ELSEVIER

NeuroImage

www.elsevier.com/locate/ynimg
NeuroImage 33 (2006) 63 – 71

Classification of functional brain images with a spatio-temporal dissimilarity map

Svetlana V. Shinkareva,[a,*] Hernando C. Ombao,[b,c,d] Bradley P. Sutton,[c,e] Aprajita Mohanty,[d] and Gregory A. Miller[c,d,f]

[a]*Center for Cognitive Brain Imaging, Carnegie Mellon University, USA*
[b]*Department of Statistics, University of Illinois at Urbana-Champaign, USA*
[c]*Beckman Institute, University of Illinois at Urbana-Champaign, USA*
[d]*Department of Psychology, University of Illinois at Urbana-Champaign, USA*
[e]*Bioengineering Department, University of Illinois at Urbana-Champaign, USA*
[f]*Department of Psychiatry, University of Illinois at Urbana-Champaign, USA*

Received 4 October 2005; revised 5 April 2006; accepted 25 June 2006
Available online 14 August 2006

Classification of subjects into predefined groups, such as patient vs. control, based on their functional MRI data is a potentially useful procedure for clinical diagnostic purposes. This paper presents an automated method for classifying subjects into groups based on their functional MRI data. The proposed methodology provides general framework using preprocessed time series for the whole brain volume. Using a training set of two groups of subjects, the new methodology identifies spatio-temporal features that distinguish the groups and uses these features to categorize new subjects. We demonstrate the method using simulations and a clinical application that classifies individuals into schizotypy and control groups.

Keywords: fMRI; Classification; Functional neuroimaging; RV-coefficient

Introduction

Functional magnetic resonance imaging (fMRI) is a noninvasive neuroimaging technique used to study functional activity in the living human brain. The task-evoked signal in fMRI is based on the blood oxygen level dependence (BOLD) effect (Ogawa et al., 1990). Neuronal activity is measured indirectly by recording hemodynamic changes (Jueptner and Weiller, 1995). Data collected on each subject are four-dimensional, with three spatial dimensions measured over time.

This paper addresses the problem of classifying subjects into groups using both temporal and spatial information in fMRI data, which has the potential to assist in early disease detection and diagnosis. For example, mild cognitive impairment (MCI) is a condition that predicts the onset of Alzheimer's disease. It has been shown via FDG-PET that MCI has a distinct functional signature that is a more accurate predictor of the disease than neuropsychological tests (Chetelat et al., 2005). Another application, from cognitive neuroscience research, is confirmation of hypothesized group differences in BOLD response for distinct cognitive states.

Developing methods for classification of subjects based on neuroimaging data poses significant challenges. Large data size and small signal intensity change inhibit accurate classification. Extracting robust features representative of both spatial and temporal aspects of the data plays a crucial role in success of classification.

There have been a number of efforts to classify subjects into groups based on functional data. Some methods proceed from activation maps generated by the general linear model. Kontos et al. (2003) introduced an approach that uses space-filling curves for mapping 3D space into a linear domain. Wang et al. (2004) used dimension reduction techniques on the space-filling curves for discriminative pattern discovery. Liow et al. (2000) applied linear discriminant analysis on the principal components of PET data to classify HIV-1 seropositive patients into AIDS dementia complex (ADC) and non-ADC groups. Ford et al. (2003) used the same approach for discriminating patients and controls based on fMRI data for Alzheimer's disease, schizophrenia, and mild traumatic brain injury.

Some techniques focus their analysis on regions of interest (ROI). Bogorodzki et al. (2005) developed a method based on differences in regional brain activity. As part of feature selection for preselected ROIs, mean time intensity curves for voxels correlated with the stimulus are computed and modeled using a mixture of time shifted Gaussian functions. New subjects are

* Corresponding author. Fax: +1 412 268 2804.
E-mail address: shinkareva@cmu.edu. (S.V. Shinkareva).
Available online on ScienceDirect (www.sciencedirect.com).

1053-8119/$ - see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.neuroimage.2006.06.032

classified based on derived feature vectors from multiple ROIs. Pokrajac et al. (2005) introduced statistical distance and neural-network-based methods for classification of brain image data contingent on measures of dissimilarity between 3D probability distributions of ROIs.

Mitchell et al. (2004) applied machine learning methods to the classification of cognitive states based on fMRI data, a topic closely related to classification of subjects. They investigated several feature selection and classification approaches. Following the same framework, Zhang et al. (2005) applied machine learning techniques to classification of subjects based on activation maps. In both applications, Gaussian Naive Bayes, Support Vector Machines, and k nearest neighbor classifiers were examined.

While these methods are promising, most rely on contrast maps or predefined regions of interest and are not designed for detecting subtle temporal differences. This paper examines the feasibility of working directly with time series either from whole brain volumes or regions of interest suggested by results from previous studies to select voxels that exhibit highly discriminating features (based on a measure of temporal dissimilarity) and assign group membership to new subjects.

Methods

For this work, functional data are assumed to be motion-corrected, smoothed, and normalized to a standard anatomical template.

We address the problem of classifying subjects by a general methodology that takes two steps. The first is a feature selection step that identifies patterns predictive of group differences by localizing areas in space where temporal behavior is most dissimilar between groups. The second step is classification, where new subjects are assigned group membership based on similarity at selected voxels to subjects with known group membership.

As part of the feature selection step, we compare spatially localized sets of time series between groups, using a permutation test to identify voxels where groups exhibit temporal dissimilarity. As an intermediate step we would like to quantify the degree of similarity in the time series between two subjects for a selected brain region. Each subject's data for a specified region can be represented in a voxel-by-time matrix. Hence, comparing two subjects at a specified brain location can be accomplished by comparing two matrices.

In the remainder of this section, the similarity measure and the test statistic will be discussed in more detail. Once all the necessary tools are developed, the feature selection algorithm is presented, and classification is discussed.

Test statistic

It is important to define 'similarity' between two time series data sets. The goal of the proposed feature selection algorithm is to identify regions in the brain that best discriminate the groups, regions where temporal patterns differ between groups during an experimental task such as a cognitive challenge.

Typically, time series from different subjects are not directly comparable due to variations in stimulus presentation order. Therefore, subject and group differences are defined relative to experimental tasks related to implementation and the cognitive research question of interest. So here we follow suit and examine task relative temporal responses by including a hemodynamic response model in our test statistic.

A good similarity measure is sensitive enough to distinguish between cases where two sets of time series data contain similar signal patterns (related to the task) or similar noise patterns versus cases where only one set of time series data contains noise and the other contains signal. Problems could arise in cases where both sets of time series are noise. From the perspective of our task, those regions are similar, but the similarity measure could fail to identify them as similar.

In general, we are interested in a distance function between two voxel by time ($n \times T$) matrices G and H, relative to an experimental task, that is of the form

$$d = \mathrm{Dist}[(G, W_G), (H, W_H)], \tag{1}$$

where the task waveform matrices W_G and W_H have a time series of stimulus response timing for G and H, respectively, obtained by convolving the event stimulus train with hemodynamic response function in each voxel. Hence, rows of W are similar to the design matrix in the general linear model with one explanatory variable (for illustration, see top panel of Fig. 2).

For a spatial neighborhood of size n centered at any voxel, we can compute pairwise d values for all combinations of subjects. For illustration, consider two groups of subjects, groups A and B, with two and three subjects, respectively. Each subject's data are in the form of an $n \times T$ matrix, where n is the number of voxels in the localized region. We can construct a dissimilarity matrix D of all the distances from Eq. (1) (Scheme 1). The group A by group B quadrant of the D matrix dissimilarity values can be combined into an observed test statistic d_{obs}. For example, d_{obs} could be the average: $d_{\mathrm{obs}} = d_{A_1B1} + d_{A_1B2} + d_{A_1B3} + d_{A_2B1} + d_{A_2B2} + d_{A_2B3})/6$. Another possibility is to consider the median value. When the observed test statistic value exceeds a predefined threshold, we conclude that the temporal dynamics of the two localized brain regions are dissimilar.

Significance testing and FDR

Statistical inference on the observed test statistic is performed using a permutation test framework. The observed test statistic d_{obs} is computed from a subject-by-subject dissimilarity matrix D. Large values of d_{obs} indicate that the two groups are different relative to the experimental paradigm. We would like to know if the observed test statistic d_{obs} is statistically significant. We can construct a nonparametric test using the distribution of the test statistic under the null hypothesis of no group difference in temporal similarity (that is, subjects are randomly allocated between the two groups) making weak distributional assumptions. Under the null hypothesis, the distribution of d_{obs} does not change with random permutation of the D matrix (Fisher, 1935; Pitman, 1937; Hubert, 1987).[1]

The permutation distribution of the test statistic consists of all possible permutations of subjects' labels, found by simultaneously permuting rows and columns of the D matrix. Individual

[1] Permutation methods have been successfully applied to fMRI data by Gamalo et al. (2005), Hayasaka and Nichols (2003), Nichols and Holmes (2001), and Raz et al. (2003).

$$D = \begin{array}{c|cc|ccc} & \text{A1} & \text{A2} & \text{B1} & \text{B2} & \text{B3} \\ \hline \text{A1} & d_{A_1A_1} & d_{A_1A_2} & d_{A_1B_1} & d_{A_1B_2} & d_{A_1B_3} \\ \text{A2} & d_{A_2A_1} & d_{A_2A_2} & d_{A_2B_1} & d_{A_2B_2} & d_{A_2B_3} \\ \hline \text{B1} & d_{B_1A_1} & d_{B_1A_2} & d_{B_1B_1} & d_{B_1B_2} & d_{B_1B_3} \\ \text{B2} & d_{B_2A_1} & d_{B_2A_2} & d_{B_2B_1} & d_{B_2B_2} & d_{B_2B_3} \\ \text{B3} & d_{B_3A_1} & d_{B_3A_2} & d_{B_3B_1} & d_{B_3B_2} & d_{B_3B_3} \end{array}$$

Scheme 1. Illustration of a D matrix constructed for two groups of subjects. Images $A1$ and $A2$ belong to one group. Images $B1$, $B2$ and $B3$ belong to the second group of subjects.

d values do not need to be recalculated as we are simply reordering rows and columns of the D matrix. The significance of the observed test statistic can be computed considering the proportion of times that a value as high or higher than d_{obs} is obtained.

The number of possible permutations can be quite large, and listing all possible permutations can be computationally costly. As an alternative, an approximate test can be made by drawing a random sample from the space of all possible permutations, with as few as 1000 permutations drawn (Wasserman and Böckenholt, 1989; Nichols and Holmes, 2001).

At each voxel, a test statistic and corresponding p value are computed. Empirically obtained p values are then thresholded through comparison to a predetermined α using a false discovery rate to account for the multiple comparisons problem (Benjamini and Hochberg, 1995; Genovese et al., 2002).

Weighting scheme

To get spatial sensitivity, increased stability, and power of the classification procedure, a test statistic can be computed over a small localized neighborhood. Both the central voxel and its neighbors are used to compute smooth spatio-temporal dissimilarity map voxel values (Fig. 1).

Within a neighborhood, n voxels can be weighted differently in the computation of the dissimilarity value. A Gaussian weighting scheme gives more weight to the central voxel and less weight to the neighboring voxels. The farther away the neighbors are from the center, the smaller the weight assigned. A Gaussian pattern of weights is characterized by its standard deviation, expressed in terms of voxel dimension. In a two-dimensional case:

$$\Omega_\sigma(x,y) = \frac{1}{2\pi\sigma^2}\exp\left(-\frac{x^2+y^2}{2\sigma^2}\right), \tag{2}$$

where standard deviation σ determines the amount of smoothing.

There are two main reasons to include neighbors in the computation of a voxel-based dissimilarity measure. The first reason is to average out high spatial frequency noise in the data. Areas of group difference with greater spatial extent will be kept, and those with small spatial extent eliminated. The second reason is to address variations in anatomy among subjects. Data across subjects are aggregated; due to variation in anatomy, functional activations may not overlap exactly.

Similarity measure and the RV-coefficient

The general framework of the proposed method requires a selection of dissimilarity measure. We adapt the RV-coefficient, first introduced by Robert and Escoufier (1976), for comparison of time series data at a specified spatial region between two subjects. The RV-coefficient has been previously applied to fMRI data for subject selection by Kherif et al. (2003).

The RV-coefficient is an ideal measure of temporal similarity between signals because it is a multivariate extension of the Pearson correlation coefficient and indicates the overall similarity of two matrices.

Let matrices $\mathcal{G}_{p\times T}$ and $\mathcal{H}_{q\times T}$ define two configurations of points $\mathcal{C}(\mathcal{G})$ and $\mathcal{C}(\mathcal{H})$ in $\Re^p$ and $\Re^q$ respectively, where p and q are the number of voxels, and T is the number of time points. Define G and H to be two mean centered matrices of the original data, such that every time series is mean centered and the $(i,j)^{th}$ element of G is $(g_{ij}-\bar{g}_j)$. Without loss of generality, let $p=q=n$, where n is a subset of image voxels. This subset can be as small as a single voxel, for a voxel-by-voxel comparison of two time series. The RV-coefficient measures the closeness of these two configurations (Robert and Escoufier, 1976):

$$\text{RV}(G,H) = \frac{\text{tr}(G'GH'H)}{\sqrt{\text{tr}(G'GG'G)\text{tr}(H'HH'H)}} = \frac{\text{tr}(GH'HG')}{\sqrt{\text{tr}(GG')^2\text{tr}(HH')^2}}. \tag{3}$$

The scalar product between $G'G$ and $H'H$ represents a generalized covariance between G and H. Denote the scalar product $\text{tr}(G'GH'H)$ by $\langle G'G,H'H\rangle$. Greater values of $\langle G'G,H'H\rangle$, relative to $\langle G'G,G'G\rangle$ and $\langle H'H,H'H\rangle$, indicate more similarity between $G'G$ and $H'H$ in terms of raw product distances. Thus, larger RV-coefficient values indicate higher similarity (high degree

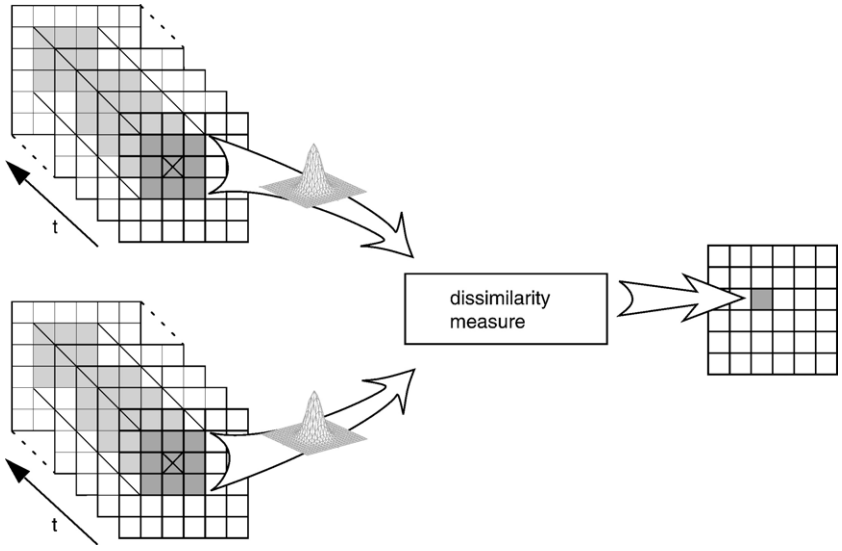

Fig. 1. Smooth spatio-temporal dissimilarity map construction is illustrated with 2D data. A mask (3 × 3 in this illustration) is applied at each pixel. Based on the time series data selected by the mask, the dissimilarity value is recorded at the location of the central pixel in the output array.

of co-structure) between configurations. Reexpressing the matrix formulation of the RV-coefficient, we derive:

$$\mathrm{RV}(G,H)=\frac{\sum_{j=1}^{n}\sum_{i=1}^{n}\left(\sum_{t=1}^{T}g_{it}h_{jt}\right)^2}{\sqrt{\sum_{j=1}^{n}\sum_{i=1}^{n}\left(\sum_{t=1}^{T}g_{it}g_{jt}\right)^2\sum_{j=1}^{n}\sum_{i=1}^{n}\left(\sum_{t=1}^{T}h_{it}h_{jt}\right)^2}}. \tag{4}$$

which is a normalized cross-covariance measure. Note that the numerator gives a measure of temporal similarity between all pairs of time series at $\{g_i\}$ and $\{h_i\}$. The RV-coefficient is defined on [0,1].

One distance function between G and H relative to an experimental task is an absolute difference between the RV-coefficient of G with W_G and the RV-coefficient of H with W_H:

$$d=|\mathrm{RV}(G,W_G)-\mathrm{RV}(H,W_H)|. \tag{5}$$

The reasoning behind considering the absolute difference of two RV-coefficients is that the RV-coefficient between the stimuli and neuronal activity is expected to be high in activated areas and low in areas with no activation. Therefore, the absolute difference of the two RV-coefficients will be low for regions where activation is present in both time series data sets or where both time series data sets contain noise. Regions where one time series data set contain signal and another time series data set contain noise will have a high absolute difference of the RV-coefficients. In addition, d allows us to compare subjects with different stimulus presentation sequence.

There are similarities between the proposed approach and the general linear model when dissimilarity measure is based on RV-coefficient. Following the above approach, the comparison between subjects is made on a localized region which consists of a set of voxels. For a special case where the localized region consists of a single voxel, d in Eq. (5) is equivalent to the general linear model. However, d does not make any distributional assumptions about the data, although we do make an assumption of exchangeability for permuting subject's labels at the thresholding step. The general framework allows the user to specify a dissimilarity measure of choice. The method based on RV-coefficient dissimilarity measure is expected to be spatially sensitive and powerful because it extracts common information from neighboring voxels and uses the common information for classification.

Feature selection algorithm

The goal of feature selection is to identify areas where groups differ relative to stimulus presentation. The algorithm searches for areas that are most dissimilar between the two groups with respect to time series data, according to the criterion defined in Eq. (1). Data from group A are labeled $A_1, \ldots, A_{n_A}$, and data from group B are labeled $B_1, \ldots, B_{n_B}$. The feature selection algorithm, based on a smooth spatio-temporal dissimilarity map, proceeds as follows:

1. For each voxel (x_0, y_0, z_0), construct the D matrix and compute the observed test statistic. For example:

$$d(x_0,y_0,z_0)_{\mathrm{obs}}=\frac{1}{n_An_B}\sum_{i=1}^{n_A}\sum_{j=1}^{n_B}\left|\mathrm{RV}_\Omega(A_i,W_{A_i})-\mathrm{RV}_\Omega\left(B_j,W_{B_j}\right)\right|, \tag{6}$$

   where Ω represents a weighted combination of the voxel and its neighbors that are used in the computation. The map of observed test statistics is referred to as a smooth spatio-temporal dissimilarity map.
2. The value of $d(x_0, y_0, z_0)_{\mathrm{obs}}$ is compared to the distribution of $d(x_0, y_0, z_0)$ under the null hypothesis of no group differences in temporal similarity, and permutation-based p values for each voxel of the spatio-temporal dissimilarity map are computed.
3. Threshold the map of p values from step 2 using a false discovery rate method.

The output of the algorithm is a reduced data set of n_r spatial locations from voxels where the difference between the two groups is significant. To classify future subjects, we use only the time series from the training data associated with these n_r voxel locations selected by the feature extraction step.

This method is easily extended to more than two groups. Suppose that our data consist of subjects from G groups and the images in group g are labeled $A_1^g, \ldots, A_{n_g}^g$, $g=1, \ldots G$. Let (g_1, g_2) be a pair of groups. For each of the pairings of two groups chosen from the G groups, we compute

$$d(x_0,y_0,z_0)_{g_1,g_2} = \frac{1}{n_{g_1},n_{g_2}} \sum_{i_1}^{n_{g_1}} \sum_{i_2}^{n_{g_2}} |\mathrm{RV}\left(A_{i_1}^{g_1}, W_{A_{i_1}^{g_1}}\right) - \mathrm{RV}\left(A_{i_2}^{g_2}, W_{A_{i_2}^{g_2}}\right)|.$$

Finally, for each voxel, we simply add the computed distances for each pair

$$\hat{\mathrm{d}}(x_0,y_0,z_0) = \sum_{g_1=1}^{G-1} \sum_{g_2=g_1+1}^{G} d(x_0,y_0,z_0)_{g_1,g_2}$$

Classification

With certain areas identified in the feature selection step, we are now ready to classify. For consistency, we will use the same similarity measure from the feature selection step. Let E be a new data set that is to be classified into one of the two previously defined groups, A or B. Let $\bar{A}$ and $\bar{B}$ be the average time series data of groups A and B, respectively. A subset of the new subject's time series data E is compared to subsets of $\bar{A}$ and $\bar{B}$ identified in the feature selection step. The same similarity measure used in the dimension reduction stage is computed between selected subsets for E and $\bar{A}$, and E and $\bar{B}$. The new subject E is classified into group A if $\mathrm{RV}(E, \bar{A}) \geq \mathrm{RV}(E, \bar{B})$. Otherwise, E is classified into group B.

In the general case of G groups, a new subject E is assigned to group g^* where

$$g^* = \mathrm{argmax}_{g=1,\ldots,G} |\mathrm{RV}(E, \bar{A}^g)|.$$

Comparing voxels-by-time matrices is only meaningful for cases where all subjects were presented with the same stimuli. In other cases, where stimulus presentation is randomized, a new subject E is assigned to group g^* where

$$g^* = \mathrm{argmin}_{g=1,\ldots,G} \sum_{i=1}^{n_g} |\mathrm{RV}(A_i^g, W_i^g) - \mathrm{RV}(E, W_E)|.$$

It is important to consider that classification accuracy could be limited by potential noise and artifact sources in the data. A classifier could pick out such artifacts as important features to be used in classification. The data quality must be assessed prior to use in this analysis, using a procedure similar to that described by Stöcker et al. (2005).

Results

Simulated data

Feature selection based on smooth spatio-temporal dissimilarity maps and classification performance were evaluated on simulated data. To ensure a simulation with realistic noise, a task waveform was superimposed on actual fMRI scans of a single subject taken during a rest condition. All computations were performed using Matlab 7.0 (http://www.mathworks.com).

The data were acquired on a 3 T Siemens Allegra scanner in accordance with the Institutional Review Board at the University of Illinois at Urbana-Champaign. A single-shot EPI sequence was used with a TE/TR of 25/2000 ms, 32 slices, FOV of 22 cm, matrix size of 64×64, and 150 acquisitions per run. Ten of these runs were acquired with the subject in a resting state. Each run in the simulation was considered as if it were obtained from a different subject. The data were realigned and motion-corrected using SPM99 (http://www.fil.ion.ucl.ac.uk/spm). For evaluation purposes, only a 2D simulation was performed. Thus, only a single slice of the realigned data sets was extracted. A task waveform for an event-related design was generated using the optseq2 program (http://surfer.nmr.mgh.harvard.edu/optseq/; Dale, 1999; Dale et al., 1999). The amplitude of the stimulus waveforms was scaled to be 5% of the mean of the brain region to which it was added. The task waveform was added to an ROI of 24 pixels in the images that form group B. Images that form group A had no signal added. For illustrative purposes, Fig. 2 shows extracted time series from a voxel outside the ROI and from a voxel within the ROI from representative images in groups A and B. Notice that visually determining which time series has a signal added is a challenging task. Next, images in groups A and B were smoothed by a Gaussian filter with a full-width at half maximum (FWHM) of 2 voxels. Finally, each pixel's mean time course was subtracted.

To examine the effect of neighborhood size smooth spatio-temporal dissimilarity, maps were constructed for Gaussian weights defined on 3×3 and 5×5 neighborhoods. Only voxels inside the brain volume were used in the computation. The D matrix was permuted 1000 times, and smooth spatio-temporal dissimilarity maps were thresholded using a 5% false discovery rate. The upper panel of Fig. 3 shows smooth spatio-temporal dissimilarity maps constructed with 3×3 and 5×5 Gaussian weights. The lower panel shows EPI data with superimposed thresholded p values. In both cases, the area with an added signal was correctly identified. There is a tradeoff between the signal and spatial resolution. The optimal filter width would match the spatial extent of the activation area.

In the next simulation, classification performance was evaluated using cross-validation technique as a function of percent change of baseline for different values of false discovery rate q. We would expect classification accuracy to increase monotonically with percent change of baseline.

For each value of percent signal change and false discovery rate q, one subject's time series data were left out for testing purposes. Based on the remaining subjects' data, regions with highest dissimilarity were identified using feature selection based on the spatio-temporal dissimilarity map, with 1000 permutations of the D matrix. The left-out subject was then classified into one of the two groups. The procedure was repeated 10 times, based on the number of available data sets, and classification rate was determined and recorded for each of the two groups.

The results of the simulation study examining classification accuracy as a function of percent change of baseline for different

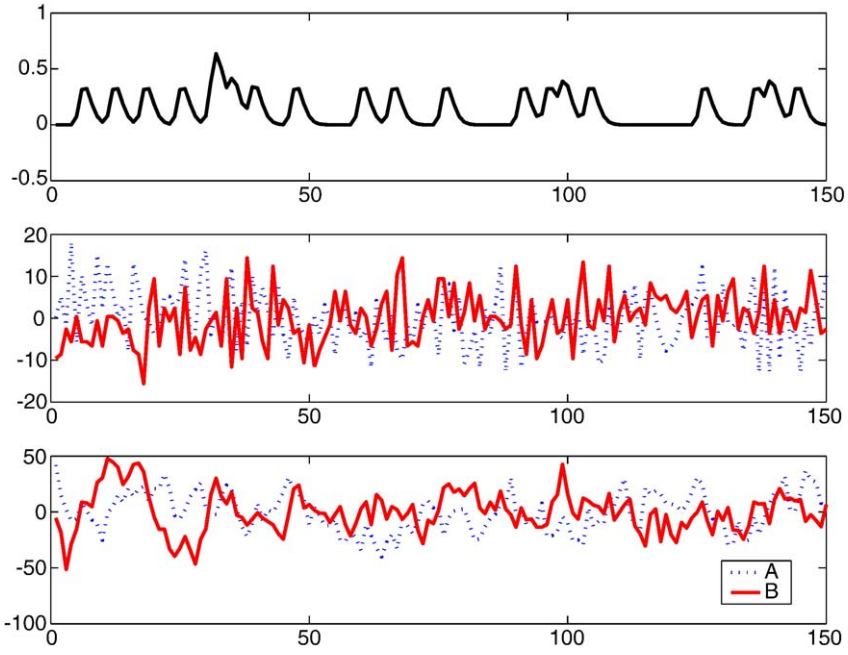

Fig. 2. Task waveform for an event-related design (top); sample time series from a pixel (20, 20) outside the ROI (middle) and a pixel (46, 35) inside the ROI (bottom) from representative images in groups *A* (no task waveform added) and *B* (task waveform added). Time series are mean centered.

values of q are shown in Fig. 4. As expected, classification rate monotonically increases with increase in percent baseline change. Classification accuracy for group *B* reaches 100% for 2% baseline change. Stringent choices of q result in more false classifications (type II error) since fewer regions are selected, even when true differences exist, giving a higher sensitivity to noise. When q is

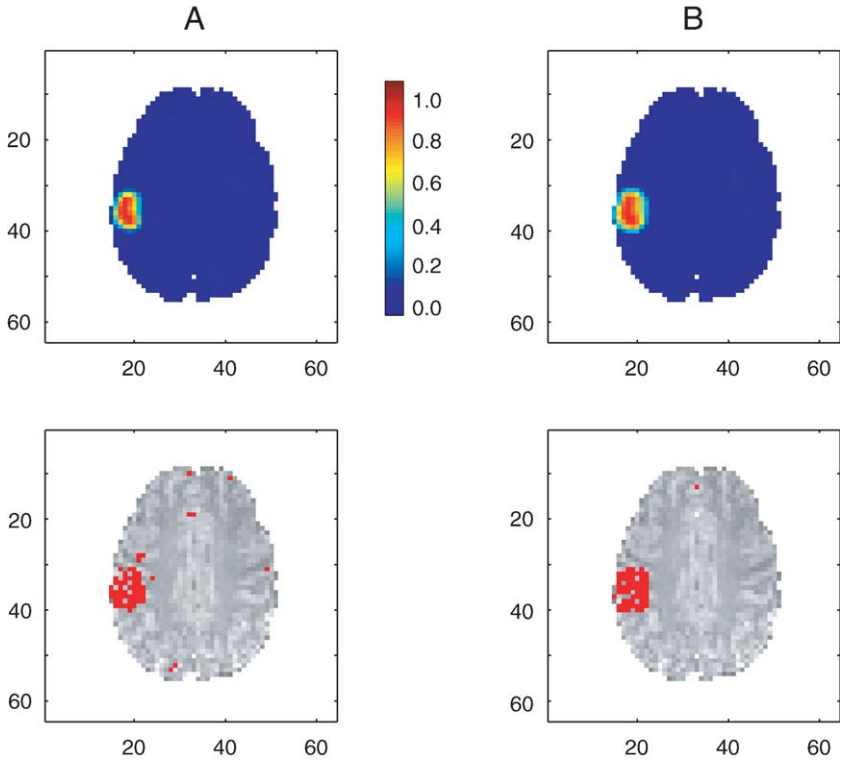

Fig. 3. The upper panel shows smooth spatio-temporal dissimilarity maps constructed with panel A 3×3 and panel B 5×5 Gaussian filters. The color scale indicates the magnitude of the dissimilarity values. The lower panel shows locations of significant group differences based on 1000 permutations and thresholded with FDR.

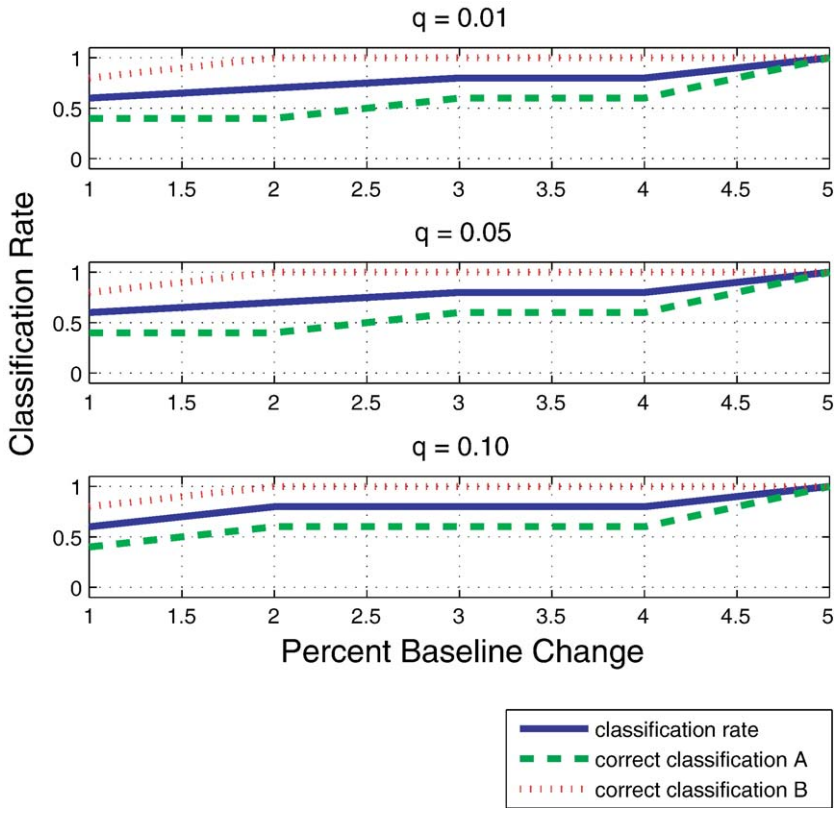

Fig. 4. Cross-validation results for realistic fMRI noise data. Classification accuracy for different values of percent signal change and false discovery rate is based on spatio-temporal dissimilarity map feature selection constructed with a 3 × 3 Gaussian filter and 1000 permutations of the *D* matrix.

large, more regions are selected, including those that are similar (type I error).

In vivo

Data description

We illustrate the methodology on a subset of individuals from a study reported by Mohanty et al. (2005) consisting of 16 subjects with high scores on a positive schizotypy behavioral measure and 16 low scoring controls (Chapman et al., 1978; Eckblad and Chapman, 1983). Subjects performed an emotional Stroop task which involves the simultaneous presentation of task-relevant (color of letters) and task-irrelevant (emotional meaning) attributes. The task is to identify the ink color of a word as quickly as possible while ignoring the word's meaning. Performance depends on selective attention to task-relevant vs. task-irrelevant stimulus features and maintenance of contextual information (e.g., Cohen and Servan-Schreiber, 1992). Each participant performed a block-design emotional Stroop task consisting of blocks of positive or negative emotion words alternating with blocks of neutral words, while 445 echo-planar (EPI) images (TR 1517 ms, echo time TE 40 ms, flip angle 90°, 15 contiguous slices 7 mm thick, no gap, in-plane resolution = 3.75 mm^2) were acquired in a 1.5 T GE Signa scanner, in accordance with the institutional review board at the University of Illinois. High-resolution three-dimensional (3D) anatomical images (T1-weighted 3D gradient echo images) of the whole brain were collected for each participant for landmark selection. T1-weighted anatomical images of the 15 functional acquisition slices were also collected for image registration purposes.

Previous findings

Mohanty et al. (2005) investigated the hypothesis that, while performing the emotional Stroop task, positive-schizotypy individuals would show abnormal activation in regions involved in emotional processing as well as maintenance of attentional set in the presence of emotional distractors. Among the cortical regions, the schizotypy group showed greater activation in the right middle frontal gyrus (Brodmann area 9),

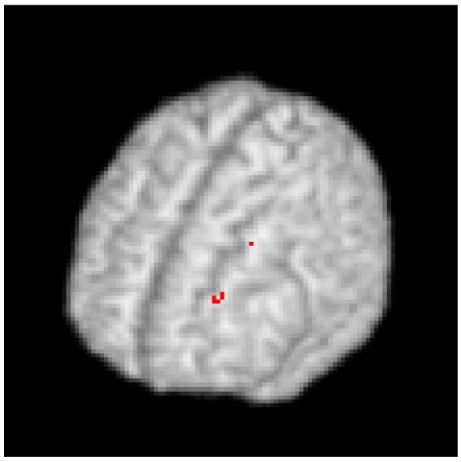

Fig. 5. Voxels selected based on spatio-temporal dissimilarity map feature selection for a representative subject. Voxels selected are in left middle frontal gyrus and left superior frontal gyrus.

right inferior frontal gyrus (Brodmann area 46), rostral anterior cingulate gyrus (BA 32), and left superior and inferior parietal lobe for negative than for neutral words. For the same contrast, the control group showed greater activation in the left middle frontal gyrus, left superior frontal gyrus, right inferior temporal gyrus, and right middle occipital gyrus. Since the dorsolateral prefrontal cortex (DLPFC) plays a critical role in maintenance of attentional set in the presence of emotional and non-emotional distractors (Compton et al., 2003), the decreased left DLPFC activity (i.e., middle frontal gyrus, BA 9) in schizotypy participants is consistent with the hypothesis that these individuals have deficits in the ability to maintain attentional set in the presence of aversive distractors.

Classification results

Participants were classified as high schizotypes or controls based on functional data, and the classification was compared to the behavioral test classification. Fourteen participants with the same order of stimulus presentation – seven in each group – were included in the analysis. Schizophrenia spectrum individuals exhibit problems in maintaining context or attentional set and monitoring for the occurrence of conflict within the attentional network (Braver et al., 1999; van Veen and Carter, 2002; Mohanty et al., 2005). These functions are implemented in frontal brain areas including DLPFC and anterior cingulate cortex (ACC) (Milham et al., 2001; MacDonald et al., 2000). Hence, most differences between schizotypes and controls for emotional Stroop task are expected to occur in frontal areas. For this reason and for computational considerations, we classified participants into schizotypes and controls based on functional activity in frontal areas.

Classification accuracy was evaluated by the leave-one-out method. For each subject that was excluded from the analysis, smooth spatio-temporal dissimilarity and p value maps were computed based on the remaining subjects (i.e., training data). Maps of p values were thresholded using 5% FDR. Identified features (voxels) were used for classification of the left-out subject. From the training data, average voxel-by-time matrices were computed for each of the two groups based on the selected features. The voxel-by-time matrix for the new subject was compared to group mean matrices using the RV-coefficient, and the subject was classified into the group with the larger RV-coefficient. The procedure was repeated for all subjects. Average classification accuracy was 85.71%, 6/7 for each group. Two frontal areas, DLPFC and ACC emerged as important in the classification process. Selected voxels for a representative subject are shown in Fig. 5.

Discussion

There is considerable literature indicating that schizophrenia spectrum individuals show problems in processing of context including deficits in maintenance of attentional set and monitoring of conflict in the attentional network (e.g., Braver et al., 1999). This impairment is attributed to abnormal activity in DLPFC and ACC (Barch et al., 2001; MacDonald et al., 2003; Mohanty et al., 2005). Results from the present study confirm that there are functional differences between schizotypy individuals and controls in the DLPFC and ACC. Selected areas where the temporal dissimilarity is large between the two groups are similar to those identified in Mohanty et al. (2005).

Conclusion

This paper presented a unified feature selection and classification procedure for classifying subjects into groups based on four dimensional spatio-temporal data. Unlike previous approaches, the present approach offers the ability to locate spatial regions with temporal differences between groups. The proposed method simultaneously accounts for and identifies intergroup spatial and temporal variability. It uses temporal similarity as a spatially localized measure of similarity, and it incorporates a permutation framework for significance testing.

To our knowledge, this is the first nonparametric feature selection and classification method that uses, simultaneously, localized spatial and temporal information of fMRI based on the whole brain volume. The proposed methodology shares similarities with a general linear model approach, but it departs from the traditional application by not making distributional assumptions, hence we expect this to be robust and to perform well under very general conditions. The methodology allows the user to specify a dissimilarity measure. Moreover, it can be easily adapted so that other classification methods can be incorporated into the methodology.

The simulation studies show that the proposed method is able to accurately identify regions of dissimilarity and gives low misclassification rates. Finally, an analysis of an *in vivo* data set demonstrates that the method gives results that are consistent with previous findings and provides the first demonstration of classification of schizotypes and controls based on their hemodynamic time series data.

Acknowledgments

We thank the anonymous reviewers for helpful comments on the earlier version of the manuscript.

References

Barch, D.M., Carter, C.S., Braver, T.S., Sabb, F., MacDonald III, A.W., Noll, D.C., Cohen, J.D., 2001. Selective deficits in prefrontal cortex function in medication-naive patients with schizophrenia. Arch. Gen. Psychiatry 58, 280–288.

Benjamini, Y., Hochberg, Y., 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. R. Stat. Soc., Ser. B Methodol. 57, 289–300.

Bogorodzki, P., Rogowska, J., Yurgelun-Todd, D.A., 2005. Structural group classification technique based on regional fMRI BOLD responses. IEEE Trans. Med. Imaging 24, 389–398.

Braver, T.S., Barch, D.M., Cohen, J.D., 1999. Cognition and control in schizophrenia: a computational model of dopamine and prefrontal function. Biol. Psychiatry 46, 312–328.

Chapman, L.J., Chapman, J.P., Raulin, M.L., 1978. Body-image aberration in schizophrenia. J. Abnorm. Psychol. 87, 399–407.

Chetelat, G., Eustache, F., Viader, F., De La Sayette, V., Pelerin, A., Mezenge, F., et al., 2005. FDG-PET measurement is more accurate than neuropsychological assessments to predict global cognitive deterioration in patients with mild cognitive impairment. Neurocase 11, 14–25.

Cohen, J.D., Servan-Schreiber, D., 1992. Context, cortex, and dopamine: a connectionist approach to behavior and biology in schizophrenia. Psychol. Rev. 99, 45–77.

Compton, R.J., Banich, M.T., Mohanty, A., Milham, M.P., Herrington, J., Miller, G.A., et al., 2003. Paying attention to emotion: an fMRI

investigation of cognitive and emotional Stroop tasks. Cogn. Affect. Behav. Neurosci. 3, 81–96.
Dale, A.M., 1999. Optimal experimental design for event-related fMRI. Hum. Brain Mapp. 8, 109–114.
Dale, A.M., Greve, D.N., Burock, M.A., 1999. Optimal stimulus sequences for event-related fMRI. 5th International Conference on Functional Mapping of the Human Brain. Duesseldorf, Germany. June 11–16.
Eckblad, M., Chapman, L.J., 1983. Magical ideation as an indicator of schizotypy. J. Consult. Clin. Psychol. 51, 215–225.
Fisher, R., 1935. The Design of Experiments, Section 21. Oliver and Boyd Ltd., Edinburgh.
Ford, J., Farid, H., Makedon, F., Flashman, L.A., McAllister, W., Megalooikonomou, V., et al., 2003. Patient classification of fMRI activation maps. Paper presented at the 6th Annual International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI'03).
Gamalo, M., Ombao, H., Jennings, R., 2005. Comparing extent of activation: a robust permutation approach. NeuroImage 24, 715–722.
Genovese, C., Lazar, N.A., Nichols, T.E., 2002. Thresholding of statistical maps in functional neuroimaging using the false discovery rate. NeuroImage 15, 870–878.
Hayasaka, S., Nichols, T.E., 2003. Validating cluster size inference: random field and permutation methods. NeuroImage 20, 2343–2356.
Hubert, L., 1987. Assignment Methods in Combinatorial Data Analysis. M. Dekker, New York.
Jueptner, M., Weiller, C., 1995. Review: does measurement of regional cerebral blood flow reflect synaptic activity?—Implications for PET and fMRI. NeuroImage 2, 148–156.
Kherif, F., Poline, J., Meriaux, S., Benali, H., Flandin, G., Brett, M., 2003. Group analysis in functional neuroimaging: selecting subjects using similarity measures. NeuroImage 20, 2197–2208.
Kontos, D., Megalooikonomou, V., Ghubade, N., Faloutsos, C., 2003. Detecting discriminative functional MRI activation patterns using space filling curves. Paper presented at the Proceedings of the 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Cancun, Mexico.
Liow, J., Rehm, K., Strother, S., Anderson, J., Morch, N., Hansen, L.K., et al., 2000. Comparison of voxel- and volume-of-interest-based analysis in FDG PET scans of HIV positive and healthy individuals. J. Nucl. Med. 41, 612–621.
MacDonald III, A.W., Cohen, J.D., Stenger, V.A., Carter, C.S., 2000. Dissociating the role of the dorsolateral prefrontal and anterior cingulate cortex in cognitive control. Science 288, 1835–1838 (June 9).
MacDonald III, A.W., Johnson, M.K., Becker, T.M., Carter, C.S., 2003. Context processing deficits associated with hypofrontality in the healthy relatives of schizophrenia patients: an event related fMRI study. Paper Presented at the IXth Meeting of International Congress of Schizophrenia Research, Colorado Springs, CO.
Milham, M.P., Banich, M.T., Webb, A., Barad, V., Cohen, N.J., Wszalek, T., Kramer, A.F., 2001. The relative involvement of anterior cingulated and prefrontal cortex in attentional control depends on nature of conflict. Cogn. Brain Res. 12, 467–473.
Mitchell, T.M., Hutchinson, R., Niculescu, R., Pereira, F., Wang, X., 2004. Learning to decode cognitive states from brain imaging. Mach. Learn. 57, 145–175.
Mohanty, A., Herrington, J.D., Koven, N.S., Fisher, J.E., Wenzel, E.A., Webb, A.G., et al., 2005. Neural mechanisms of affective interference in schizotypy. J. Abnorm. Psychol. 114, 16–27.
Nichols, T.E., Holmes, A.P., 2001. Nonparametric permutation tests for functional neuroimaging: a primer with examples. Hum. Brain Mapp. 15, 1–25.
Ogawa, S., Lee, T.M., Kay, A.R., Tank, D.W., 1990. Brain magnetic resonance imaging with contrast dependent on blood oxygenation. Proc. Natl. Acad. Sci. 87, 9868–9872.
Pitman, E.J.G., 1937. Significance tests which may be applied to samples from any populations. Suppl. J. R. Stat. Soc. 4, 119–130.
Pokrajac, D., Megalooikonomou, V., Lazarevic, A., Kontos, D., Obradovic, Z., 2005. Applying spatial distribution analysis techniques to classification of 3D medical images. Artif. Intell. Med. 33, 261–280.
Raz, J., Zheng, H., Ombao, H., Turetsky, B., 2003. Statistical tests for fMRI based on experimental randomization. NeuroImage 19, 226–232.
Robert, P., Escoufier, Y., 1976. A unifying tool for linear multivariate statistical methods: the RV-coefficient. Appl. Stat. 25, 257–265.
Stöcker, T., Schneider, F., Klein, M., Habel, U., Kellermann, T., Zilles, K., Shah, N., 2005. Automated quality assurance routines for fMRI data applied to a multicenter study. Hum. Brain Mapp. 25, 237–246.
van Veen, V., Carter, C.S., 2002. The anterior cingulate as a conflict monitor: fMRI and ERP studies. Physiol. Behav. 77, 477–482.
Wang, Q., Kontos, D., Li, G., Megalooikonomou, V., 2004. Application of time series techniques to data mining and analysis of spatial patterns in 3D images. Proceedings of the 2004 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2004), Montreal, Canada, pp. 525–528.
Wasserman, S., Böckenholt, U., 1989. Bootstrapping: applications to psychophysiology. Psychophysiology 26, 208–221.
Zhang, L., Samaras, D., Tomasi, D., Volkow, N., Goldstein, R., 2005. Machine learning for clinical diagnosis from functional magnetic resonance imaging. IEEE Proc. Comput. Vis. Pattern Recogn. 1, 1211–1217.