Group learning using contrast NMF : Application to functional and structural MRI of schizophrenia

Vamsi K. Potluru
MIND Research Network
Department of Computer Science
University of New Mexico
Albuquerque, New Mexico 87131
Email: ismav@cs.unm.edu

Vince D. Calhoun
MIND Research Network
Dept. of Electrical & Computer Engg
University of New Mexico
Albuquerque, New Mexico 87131
Email: vcalhoun@unm.edu

***Abstract*— Non-negative Matrix factorization (NMF) has increasingly been used as a tool in signal processing in the last couple of years. NMF, like independent component analysis (ICA) is useful for decomposing high dimensional data sets into a lower dimensional space. Here, we use NMF to learn the features of both structural and functional magnetic resonance imaging (sMRI/fMRI) data. NMF can be applied to perform group analysis of imaging data and we apply it to learn the spatial patterns which linearly covary among subjects for both sMRI and fMRI. We add an additional contrast term to NMF (called co-NMF) to identify features distinctive between two groups. We apply our approach to a dataset consisting of schizophrenia patients and healthy controls. The results from co-NMF make sense in light of expectations and are improved compared to the NMF results. Our method is general and may prove to be a useful tool for identifying differences between multiple groups.**

I. Introduction

Non-negative matrix factorization (NMF) is a tool to split the given data matrix into a product of two non-negative matrix factors. This process can be used to identify useful features in the dataset. Another tool used commonly to find features is independent component analysis (ICA) [1]. ICA assumes that the features found are statistically independent [2]. We will explore the potential of NMF to model spatio-temporal features across subjects. The use of NMF is attractive for analysis as it naturally incorporates the non-negativeness of the images and is quite easily adapted to learn across subjects. Also, the constraint of independence among the learnt features is not required for NMF . This strong assumption of ICA is thus relaxed for NMF. Adding different constraints to NMF is relatively straightforward as can be seen by the wide range of NMF extensions. For example, [3] extends plain NMF by having the learnt factors to be convolutive instead of being bilinear and [4] adds a sparsity constraint on one of the learnt factors.

Our main motivation is to find distinctive features by group analysis of MRI data using NMF. Its extension, non-negative tensor factorization (NTF) [5], is a rich framework to model spatio-temporal patterns across subjects . NMF algorithms have been previously applied to fMRI data in [6]. They use it to find a suitable representation of the data to detect task-related brain activations. Our goal is to identify features distinctive between healthy controls and patients who have been diagnosed with schizophrenia using a version of NMF where we explicitly optimize for this. We call this approach contrast NMF or co-NMF.

Schizophrenia is a disease that involves disruption of a variety of cognitive functions such as memory, perception, executive function and emotion. It can be characterized by disturbances in thought, disorganized speech with poor content, delusions, hallucinations, impairment of personal/occupational relationships, poor self-care and impersistence at work. Currently, diagnoses for major psychiatric disorders like schizophrenia are based solely on clinical manifestations and observed psycho-social impairments. Biological indices, if they can be discovered, would be beneficial in providing more objective methods of classification.

In the following sections we present NMF and co-NMF along with their corresponding update rules. Then, we apply the algorithms on both sMRI and fMRI datasets. In the discussion section, we summarize the results and future work to be done.

II. NMF model

Given a non-negative matrix $\mathbf{X}$ of size $M \times N$, the task is to split it into a product of two non-negative matrices $\mathbf{W} \in R^{\geq 0, M \times R}$ and $\mathbf{H} \in R^{\geq 0, R \times N}$. That is,

$$\mathbf{X} \approx \mathbf{WH} \tag{1}$$

Lee and Seung [7] gave two different update rules for miniziming $\|\mathbf{X} - \mathbf{WH}\|$ corresponding to two different cost functions. Here, we use the Euclidean distance as the cost function. The function and its update are given by:

$$E = \|\mathbf{X} - \mathbf{WH}\|_F \tag{2}$$

$$\mathbf{W} = \mathbf{W} \otimes \frac{\mathbf{XH}^\top}{\mathbf{WHH}^\top} \tag{3}$$

$$\mathbf{H} = \mathbf{H} \otimes \frac{\mathbf{W}^\top\mathbf{X}}{\mathbf{W}^\top\mathbf{WH}} \tag{4}$$

Here, $\otimes$ is used to represent element-wise multiplication. Also, the division in the above equations is element-wise. It should be noted that the cost function to be minimized is convex in either $\mathbf{W}$ or $\mathbf{H}$ but not in both. The rank parameter R on which the sizes of both the matrices $\mathbf{W}, \mathbf{H}$ depend is usually based on prior knowledge of the data being decomposed. In practice, the matrices $\mathbf{W}, \mathbf{H}$ are initialized to positive random matrices.

III. co-NMF

We now introduce a variation of standard NMF. The motivation comes from a brain imaging problem where we are given images of two groups namely healthy controls and patients with schizophrenia. NMF can be used to represent the data. However, we are looking for features that not only represent the data but also maximize the difference between the two groups thereby identifying the differences. This is accomplished by introducing an additional constraint that the difference of mean activations of two groups in the learnt factor is maximized.

This is given by the following objective :

$$\min_{\mathbf{W},\mathbf{H}_1,\mathbf{H}_2} \frac{1}{2}\|\mathbf{X}_1 - \mathbf{WH}_1\|_F^2 + \frac{1}{2}\|\mathbf{X}_2 - \mathbf{WH}_2\|_F^2 - \|\lambda \otimes (\mu(\mathbf{H}_1) - \mu(\mathbf{H}_2))\|_1 \tag{5}$$

The matrices $\mathbf{X}_1, \mathbf{X}_2$ correspond to the observation data from two groups, $\mathbf{W}$ is the common feature space of the groups and $\mathbf{H}_1, \mathbf{H}_2$ the corresponding activations. The function μ computes the mean activation by taking a matrix of column vectors and producing a single column vector by averaging across the entire group.

We use the following update equations to minimize the objective:

$$\mathbf{W} = \mathbf{W} - \eta[-\mathbf{X}_1\mathbf{H}_1^\top - \mathbf{X}_2\mathbf{H}_2^\top + \mathbf{WH}_1\mathbf{H}_1^\top + \mathbf{WH}_2\mathbf{H}_2^\top] \tag{6}$$

$$\mathbf{H}_1 = \mathbf{H}_1 \otimes \frac{\mathbf{W}^\top\mathbf{X}_1 + ((\lambda \otimes \mathbf{d})\mathbf{1})^-}{\mathbf{W}^\top\mathbf{WH}_1 + ((\lambda \otimes \mathbf{d})\mathbf{1})^+} \tag{7}$$

$$\mathbf{H}_2 = \mathbf{H}_2 \otimes \frac{\mathbf{W}^\top\mathbf{X}_2 + ((\lambda \otimes \mathbf{d})\mathbf{1})^+}{\mathbf{W}^\top\mathbf{WH}_2 + ((\lambda \otimes \mathbf{d})\mathbf{1})^-} \tag{8}$$

where $\mathbf{1}$ is row vector of ones of appropriate dimension, λ is the weight vector and $\mathbf{d}$ is the mean(along columns) difference of matrices $\mathbf{H}_1, \mathbf{H}_2$. We use $\pm$ in the superscript to denote the absolute values of positive and negative elements of matrix with the rest set to zero respectively. As has already been noted in [4], the objective function in our case is also not scale free. Therefore, we use gradient descent for updating $\mathbf{W}$ and then rescale its column vectors to norm unity. We note that multiplicative update rules for $\mathbf{W}$ could have been employed as in [8].

IV. Methods

We now apply the NMF and co-NMF algorithms on structural and functional MRI datasets.

A. Group analysis

Group analysis of fMRI is important to study specific clinical or experimental conditions within or between groups of subjects. In [9] , a method to do group ICA was introduced and has been implemented in a user-friendly environment (GIFT). It can be found at http://icatb.sourceforge.net. GIFT has already been used to study structural MRI datasets [10] . There has to date been no application of NMF to group brain imaging data.

Group analysis using NMF is done as follows. We stack the preprocessed datasets(one image per subject) for each subject column wise to get the new dataset. This gives us a common image of gray matter concentration or activation for each subject.

In the following experiments, $\mathbf{X}_1$ and $\mathbf{X}_2$ correspond to the set of images from healthy controls and patients. $\mathbf{H}_1, \mathbf{H}_2$ represennt their corresponding activations and $\mathbf{W}$ the learnt features.

B. Structural MRI

Structural MRI scans of 136 healthy controls and 133 schizophrenia patients were taken at the John Hopkins University. The scans were taken on a single $1.5T$ scanner with the imaging parameters($35mm$ TR , $5ms$ TE , matrix size of 256×256) . We segment these images into gray matter, white matter and cerebral spinal fluid images, using the the software program SPM5 (http://www.fil.ion.ucl.ac.uk/spm/software/spm5/), followed by spatial smoothing with a Gaussian kernel

of $10 \times 10 \times 10$ mm. This results in images which are $105 \times 127 \times 46$, with each voxel containing a value between 0 and 1. We use the gray matter concentration images to create the matrices X_1 and X_2 from the resulting 136 and 133 segmented images each of which is $105 \times 127 \times 69$. We then analyzed this dataset using both NMF and co-NMF. We plot sample images and plot the feature which have the maximum difference between the control and patients. The number of components extracted in both cases is 9.

C. Functional MRI

We utilized the MIND clinical imaging consortium research data to analyze differences in brain activity for first episode and chronic schizophrenia patients versus healthy controls. The analysis involves a dataset of fMRI images totaling 74 consisting of 38 healthy controls and 36 patients with schizophrenia. Each test subject underwent an fMRI scan on a Siemens 1.5T Sonata scanner while performing an auditory 'oddball' task. The auditory stimuli consisted of a standard tone ($1000Hz$), a target tone ($1200Hz$) and novel complex complex generated tones. fMRI data were preprocessed with the SPM5 software. Data for each subject were analyzed by multiple regression incorporating regressors for the novel, target and standard stimuli and their temporal derivatives plus an intercept term. Regressors were created by modeling the stimuli as delta functions convolved with the default SPM hemodynamic response function. Only correct responses were modeled. The target versus standard contrast image was used in the NMF analyses. We create the matrices X_1 and X_2 from the 38 healthy controls and 36 patients images respectively, each with dimension $53 \times 63 \times 46$.

D. Preprocessing

The datasets of both structural and functional MRI images were processed in a similar manner as follows. A brain mask was created by selecting only those voxels with value greater than 0.1. For the structural images, the average across subjects was taken and removed from the images. The images were then kept non-negative by subtracting the least value among them. A brain mask was created by selecting only those voxels with value greater than 0.1. The parameter λ was arbitrarily chosen to be 0.04.

For the functional images, brain mask was limited to in-brain voxels and provided with SPM software. Also, the mean was not removed though the least value among them was subtracted across all subjects. Each individual subject has the same preprocessing done separately.

V. Results

We now present the results from both the sMRI and fMRI datasets.

A. Structural MRI

We now show the feature which has the largest mean difference between patients and controls as found by both co-NMF and NMF.

There are a couple important points to note about the results. First, as shown in Figure 3 (left) the difference between patients and controls is larger for the co-NMF results, as expected. Also, the difference appears to be evenly spread in all the features. In Figure 1, both algorithms are finding regions in frontal and temporal lobe (known to be affected in schizophrenia). However the co-NMF results shows much more temporal lobe area and also has a maximum in the anterior cingulate, another regions often implicated in schizophrenia. These results are encouraging and suggest that co-NMF is more sensitive to the expected difference in controls versus patients.

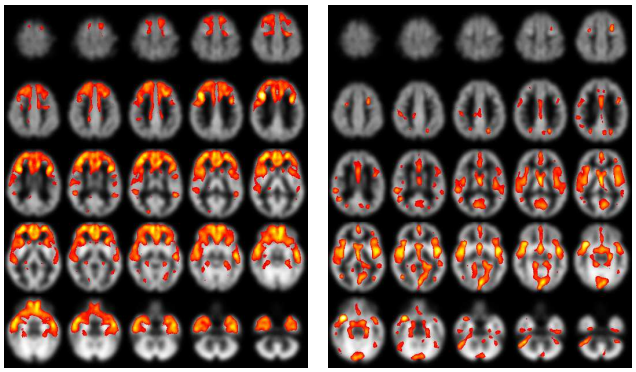

Fig. 1. Left corresponds to NMF applied to sMRI data and right corresponds to co-NMF and in both we plot the feature which has the maximum mean difference of activation between controls and patients

B. Functional MRI

We do the same for the fMRI dataset i.e. the feature with the largest mean difference between the groups.

For the fMRI results, as shown in Figure 3, the difference between patients and controls are plotted. Figure 2 again show more regions concentrated in the temporal lobe area for co-NMF compared to NMF.

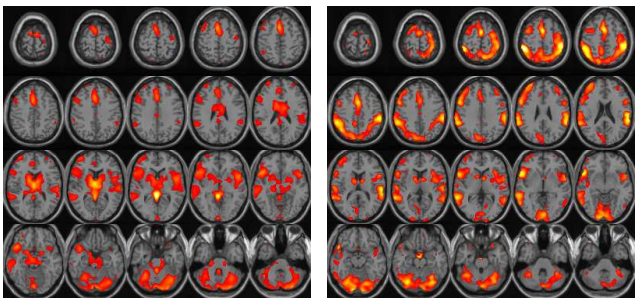

Fig. 2. NMF feature which shows most difference between controls and patients(left) and corresponding co-NMF learnt feature (right)

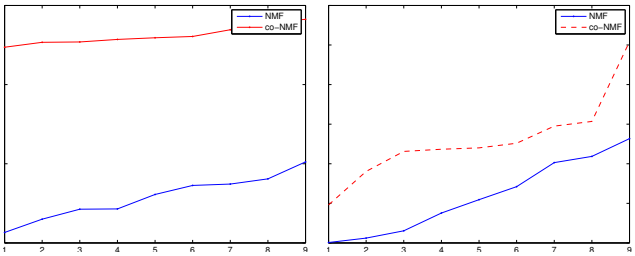

Fig. 3. Comparision of difference of mean activations of 9 features using NMF and co-NMF ,where the left figure is that of sMRI data and the right to that of fMRI data.

The difference between groups was more significant for co-NMF than NMF for both sMRI and fMRI.

We did a t-test comparing the mean values of the loading parameters for controls and patients. The p value for the feature identified by co-NMF is $1.22e-15$ and that for NMF is $6.91e-07$.

VI. Classification

We also applied a classification method to the activations learnt by both co-NMF and NMF. Since, the dataset is small, we used the leave-one-out method. NMF and co-NMF were applied on the fMRI dataset to obtain the corresponding activations. R is chosen as before to be 9. Classification rate of 55% and 68% were obtained respectively. Nearest neighbours algorithm was used for classification. We choose the vote based on 5 nearest neigbours.

VII. Discussion and Future work

We have shown NMF to be an effective tool for use in both sMRI and fMRI data analysis. In this paper, we focused upon the difference between NMF and co-NMF which was adapted to maximize the patient versus control differences. NMF is easily adapted to add additional constraints and not be restricted to the components being independent.

We also relax the constraint that the features learnt are independent as is required by ICA. It is possible that you have multiple features with statistical dependency that account for the data in physically meaningful manner. In contrast, ICA assumes spatial independence for the learnt sources . The other advantage of NMF is that it assumes the data is an additive mixture of the features which corresponds to different parts of the brain activating. Drawing a similar analysis with ICA is not straightforward.

In future work, it would be interesting to compare NMF with the more widely used ICA. We would also like to model fMRI data time series using non-negative tensor factorization (NTF) which would allow for a richer set of parameters and incorporation of prior information into the modeling process. The choice of λ and factorization rank of $\mathbf{W}$ matrix from data is another direction for research.

References

[1] P. Comon, "Independent component analysis: A new concept," *Signal Processing*, vol. 36, pp. 287–314, 1994.

[2] Anthony J. Bell and Terrence J. Sejnowski, "An information-maximization approach to blind separation and blind deconvolution," *Neu. Comp.*, vol. 7, no. 6, pp. 1129–59, 1995.

[3] Paris Smaragdis, "Non-negative matrix factor deconvolution; extraction of multiple sound sources from monophonic inputs," in *Fifth International Conference on Independent Component Analysis*, Granada, Spain, Sept. 22–24 2004, LNCS 3195, pp. 494–9, Springer-Verlag.

[4] Patrik O. Hoyer, "Non-negative sparse coding," in *IEEE Workshop on Neural Networks for Signal Processing*, 2002.

[5] A.Cichocki, R.Zdunek, S.Choi, R.Plemmons, and S.Amari, "Non-negative tensor factorization using alpha and beta divergences," *International Conference on Acoustics,Speech and Signal Processing (ICASSP 07), Honolulu,Hawaii,USA*, 2007.

[6] X. Wang, J. Tian, X. Li, J. Dai, and L. Ai, "Detecting brain activations by constrained non-negative matrix factorization from task-related BOLD fMRI," in *Medical Imaging 2004: Physiology, Function, and Structure from Medical Images. Edited by Amini, Amir A.; Manduca, Armando. Proceedings of the SPIE, Volume 5369, pp. 675-682 (2004).*, A. A. Amini and A. Manduca, Eds., Apr. 2004, vol. 5369 of *Presented at the Society of Photo-Optical Instrumentation Engineers (SPIE) Conference*, pp. 675–682.

[7] Daniel D. Lee and H. Sebastian Seung, "Learning the parts of objects with nonnegative matric factorization," *Nature*, vol. 401, pp. 788–91, 1999.

[8] Julian Eggert and Edgar Körner, "Sparse coding and NMF," in *IEEE International Joint Conference on Neural Networks, 2004. Proceedings*. IEEE, July 2004, vol. 4, pp. 2529–33.

[9] V. D. Calhoun and T.Adali, "Unmixing functional magnetic resonance imaging with independent component analysis," *IEEE Eng. in Medicine and Biology*, vol. 25, pp. 79–90, 2006.

[10] V. D. Calhoun, T.Adali, N. Giulaiani, G.D. Pearlson J.J. Pekar, and K. A. Kiehl, "A method for multimodal analysis of independent source differences in schizophrenia: Combining gray matter structural and auditory oddball functional data," *Hum.Brain Map.*, vol. 27, pp. 47–62, 2006.