SEPARATING SPATIAL AND TEMPORAL ACTIVATION PATTERNS IN FMRI USING COMPETITIVE SUBSPACE PROJECTION*

Rui Yan[1], *Guojun He*[2], *Deniz Erdogmus*[3], *Sung-Phil Kim*[1], *Jose C. Principe*[1], *Yijun Liu*[2]

[1]Dept. of ECE, University of Florida, Gainesville, FL 32611, USA
[2]Dept. of Psychiatry, University of Florida, Gainesville, FL 32611, USA
[3]Dept. of CSE and BE, OGI School of Engineering and Science, OHSU, Portland, OR 97006, USA

ABSTRACT

The challenge for functional Magnetic Resonance Imaging (fMRI) is to determine when and where the response occurs in the images due to the external stimulus. The temporal clustering analysis (TCA) method has been used to study the brain activities after eating and drinking both in time and spatial domains. In this paper, we propose a new method, competitive subspace projection (CSP) to optimally represent data compared to other subspace projection methods. This method is used to detect spatial and temporal activation patterns in fMRI associated with such behavior. The CSP and TCA methods are compared using both synthetic and real fMRI data. The results on both data sets show consistent conclusions can be drawn from these two methods while CSP is observed to have better noise rejection capability than TCA.

1. INTRODUCTION

The rapid development of functional magnetic resonance imaging (fMRI) techniques allows dynamic mapping of the brain processes with finer spatial resolution. However, the challenges remain in localizing brain function when there is no a priori knowledge available about the time a stimulus may elicit response. In addition, the fMRI signal is contaminated by high level of noise, especially for non-repeatable physiological events or relatively long events (compared to cognitive processes) in the brain, such as those following eating and drinking. Liu *et al.* proposed temporal clustering analysis (TCA) to quantify absorption of food in brain imaging [1]. Still, new methods are required for dealing with these challenges.

The subspace projection method has its advantage of data compression and noise cancellation. Now it is widely implemented in image processing, e.g., image compression [2] and hyperspectral image classification [3]. The optimal subspace projection method in terms of preserving energy is principal component analysis (PCA) [4]. However, in many cases global PCA is not optimal due to the complicated data distribution. Simultaneously competitive learning is known by its powerful local feature extraction. It is first used in supervised learning scheme [5]. Then vector quantification is combined into PCA for the unsupervised learning strategy [6],[7]. Haykin *et al.* incorporates unsupervised competitive PCA network and proposes a OIAL (optimally integrated adaptive learning) method, which gives lower MSE and higher compression ratio [8]. However, OIAL doesn't take the bias effect among models into consideration, which leads to suboptimal results. Fancourt *et al.* combine the mixture of experts and PCA into a cooperative network to segment time series and images [9]. However, image segmentation is done in a winner-take-all fashion. Cooperative network assigns multiple probabilities for one input (multiple experts contribute to the input based on the posterior probabilities), which makes the classification unnecessarily complex. We propose in this paper a hard competition method, named competitive subspace projection (CSP) method, to optimally project data into subspace and incorporate this method into fMRI analysis. This method performs well since it bridges both the pixel time distribution and spatial distribution and its advantage is that it doesn't need any prior information of time course modeling. The unsupervised vector space representation optimally clusters vectors of time series, which gives optimal spatial task-oriented segmentation. This segmentation is uncorrelated with image content and has a good noise rejection performance.

Specifically, this paper deals with segmenting the activation region in terms of their temporal response given a single stimulus. As a first step, the existing temporal clustering analysis method is reviewed and analyzed in a probabilistic point of view. The competitive subspace projection method, which optimally represents data space, is configured to the alcohol detection fMRI problem. The effectiveness and performance of the proposed method are verified on experiments using both synthetic data and alcohol ingestion fMRI images.

*THIS WORK WAS PARTIALLY SUPPORTED BY THE NSF GRANT ECS-0300340, AND ABMRF AND NIH/DA16221 (TO YL).

2. METHODS

2.1. Temporal Clustering Analysis (TCA)

The proposed TCA method effectively extracts the statistical properties from a 3-dimensional data space (the 2-dimensional spatial image plus the time dimension, considering one brain slice only) and forms a probabilistic sequence over time where each element $N_{max}(n)$ of the sequence represents the number of pixels which reach its maximum throughout the time series. Given the fMRI image of size $M \times N$ at discrete time n, where $n = 1, \cdots, L$, and the pixel value $\rho_{i,j}(n)$ at instant n with $i = 1, \cdots, M$ and $j = 1, \dots, N$, the temporal maxima response $N_{max}(n)$ can be written as

$$N_{max}(n) = \sum_{i=1}^{M} \sum_{j=1}^{N} f(\rho_{i,j}(n)) \tag{1}$$

where $f(\rho_{i,j}(n)) = 1$ if $\rho_{i,j}(n) \geq \rho_{i,j}(n^*), \forall n^*, n^* \neq n$; 0 otherwise. This method implicitly assigns probability

$P(i, j, \hat{n}) = 1$ to pixel (i, j) at the peaking time of $\hat{n}$ and probability $P(i, j, n) = 0$ for all other time instants considering the time course as a probability distribution function distribution; essentially, this is a delta function approximation for each pixel. Next, for each time instant, the probabilistic masses of every pixel at the time instant being considered is summed to obtain the temporal maxima response $N_{max}(n) = \sum_{i=1}^{M} \sum_{j=1}^{N} P(i, j, n)$. This quantity is a measure of activation strength (possibly due to a common cause) since it will take large values when many pixels are activated simultaneously. The large value of $N_{max}(n)$ at time t is considered to be the voxel response by the stimulus, and the noise effect is assumed to be wide sense stationary (WSS) in time and space. The voxels with large values are called the region of interest (ROI).

This method has been successfully applied to mapping the brain activities following glucose ingestion. It provides a deterministic analytical solution with straightforward computations. However, there still exist room for improvement for this method. Firstly, it is strictly constrained by the WSS assumption on the noise. In the case of impulsive noise, where noise energy is concentrated in time, the occurrences of such noise may result in false temporal maxima, thus yield wrong estimates for the response time and region. Secondly, since TCA only emphasizes the largest maxima for a selected pixel, it exhibits a large chance of ignoring any secondary activation peak times whose intensity response is lower than the selected maxima. This will cause the observer to miss multiple activations in the same region.

2.2. Competitive Subspace Projection (CSP)

As we know, if the data is naturally modeled by a Gaussian distribution, PCA optimally represents the data structure information in terms of minimum MSE between input and reconstructed projections. However, in most real data, such as images, are not well modeled by a single Gaussian distribution. PCA is not optimal in this case. The productive alternative is to project each cluster by a projection network where the network parameters are determined locally by the clustered data. This subspace projection is superior to other classification methods based on minimizing the distance between input and cluster centers, such as LBG and k-means algorithm, because it preserves the input structure better and has less effect of the vector scale ambiguity problem. To achieve this classification goal, competition or cooperation among multiple expert networks is needed. Haykin proposed the OIAL method incorporating competition among each PCA network using generalized Hebbian adaptation. This method is optimal in terms of MSE only if each cluster approximates the same cluster centroid.

We propose the CSP strategy to classify clusters using subspace projection while explaining the different centroids in each clusters. The block diagram of the CSP network is shown in Fig. 1. It consists of multiple (K) autoassociative networks corresponding to K patterns to be classified. When a input vector $\boldsymbol{x}$ enters the system, the K experts compete in a winner-take-all fashion in terms of MSE between input $\boldsymbol{x}$ and the reconstruction $\tilde{\boldsymbol{x}}$. The expert with least MSE wins the adaptation. The adaptation in each expert is given by supervised LMS strategy using the reconstruction $\tilde{\boldsymbol{x}}$ as the desired response. After the adaptation for the whole CSP network converges, the input data is classified to K patterns corresponding to K autoassociative networks.

To explain the different centroid location of each cluster, an intuitive solution is to add bias in the hidden layer of the network

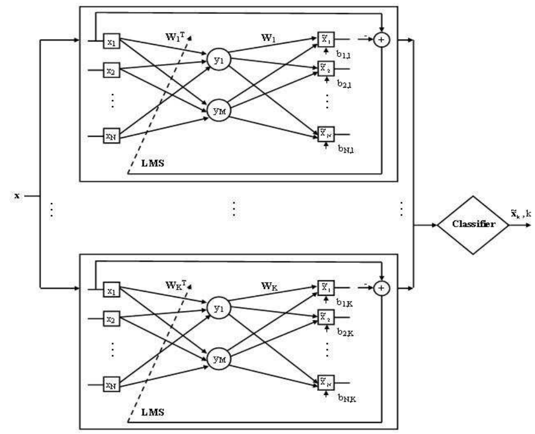


Fig. 1. The block diagram of competitive subspace projection methodology.

to account for the mean of projection variable. However, adding bias in the hidden layer introduces the 4^{th} order product between projection weights and bias into the cost function. To simplify the mathematical derivation, the bias is equivalently shifted to the output layer. Thus the cost function $J(\boldsymbol{W}, \boldsymbol{b})$ for each expert is defined as

$$\begin{aligned} J(\boldsymbol{W}, \boldsymbol{b}) =& \frac{1}{2}||\boldsymbol{x} - \tilde{\boldsymbol{x}}||^2 \\ =& \frac{1}{2}(||\boldsymbol{x}||^2 - 2\boldsymbol{x}^T\boldsymbol{W}\boldsymbol{W}^T\boldsymbol{x} + \boldsymbol{x}^T\boldsymbol{W}\boldsymbol{W}^T\boldsymbol{W}\boldsymbol{W}^T\boldsymbol{x} \\ &+ 2\boldsymbol{x}^T\boldsymbol{W}\boldsymbol{W}^T\boldsymbol{b} - 2\boldsymbol{x}^T\boldsymbol{b} + \boldsymbol{b}^T\boldsymbol{b}) \end{aligned} \tag{2}$$

where $\boldsymbol{x}$ is the input vector containing neighborhood pixels, $\tilde{\boldsymbol{x}} = \boldsymbol{W}\tilde{\boldsymbol{y}} + \boldsymbol{b} = \boldsymbol{W}\boldsymbol{W}^T\boldsymbol{x} + \boldsymbol{b}$ is the reconstruction and $\tilde{\boldsymbol{y}} = \boldsymbol{W}^T\boldsymbol{x}$ is the projection vector which has less dimension than that of $\boldsymbol{x}$. Based on the matrix lemma of $\partial\boldsymbol{a}^T\boldsymbol{a}/\partial\boldsymbol{w} = 2\boldsymbol{J}(\boldsymbol{a}, \boldsymbol{w})\boldsymbol{a}$ where $\boldsymbol{J}_{i,j}(\boldsymbol{a}, \boldsymbol{w}) = \partial a_j/\partial w_i$, the adaptation criterion is written as

$$\begin{aligned} \Delta\boldsymbol{w}_i &= \eta[\tilde{y}_i(\tilde{\boldsymbol{x}} - \boldsymbol{x}) + \boldsymbol{x}(\boldsymbol{w}_i^T\tilde{\boldsymbol{x}} - \tilde{y}_i)] \\ \Delta\boldsymbol{b} &= 10\eta[\boldsymbol{x} - \tilde{\boldsymbol{x}}] \end{aligned} \tag{3}$$

where $\boldsymbol{w}_i$ is the i^{th} column of $\boldsymbol{W}$.

The advantage of competitive subspace projection method lies in three aspects. First, the adapted bias simultaneously with projection axis adaptation gives optimal data representation. Without bias incorporated into the multiple autoassociators, the competition gives the same clustering as OIAL does. This kind of clustering actually separates the space into multiple cones with the same vertices at the origin. This kind of clustering neglects varied cluster locations while the proposed CSP methodology regards the cluster as a combination of its spatial location and its shape linearly represented by the projected axes. Furthermore, unlike some other methods as local PCA [7] which treats finding spatial locations and shapes of clusters as two independent processes, CSP couples the two aspects of cluster representation and adapts them simultaneously to optimally represent data space. Secondly, this subspace projection method is well-performed in noise-suppression. Finally, this method trains the competitive system by supervised

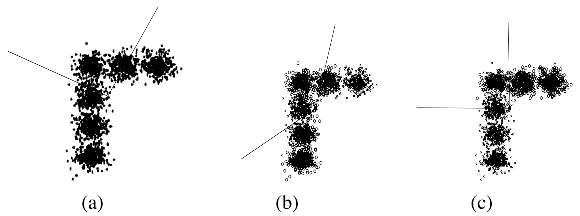


Fig. 2. (a) Synthetic data and its PCA projection, (b) clustered to two sets by competitive Hebbian learning, (c) clustered to two sets by competitive biased subspace projection, where the two dark lines represent the two projection directions.

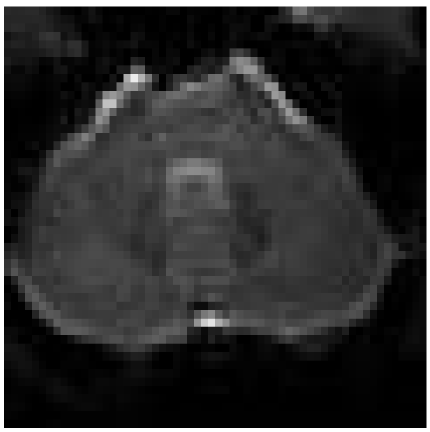


Fig. 3. Raw functional transverse brain image for detection of alcohol effects with size of 50×60.

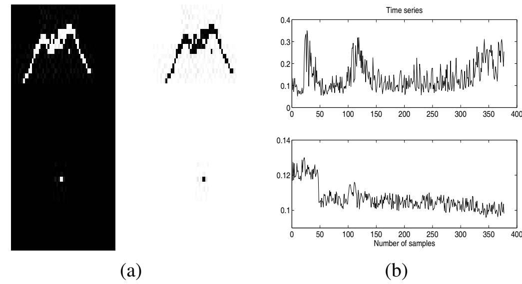


Fig. 4. (a) Pixel classification by competitive subspace projection, (b) The time series corresponding to the classified pixels in (a).

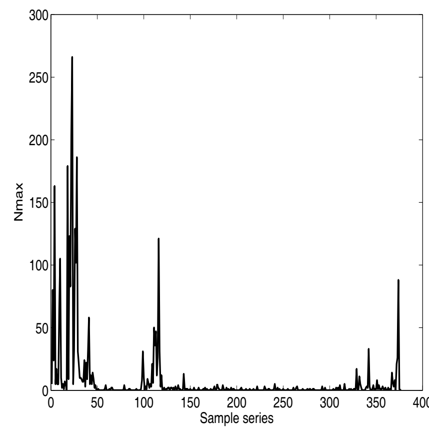


Fig. 5. The temporal maxima plot for temporal clustering analysis (TCA) method.

training instead of unsupervised training. This is done by estimate of desired response by using autoassociator. Then the computational load is greatly reduced.

2.3. Application of CSP to fMRI

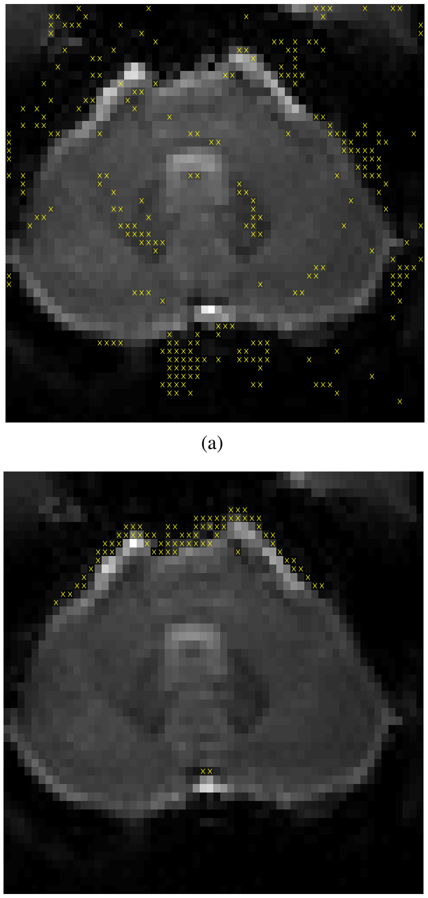
The purpose is detect when and where alcohol takes effect inside the brain after drinking for human beings. The data consists of a time series of 2D fMRI images (considering a single brain slice). While most pixels correspond to noisy background images, some pixels show temporal activation as a response to the alcohol in the system. The classification using CSP can segment different patterns based on this spatio-temporal structure of the activation. The patterns which have specific time structures give information about the spatial location and the temporal response of the alcohol effect.

3. NUMERICAL RESULTS

Algorithm Effectiveness: We compare the projection performance of the competitive biased subspace projection and the competitive Hebbian learning by the two dimensional synthetic data with two natural experts shown in Fig 2(a), where the two lines shows the global PCA projections due to the two eigenvalues. We can see in this case PCA is not optimal since it doesn't find the two orthogonal projection axes. In the two competitive learning strategies, each expert preserves only one projection axis for a fair comparison with global PCA preserving the two eigenvectors. The proposed method has a final MSE of 29.4% of the total energy while the competitive Hebbian learning conserves that of 52.4%. It is clearly demonstrated in Fig. 2 that the competitive Hebbian learning doesn't find the correct projection direction (x, y coordinate) due to the bias effect while the competitive biased subspace projection method does.

Alcohol Detection: The fMRI brain images detecting functional effects before and after drinking alcohol are collected on eight human volunteers using a 3T MRI scanner(GE Medical Systems Milwaukee, WI) at UF. A gradient-echo EPIBOLD (or EPI-RT according the current update) pulse sequence was used with the following scan parameters: TR/TE/FA = 6s/30ms/90o, Field of View = 240mm, matrix size = 128x128 with an in-plane resolution of 1.875 x 1.875 mm2 and 36 slices (3.5 mm thick) without gap covering the whole brain (in this case, only one slice (8) is used for simplicity). The functional images consist of 50 reference samples (at 6s per sample) before drinking alcohol and 360 samples after drinking alcohol.

The first 3 and last 30 images are discarded due to initial and final instability. Thus total 377 samples are used. Each sample has the interested center region of 50×60 pixels shown in Fig. 3. The competitive subspace projection method classifies the pixels into



(a)

(b)

Fig. 6. (a) Pixel activation location by TCA, (b) Pixel activation location by CSP.

two categories based on their time series response in Fig. 4. The time series for each category, given by the centroid of the pixel time series, clearly show that one class has two peaks, while the first one corresponds to the drinking trembling and the second peak reflects the alcohol effect some time after. At the same time another class keeps almost constant time response except for a abrupt change at the drinking time. The time peak, denoting the alcohol effect, is between 117 to 119 samples, which is quite close to the that of the TCA method (sample 116) shown in Fig. 5. The spatial activation locations by these two methods are shown in Fig. 6. The activation pixels classified from the competitive subspace projection method focus on the both upper boundary of the brain and the lower center. The TCA method has more expanded pixel distribution even into the background because of the noise inference. It demonstrates the proposed method has more robust to noise rejection effect.

4. CONCLUDING REMARKS

The proposed competitive subspace projection method avoids bias effect of each cluster centroids when projecting data into its subspace. It optimally clusters in second order statistics. The future work can be focused on orthogonal constraint and some alternative optimization criteria.

Functional MRI analysis provides a valuable tool for understanding brain activity in response to external stimuli. In this paper, we firstly incorporate competitive learning as a tool for extracting temporal and spatial activations in sequences of fMRI images that are taken from subjects who are exposed to alcohol. It has been shown that the conclusions drawn from CSP and a previously proposed method (TCA) are convincingly consistent that the estimated timing of alcohol's effects on the human brain are accurate.

On a simulated case study from the real fMRI data, it was shown that CSP has less noise effect than TCA. Further study to understand the effects of alcohol and other stimuli on the human brain using fMRI measurements need to be conducted to arrive at some conclusions with certainty.

5. REFERENCES

[1] Y. Liu, J-H Gao, H-L Liu and P. T. Fox, "The Temporal Response of the Brain after Eating Revealed by Functional MRI," Nature, vol. 405, pp. 1058-1062, 2000.

[2] S-H Seo, and M. R. Azimi-Sadjadi, "Orthogonal Subspace Projectionn Filtering for Stereo Image Compression," Proceedings of the 1998 IEEE International Conference, vol. 5, pp. 2577-2580, 1998.

[3] C-I Chang, X. L. Zhao, M. L. G. Althouse, and J. J. Pan, "Least Squares Subspace Projection Approach to Mixed Pixel Classification for Hyperspectral Images," IEEE Transactions, Geoscience and Remote Sensing, vol. 2, pp. 69-79, 1994.

[4] S. Haykin, "Neural Networks: A Comprehensive Foundation," New York, Macmillan, 1994.

[5] R. A. Jacobs, M. I. Jordan, S. J. Nowlan, and G. E. Hinton, "Adaptive mixtures of local experts," Neural Computation, vol. 3, pp. 79-87, 1991.

[6] T. K. Leen and N. Kambhatla, "Fast Non-Linear Dimension Reduction," Advances in Neural Information Processing Systems 6, Morgan Kauffman Publishers, 1994.

[7] N. Kambhatla, and T. K. Leen, "Dimension Reduction by Local Principal Component Analysis," Neural Computation, vol. 9, pp. 1493, 1997.

[8] R. D. Dony, and S. Haykin, "Optimally Adaptive Transform Coding," IEEE Trans., Image Processing, vol. 4, pp. 1358-1370, 1995.

[9] C. L. Fancourt, and J. S. Principe, "Soft Competitive Principal Component Analysis Using the Mixture of Experts," PhD Dissertation, University of Florida, 1998.

[10] R. D. Dony, and S. Haykin, "Neural Netowrk Approaches to Image Compression," Proceedings of the IEEE, vol. 83. pp. 288-303, 1995.

[11] G. He, G. A. James, J. Mao, S. M. Gold, and Y. Liu, "Time-dependent Effects of Alcohol Administration in the Brain," Proceedings of 8th International Conference on Functional Mapping of Human Brain, Order. 561, 2002.