, et al., "Decomposing ERP time–frequency energy using PCA," *Clinical Neurophysiology*, vol. 116, no. 6, pp. 1314–1334, 2005.

[3] A. Delorme, S. Makeig, M. Fabre-Thorpe, and T. Sejnowski, "From single-trial eeg to brain area dynamics," *Neurocomputing*, vol. 44, pp. 1057–1064, 2002.

[4] B. Ghoraani and S. Krishnan, "A joint time-frequency and matrix decomposition feature extraction methodology for pathological voice classification," *EURASIP Journal on Advances in Signal Processing*, vol. 2009, pp. 10, 2009.

[5] K. Englehart, B. Hudgins, P.A. Parker, and M. Stevenson, "Classification of the myoelectric signal using time-frequency based representations," *Medical engineering & physics*, vol. 21, no. 6, pp. 431–438, 1999.

[6] F. Auger, P. Flandrin, P. Gonçalvès, and O. Lemoine, "Time-frequency toolbox," *CNRS France-Rice University*, 1996.

[7] P.O. Hoyer, "Non-negative matrix factorization with sparseness constraints," *The Journal of Machine Learning Research*, vol. 5, pp. 1457–1469, 2004.

[8] M.J. Lee, A.S. Lee, D.K. Lee, and S.Y. Lee, "Video representation with dynamic features from multi-frame frame-difference images," in *Motion and Video Computing, 2007. WMVC'07. IEEE Workshop on*. IEEE, 2007, pp. 28–28.

[9] N.F. Chikhi, B. Rothenburger, and N. Aussenac-Gilles, "A comparison of dimensionality reduction techniques for web structure mining," in *Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence*. IEEE Computer Society, 2007, pp. 116–119.

[10] I. Tosic and P. Frossard, "Dictionary learning," *Signal Processing Magazine, IEEE*, vol. 28, no. 2, pp. 27–38, 2011.

[11] D. Zonoobi, A.A. Kassim, and Y.V. Venkatesh, "Gini index as sparsity measure for signal reconstruction from compressive samples," *Selected Topics in Signal Processing, IEEE Journal of*, vol. 5, no. 5, pp. 927–932, 2011.

[12] M. Shokrollahi and S. Krishnan, "Sleep EMG analysis using sparse signal representation and classification," in *Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE*. IEEE, 2012, pp. 3480–3483.

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

[14] 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.

[15] 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.

[16] M. Shokrollahi, S. Krishnan, D. Jewell, and B. Murray, "Analysis of the electromyogram of rapid eye movement sleep using wavelet techniques," in *Engineering in Medicine and Biology Society, 2009. EMBC 2009. Annual International Conference of the IEEE*. IEEE, 2009, pp. 2659–2662.