Constrained CCA with Different Novel Linear Constraints and a Nonlinear Constraint in fMRI

D. Cordes[1], R. Nandy[2], and M. Jin[1]

[1]Radiology, University of Colorado Denver, Aurora, Colorado, United States, [2]Biostatistics and Psychology, UCLA, Los Angeles, California, United States

Introduction

The importance of multivariate statistical methods has been realized in fMRI in recent years as they are more proficient in detecting brain activations in a noisy environment. One such popular method is local canonical correlation analysis (CCA), where instead of looking at the single voxel time course, the joint time courses of a group of neighboring voxels are investigated. The value of a suitable test statistic is used as a measure of activation. It is customary to assign the value to the center voxel. However, this is a choice of convenience and the method is prone to false activations especially in a region of localized strong activation. The reason for the increase in false activations is due to two different deficiencies of CCA: The first deficiency is due to too much freedom of the spatial weights allowing for both positive and negative linear combinations of voxels which results in an improper spatial smoothing. The second deficiency is the so-called bleeding artifact [1] due to the center voxel assignment scheme without any constraint on the spatial weights. To rectify these deficiencies we investigate different novel linear constraints and a nonlinear constraint for the spatial weights on the ability to detect activation patterns in a standard fMRI motor task and a more complicated episodic memory task.

Theory

Let Y be the matrix representing p voxel time courses with dimension $t \times p$ and X the conventional design matrix of size $t \times r$ for the r temporal regressors. Furthermore, let α and β be two unknown vectors of size $p \times 1$ and $r \times 1$, respectively. In CCA, we look for the linear combinations of voxel time courses $Y\alpha$ and temporal regressors $X\beta$ such that the correlation between both quantities is maximum. This leads to an eigenvalue problem with min*(p,r)* solutions from which the solution with the largest eigenvalue (i.e. maximum canonical correlation) is being chosen. To put constraints on the spatial weights α, we consider the following four scenarios for the components α_i of α where α_1 is the weight for the center voxel and the other $\alpha_i's$ represent the weights for the neighborhood voxels.

#1. $\alpha_i > 0 \; \forall \, i$ # 2. $\alpha_1 \geq \frac{1}{p-1}\sum_{i=2}^{p} \alpha_i > 0$ and $\alpha_i > 0 \; \forall \, i$ #3. $\alpha_1 \geq \sum_{i=2}^{p} \alpha_i > 0$ and $\alpha_i > 0 \; \forall \, i$ #4. $\alpha_1 \geq \max(\alpha_i) > 0$ and $\alpha_i > 0 \; \forall \, i$.

All of these constraints lead to a proper spatial smoothing (low pass filtering) because all components of α are larger than zero. Method 1 has been investigated by Friman et al [2] in fMRI and found to be superior than unconstraint CCA. It can easily be shown [3] that constraint #1 leads to the same equation as unconstrained CCA. Thus, the solutions of unconstrained CCA need to be only searched if constraint #1 is in addition satisfied. Computationally, method 1 is fairly easy to implement. The disadvantage of method 1 (and also of unconstrained CCA) is the bleeding artifact because the center voxel for which the activation is assigned to, is not guaranteed to have the maximum spatial weight; similarly method 2. Method 3 and method 4 are quite different from the previous methods because the constraints do not lead by construction to any bleeding artifact. Method 4 has the nice property that the coefficients are minimally restricted and at the same time satisfying the necessary positivity conditions and providing maximum weight for the center voxel.

To solve for linear constraints in α (methods 2-3), it is possible to linearly transform the α to $\tilde{\alpha} = M\alpha$ to condition 1 such that $\tilde{\alpha}$ satisfies the simple condition $\tilde{\alpha}_i > 0 \; \forall \, i$. For example, for method 3 matrix $M=(M_{ij})$ is given by $M_{ij} = 1$ for $i = j$, $M_{1j} = -1$ for j>2, all other $M_{ij} = 0$. This transform of α then leads to a transformation of the data Y so that the spatial covariate becomes $YM^{-1}\tilde{\alpha} = \tilde{Y}\tilde{\alpha}$, and the maximum correlation between this covariate and $X\beta$ need to be found subject to the constraint on $\tilde{\alpha}$. For the nonlinear constraint (method 4), such a transformation is not possible. We derived the solution for this case from first principles using Lagrange multipliers by converting the CCA problem into a multivariate regression problem $Y\alpha = X\beta + \varepsilon$ where $\beta = X^{+}Y\alpha$. Note that this β is only up to a scale factor equivalent to the β in CCA. Using the least squares approach then leads to the objective function

$$f(\alpha,\lambda,\mu,\nu) = \alpha' A\alpha + \lambda(\alpha' B\alpha - 1) + \mu_1\left(\alpha_1 - \max(\alpha_2, \dots, \alpha_p) - \nu_1^2\right) + \mu_2(\alpha_2 - \nu_1^2) + \cdots + \mu_p\left(\alpha_p - \nu_p^2\right)$$

where $A = Y'(1 - XX^{+})'(1 - XX^{+})Y$, $B = cov(Y)$, and $\lambda, \mu = (\mu_1, \dots, \mu_p), \nu = (\nu_1, \dots, \nu_p)$ are Lagrange multipliers. Differentiation leads to different scenarios:

A) The solution is equivalent to a solution for the unconstrained CCA problem that also satisfy the constraints in #4.

B) Another solution for α is obtained by the following general eigenvalue equation (here we use Einstein summation convention for convenience):

$\mathrm{H}_{mj}\alpha_j = -\lambda \mathrm{S}_{mj}\alpha_j$ where $\alpha_1 = \alpha_l$, $\mathrm{H}_{1j} = \mathrm{A}_{1j} + \mathrm{A}_{lj} + (\mathrm{A}_{11} + \mathrm{A}_{l1})\delta_{jl}$, $\mathrm{S}_{1j} = \mathrm{B}_{1j} + \mathrm{B}_{lj} + (\mathrm{B}_{11} + \mathrm{B}_{l1})\delta_{jl}$, $\mathrm{H}_{mj} = \mathrm{A}_{ij} + \mathrm{A}_{i1}\delta_{jl}$, $\mathrm{S}_{mj} = \mathrm{B}_{ij} + \mathrm{B}_{i1}\delta_{jl}$

for $l \in \{2, \dots, p\}$ but fixed, $j = 2, \dots, p$, $i = 2, \dots, p \backslash \{l\}$, $m = i$ if $i < l$, $m = i - 1$ if $i > l$, and α is normalized by $\alpha' A\alpha = -\lambda$.

If these conditions are satisfied, f achieves an extremum in the interior region of the α domain. A full search need to be carried out over all possible l.

C) A third possible solution is given for α at the boundary, i.e. $\alpha = (c, c, \dots, c)'$ such that the usual normalization condition $\alpha' B\alpha = 1$ is satisfied.

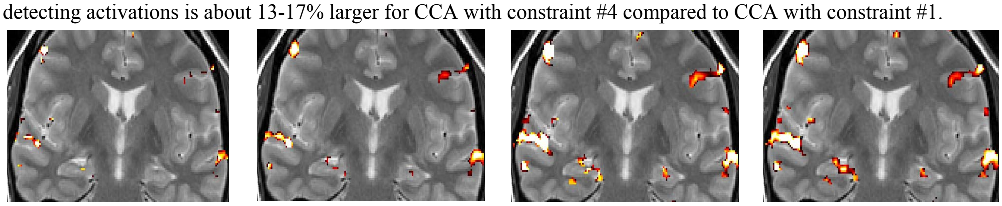
In order to find the optimum solution, the canonical correlation is determined for all possible solutions and the best is chosen.

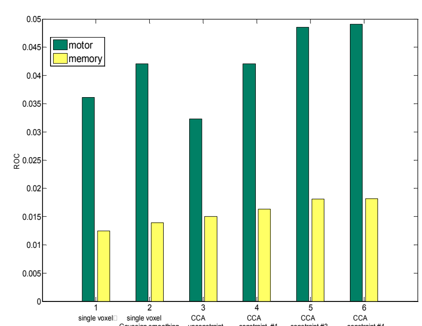
Methods

FMRI was performed in a 3.0T GE HDx MRI scanner equipped with an 8-channel head coil and parallel imaging acquisition using the following parameters: ASSET=2, ramp sampling, TR/TE=2sec/30ms, FA= 70deg, FOV=22cmx22cm, thickness/gap=4mm/1mm, 25 slices, in-plane resolution 96x96 interpolated to 128x128. We acquired three fMRI data sets. The first data set was collected during resting-state where the subject tried to relax and refrain from executing any overt task with eyes closed. The second data set was collected while the subject was performing an episodic memory paradigm. Briefly, this paradigm consisted of memorization of novel faces paired with occupation, containing 6 periods of encoding, distraction, and recall tasks (slices parallel to the long axis of hippocampus). The third data set was obtained from a conventional motor task (axial slices, four 30 sec periods with on/off finger tapping). We computed activation maps and modified ROC curves using real (non simulated) data as outlined in [4] by estimating the fraction of positives from activation data and the fraction of false positives (FFP) from resting-state data. Furthermore, as a measure of performance we computed the area under the ROC curves (aROC), integrated over FFP∈[0,0.1]. As local neighborhood for CCA we use every 3x3 in-plane pixel patch and compute CCA for all 256 configurations involving the center voxel and its 8 neighbors.

Results

In Fig.1 below (from left to right), we show the activation pattern for the memory paradigm for the contrast encoding-distraction at a p-value of 10^{-5} for single voxel GLM, single voxel GLM with Gaussian smoothing (FWHM 2.3 pixels), CCA with constraint #1, CCA with constraint #4. In Fig.2 (right) we show aROC bar graphs for both motor (green) and memory data (yellow) for (from left to right) single voxel GLM, single voxel with Gaussian smoothing, unconstraint CCA, CCA with constraints #1, #3, and #4. Notice that the improvement in detecting activations is about 13-17% larger for CCA with constraint #4 compared to CCA with constraint #1.





References and Acknowledgments [1] Nandy, R., et al., 2004. Magn. Reson. Med., 52, 947-952. [2] Friman, O., et al., 2003. NeuroImage, 19, 837-845. [3] Das, S., et al., 1994. Lin. Algebra Appl, 210, 29-47. [4] Nandy, R., et al., 2004. Magn. Reson. Med., 52, 1424-1431. This work is supported by NIH R21NIA.