

ODVBA-C: Optimally-Discriminative Voxel-Based Analysis of Continuous Variables

Tianhao Zhang¹, Theodore D. Satterthwaite², and Christos Davatzikos¹

¹Department of Radiology, University of Pennsylvania, Philadelphia, PA 19104, USA

²Department of Psychiatry, University of Pennsylvania, Philadelphia, PA 19104, USA

tianhao.zhang@uphs.upenn.edu, sattertt@mail.med.upenn.edu christos.davatzikos@uphs.upenn.edu

Abstract—In this paper, we propose a new method that utilizes a novel spatially adaptive scheme for detection of multivariate neuroimaging patterns relating to a continuous subject-level variable, aiming to effectively determine the optimal spatially adaptive filtering of neuroimaging data from the perspective of finding relationships between imaging and continuous (e.g. clinical and cognitive) variables. Analyses employ local pattern analysis using regularized least square regression with nonnegativity constraints within a spatial neighborhood around each voxel. Within each neighborhood, we determine the optimal regression coefficients that relate local patterns to the continuous variable of interest. As each voxel belongs to multiple overlapping neighborhoods, the statistic for a given voxel is determined by combining weights from all neighborhoods to which the voxel participates. Finally, non-parametric permutation testing is used to obtain a voxelwise significance map. Using both simulated and real fMRI data, we demonstrate the effectiveness of the proposed method.

Keywords—fMRI; Regression; Nonnegativity; ODVBA

I. INTRODUCTION

Many variables of interest are continuous. Despite this, multivariate pattern analysis techniques are frequently configured for classification of dichotomous or categorical variables. Sophisticated modeling of continuous variables has become more important as lifespan studies of brain development and aging have gained prominence. Furthermore, as part of the recent Research Domain Criteria [3], psychiatric illnesses are increasingly conceptualized as inter-related dimensional impairments of specific brain systems, which are best modeled as continuous variables.

Regression using a general linear model (GLM) [2] is the most commonly used technique to relate neuroimaging data to continuous subject variables of interest. However, GLM suffers two principal limitations. First, it is a fundamentally univariate approach: GLM is insensitive to relationships that are found jointly but not individually. Second, GLM is limited by the Gaussian smoothing that is routinely employed to improve the signal-to-noise ratio in the data. Uniform application of a given Gaussian smoothing kernel inevitably involves an undesirable trade-off: a small kernel size will limit statistical power, whereas a large kernel will blur across real boundaries and sometimes produce false positive results.

In this paper, we propose a novel method that overcomes

both of these limitations, allowing the precise identification of multivariate imaging patterns that relate to continuous outcomes. This is accomplished using regularized least squares regression with nonnegativity constraints of multivariate patterns within a spatial neighborhood around each voxel. The resultant components can be viewed as weights of a locally adaptive spatial filter that is optimally designed to highlight the areas that are significantly related to the continuous outcome variable. Because a neighborhood is created around every voxel, multiple overlapping neighborhoods result; the statistical value for each voxel is determined by combining the weights from each neighborhood to which it belongs. Finally, significance levels are assessed using non-parametric permutation testing.

This method is a regression-based extension of the previous implementation of ODVBA, which was originally created in the context of two-group classification [10]. We therefore name it ODVBA-C (that is, Optimally-Discriminative Voxel-Based Analysis of Continuous Variables). Below, we describe the formulation of the method and present results from experiments using both simulated data and fMRI data from a sample of 32 healthy subjects who completed a standard working memory task. Results reveal that ODVBA-C performs better than the standard GLM method, for detection of multivariate relationships between neuroimaging data and continuous outcome variables.

II. THE METHOD

The proposed method works in three stages:

A. Regional nonnegative regression analysis

We first conduct a regional regression analysis with a nonnegativity constraint in a spatial neighborhood around each voxel to determine the optimal coefficients, which can also be considered as the optimal non-isotropic smoothing in the vicinity of a voxel in a local regression problem.

In particular, for each given voxel x in the volume X (neighborhood defined by $\|x - x_i\| < \xi$), ODVBA-C gets a vector: $\vec{\theta} = [x, x_1, \dots, x_{k-1}]^T$, where x_1, \dots, x_{k-1} are the $k - 1$ neighbors of x , and then it constructs a learning set $\Theta = [\vec{\theta}_1, \vec{\theta}_2, \dots, \vec{\theta}_N]^T$ from N subjects to be investigated.

A single response variable y is related to a single subvolume vector $\vec{\theta}$:

$$y = (\vec{w})_1 x + (\vec{w})_2 x_1 + \dots + (\vec{w})_k x_{k-1} + \varepsilon = \vec{w}^T \vec{\theta} + \varepsilon, \quad (1)$$

where, \vec{w} is the regression coefficients and ε is the error. The regularized least square fit is a common way to access the coefficients on the predictor variable as follows:

$$\min_{\vec{w}} \left(\sum_{i=1}^k \left\| y_i - \vec{w}^T \vec{\theta}_i \right\|^2 + \tau \|f(\vec{w})\|^2 \right), \quad (2)$$

where $\|f(\vec{w})\|^2$ is a regularization term to be defined. For example, the ridge regression has the term: $\|\vec{w}\|^2$.

The above formulation of optimization considers the subjects globally. However, it does not take into account the local discriminative information in the sample space that can be beneficial to find the optimal direction highlighting the dynamic coefficients and eliminating the effects of outliers. Therefore, we also focus on the separation between the subjects with the lowest values of variable and those with highest values. It can be defined as:

$$\begin{aligned} & \max_{\vec{w}} (\vec{w}^T \vec{m}_H - \vec{w}^T \vec{m}_L)^2 \\ & = \max_{\vec{w}} \vec{w}^T S \vec{w}, \end{aligned} \quad (3)$$

where, \vec{m}_H and \vec{m}_L are the means of the samples with the highest and lowest values of variable, respectively: $\vec{m}_{H,L} = \frac{1}{N_{H,L}} \sum_{\vec{\theta} \in P_{H,L}} \vec{\theta}$; P_H and P_L are the two parts of subjects, respectively; N_H and N_L denote the numbers of samples; $S = (\vec{m}_H - \vec{m}_L)(\vec{m}_H - \vec{m}_L)^T$.

Moreover, we expect that the resultant vector \vec{w} to be nonnegative, which can help us interpret the brain activity changes: the larger the value of $(\vec{w})_i$ is, the more the corresponding element $(\vec{\theta})_i$ contributes to the regional regression. Therefore, we consider the above two items under the Nonnegative Quadratic Programming [8] with incorporating (3) into the regularization term of (2), then we have:

$$\begin{aligned} & \min_{\vec{w}} \left(\sum_{i=1}^k \left\| y_i - \vec{w}^T \vec{\theta}_i \right\|^2 + \gamma \vec{w}^T (-S + |\lambda_{min}| + \tau) \vec{w} \right) \\ & \text{subject to } (\vec{w})_i \geq 0, \quad i = 1, \dots, k. \end{aligned} \quad (4)$$

Equation (4) can be reduced to:

$$\min_{\vec{w}} (\vec{w}^T D \vec{w} - b \vec{w}), \quad \text{s.t. } (\vec{w})_i \geq 0, \quad (5)$$

where $D = \vec{\Theta}^T \vec{\Theta} - \gamma S + (|\lambda_{min}| + \tau) I$; γ is the controlling parameter; $b = 2Y^T \vec{\Theta}$; $|\lambda_{min}|$ is the absolute of the smallest eigenvalue of $\vec{\Theta}^T \vec{\Theta} - \gamma S$; $0 < \tau \ll 1$; I is the identity matrix; Y is the vector of variable: $Y = [y_1, y_2, \dots, y_N]^T$.

It can be inferred that D is positive definite, so (5) is a convex function and has the unique global minimum, and the nonnegative vector \vec{w} is estimated by multiplicative updates which iteratively minimize the objective function [8][10]:

$$(\vec{w})_i \leftarrow \left(\frac{(Y^T \vec{\Theta})_i + \sqrt{(Y^T \vec{\Theta})_i^2 + 4(D^+ \vec{w})_i (D^- \vec{w})_i}}{2(D^+ \vec{w})_i} \right) (\vec{w})_i, \quad (6)$$

where $i = 1, \dots, k$; the nonnegative matrices D^+ and D^- are defined as follows: $D_{ij}^+ = D_{ij}$, if $D_{ij} > 0$; otherwise 0, and $D_{ij}^- = |D_{ij}|$, if $D_{ij} < 0$; otherwise 0.

B. Determining each voxels statistic

Since each voxel belongs to a large number of neighborhoods, each centered on one of its neighboring voxels, the statistical value at each voxel is determined by tallying contributions from all neighborhoods to which it participates:

$$T_x = \sum_{\mathbb{N} \in \Delta} \delta_{\mathbb{N}}(\vec{w}_{\mathbb{N}})_i, \quad i \in \{1, \dots, k\}, \quad (7)$$

where $\vec{w}_{\mathbb{N}}$ denotes the coefficients in neighborhood \mathbb{N} , $\Delta = \{\mathbb{N} | x \in \mathbb{N}\}$, $(\vec{w}_{\mathbb{N}})_i$ denotes that x is the i^{th} element in \mathbb{N} , and $\delta_{\mathbb{N}}$ denotes the *discrimination degree*:

$$\delta_{\mathbb{N}} = \|\vec{w}^T (\vec{m}_H - \vec{m}_L)\|^\phi, \quad (8)$$

where ϕ is a tuning parameter.

Because the coefficients obtained from the first stage are nonnegative, the possible cancelations between negative and positive loadings are avoided.

C. Permutation tests

Permutation tests [5] are used to obtain the statistical significance (p values) of the resulting statistical maps, assuming that the null hypothesis is that there is no changes relationship between the neuroimaging data and the continuous outcome variable. We randomly re-assign the outcome variable values to each subject, and then recalculate the statistic as described in Equation (7) for each voxel for N_p times. The p value for the given voxel can be calculated according to: the proportion of the permutation distribution (T_1, \dots, T_{N_p}) greater than or equal to the initial one (T_0).

III. EXPERIMENTS

Before applying ODVBA-C on the real fMRI data, we evaluate it using the simulated data.

A. Simulated data

In this experiment, we generated a series of images (forty) shown in Figure 1, in which there are two squares. In the left square, the values of intensity are increasing gradually with the number of the image; however, in the right square, the values are kept in a stable level. Figure 2B shows the obtained \vec{w} from the learning set constructed according to the neighborhood containing part of the left square; the estimated optimal filter is well aligned with the underlying intensity changes, within which it has high values. Figure 2C shows the \vec{w} obtained from the learning set constructed according to the neighborhood containing part of the right square; it is basically noise with very small values of \vec{w} ,

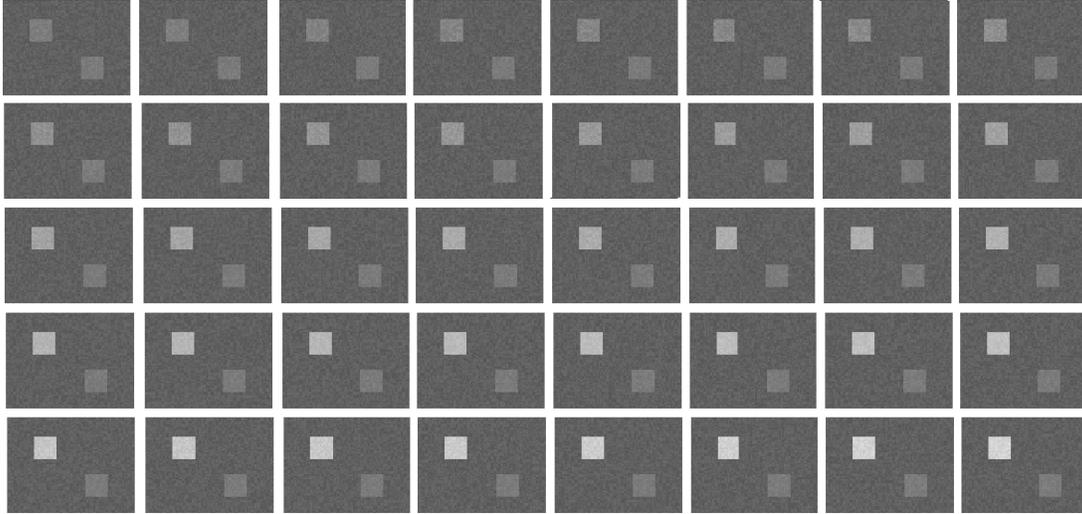


Figure 1. A series of (forty) images, with two squares inside. In the left square, the values of intensity are increasing gradually with the number of the image; however, in the right square, the values are kept in a stable level.

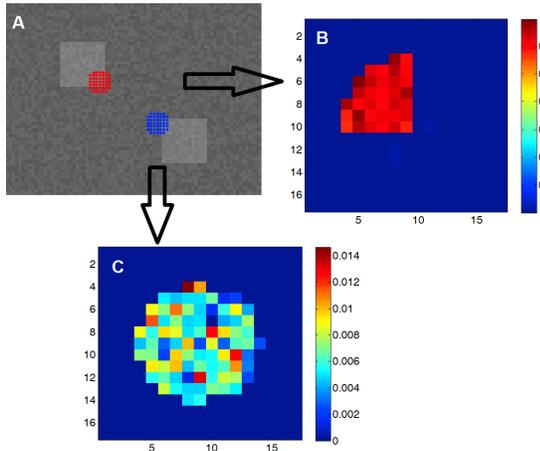


Figure 2. The demonstrations of the coefficients \vec{w} obtained by the local regression analysis on the simulated data. A) a sample of the simulated data; B) the coefficients obtained from the learning set constructed according to the neighborhood on the left square; C) the coefficients obtained from the learning set constructed according to the neighborhood on the right square.

indicating that no local filter can be found that reflects the continuous changes at that neighborhood.

The experiment verified the premise in the formulation of ODVBA-C: the filter can locally highlight continuous relationships.

B. Real fMRI data

Next, we evaluated the relative performance of ODVBA-C versus standard GLM using fMRI data from 32 healthy subjects ages 19-21 (12 male) who performed a standard N-back working memory task [11] as part of a study of neurodevelopment. In this task, working memory load was parametrically varied under three conditions, ranging from easy (0-back) to difficult (2-back). Here, we tested

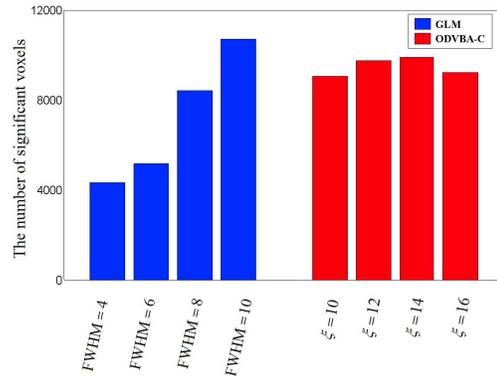


Figure 3. The number of significant voxels versus smoothing kernel sizes (mm), according to the results of GLM (left) and ODVBA-C (right).

the hypothesis that greater executive network activation is associated with better task performance. WM performance was summarized using the signal detection measure d' [9]. fMRI data was acquired using the following parameters: 231 volumes, matrix = 64×64 , slice thickness/gap = 3/0 mm, flip = 90° , TR/TE = 3000/32 ms, FOV = 192×192 mm. The data were realigned and high-pass filtered before transformation to MNI coordinates. The outcome of interest was the relationship between activation in the contrast of 2-back > 0-back and behavioral performance. Prior work [1] has established that in healthy subjects greater executive network activation is associated with better performance.

An important advantage of ODVBA-C is its spatially adaptive kernel, which has stable performance across varying neighborhood sizes. We illustrate this by reporting the number of significant voxels detected at a range of kernel sizes (the parameter ξ). As ODVBA-C was substantially more sensitive than GLM (see below), kernel sizes are compared using an uncorrected threshold of $p < 0.05$ for GLM and

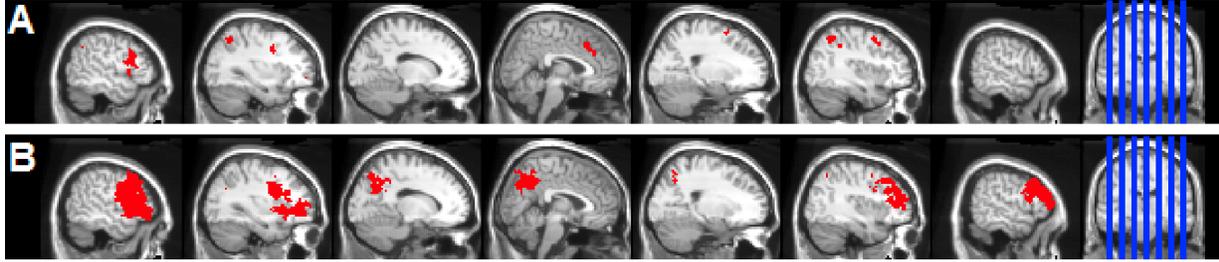


Figure 4. Significant areas of brain activity changes detected by GLM (A) and ODVBA-C (B), for the N-back working memory fMRI experiment. Multiple voxel comparisons were corrected using the FDR threshold of $q < 0.05$.

$p < 0.01$ for ODVBA-C. As seen in Figure 3, ODVBA-C is insensitive to the kernel size selected, in stark contrast to the relationship seen with GLM.

As a final step, we compare the results found by ODVBA-C ($\xi = 14$) and GLM with 8 mm FWHM. Other parameters of ODVBA-C are set as $\phi = 2$, $\gamma = 10^3$, $\tau = 10^{-5}$, and $N_p = 2000$. In both methods, type I error was controlled using the false discovery rate ($q < 0.05$) [4]. As displayed in Figure 4, ODVBA was substantially more sensitive than GLM in detecting the relationship between working memory performance and executive network activation, revealing significant clusters in the DLPFC and superior parietal cortex.

IV. DISCUSSION AND CONCLUSIONS

In this paper, we proposed a method for relating continuous outcome variables of interest to multivariate patterns of neuroimaging data by conducting regularized least square regression with nonnegativity constraint as spatially adaptive smoothing. Simulated data provided support for the method under controlled conditions with a known ground truth. Furthermore, application of ODVBA-C to real fMRI data illustrated the dramatic improvement in sensitivity afforded over GLM in detecting significant relationships between working memory performance and executive network activation. Critically, results from ODVBA-C are stable across a range of kernel sizes, in contrast to standard Gaussian smoothing that is typically applied prior to GLM. This approach is somewhat akin to spatially adaptive smoothing methods such as propagation-separation [6] and wavelet denoising [7]. However, ODVBA-C constructs an optimal smoothing kernel across the entire sample of subjects as a manner of multivariate machine learning; other methods smooth each subjects individually, which may both introduce bias and may not necessarily be optimal for detecting relationships within a given sample. Taken together, ODVBA-C represents a substantial improvement over standard methods, and has many potential applications to studies of brain development, aging, and psychopathology.

V. ACKNOWLEDGMENT

Data from studies funded by grants from the National Institute of Mental Health (NIMH) RC2 MH089983 and RC2 MH089924. The work was supported by NIH grant

R01AG14971. Dr. Satterthwaite was supported by NIMH K23MH098130 and the Marc Rapport Family Investigator grant through the Brain & Behavior Research Foundation.

REFERENCES

- [1] J.H. Callicott, V.S. Mattay, A. Bertolino, K. Finn, R. Coppola, J.A. Frank, T.E. Goldberg, and D.R. Weinberger, "Physiological characteristics of capacity constraints in working memory as revealed by functional MRI", *Cereb. Cortex*, vol. 9, pp. 20–26, 1999.
- [2] K. Friston, A. Holmes, K. Worsley, J. Poline, C.D. Frith, and R.S.J. Frackowiak, "Statistical parametric maps in functional imaging: a general linear approach", *Hum. Brain Mapp.*, vol. 2, pp. 189–210, 1994.
- [3] T. Insel, B. Cuthbert, M. Garvey, R. Heinssen, D.S. Pine, K. Quinn, C. Sanislow, and P. Wang, "Research domain criteria (RDoC): toward a new classification framework for research on mental disorders", *Am. J. Psychiatry*, vol. 167, pp. 748–751, 2010.
- [4] T.E. Nichols, and S. Hayasaka, "Controlling the familywise error rate in functional neuroimaging: a comparative review", *Hum. Brain Mapp.*, vol. 12, no. 5, pp. 419–446, 2003.
- [5] T.E. Nichols, and A.P. Holmes, "Nonparametric permutation tests for functional neuroimaging: a primer with examples", *Hum. Brain Mapp.*, vol. 15, pp. 1–25, 2002.
- [6] J. Polzehl, and V. Spokoiny, "Adaptive weights smoothing with applications to image restoration", *J. R. Statist. Soc. B*, vol. 62, pp. 335–354, 2000.
- [7] A. Pizurica, W. Philips, I. Lemahieu, and M. Acheroy, "A versatile wavelet domain noise iteration technique for medical imaging", *IEEE Trans. Med. Imaging*, vol. 22, pp. 323–331, 2003.
- [8] F. Sha, Y. Lin, L.K. Saul, and D.D. Lee, "Multiplicative updates for nonnegative quadratic programming", *Neural Computation*, vol. 19, pp. 2004–2031, 2007.
- [9] J.G. Snodgrass, and J. Corwin, "Pragmatics of measuring recognition memory: Applications to dementia and amnesia", *J. Exp. Psychol. Gen.*, vol. 117, pp. 34–50, 1988.
- [10] T. Zhang and C. Davatzikos, "ODVBA: Optimally-Discriminative Voxel-Based Analysis", *IEEE Transaction on Medical Imaging*, vol. 30, no. 8, pp.1441–1454, 2011.
- [11] J.D. Ragland, B.I. Turetsky, R.C. Gur, F. Gunning-Dixon, and T.Turner "Working memory for complex figures: An fMRI comparison of letter and fractal n-back tasks", *Neuropsychology*, vol. 16, no. 3, pp. 370–379, 2002.