Dynamic adjustment of stimuli in real time fMRI

I Jung Feng[1], Anthony Jack[2], Curtis Tatsuoka[3,1]

[1] Department of Epidemiology and Biostatistics, [2]Department of Cognitive Science, [3]Department of Neurology

Case Western Reserve University, 10900 Euclid Ave., Cleveland, OH 44106

ABSTRACT

Conventional functional magnetic resonance imaging (fMRI) image analysis is performed by carrying out a massive number of parallel regression analyses. fMRI signal is known for its low signal-to-noise ratio (SNR), and its complexity, such as reflected by spatial and temporal autocorrelation. In order to ensure accurate localization of brain activity, stimuli administration in an fMRI session is often lengthy and repetitive. In real time fMRI (rtfMRI), signal processing is carried out while the signal is being observed. This method allows for the dynamic adjustment of stimuli through sequential experimental designs. We have developed a voxel-wise sequential probability ratio test (voxel-wise SPRT) approach[1] for dynamically localizing activation associated with stimuli, as well as decision rules for the stopping of experimentation. Stopping is dynamically determined when sufficient statistical evidence is collected to assess the activation status of voxels across regions of interest (ROIs). Simulation studies show that the number of scan units can be reduced substantially compared to standard fMRI experimental designs that are fixed and predetermined, while still achieving comparably high levels of detection accuracy. An analysis based on actual brain imaging confirms the promise of the presented approach.

INTRODUCTION

Functional neuroimaging technology is widely used for clinical applications. fMRI provides neural images with high spatial resolution by non-invasively detecting task-related blood oxygen level dependent (BOLD) signal changes which are associated with neural activity in the brain. Importantly, then, fMRI allows for precise localization of brain functioning. However, fMRI signals present analytical challenges because they are abundant, noisy and highly correlated, both spatially and temporally. The task related signal, or what investigators want to detect, is only 0.5 to 2% signal change in measured BOLD signal. In order to ensure accurate spatial localization, a pre-determined, redundant and lengthy fMRI session is usually implemented. This leads to high costs for fMRI implementation, and exposes the signal data to fatigue and learning effects.

With progress in fMRI acquisition and computational processing, it has become feasible to observe brain activity during experimentation. This advance is known as real-time fMRI. This advent of real-time processing of BOLD signals, creates an important opportunity: the ability to adapt experimental stimuli in real-time according to individual response and variability. Through real-time signal processing, when the activation status of voxels within a ROI becomes clear, experimentation can be terminated early without having to fully proceed through a pre-determined sequence of stimuli. As a result, a precise detection of neural activity can be obtained in less fMRI scan time units as compared to the number of scan units required in conventional fMRI experimentation. An innovative statistical method, a dynamic adjustment of stimuli via sequential analysis, applied on analyzing real time fMRI signal, is proposed.

METHODS

General linear model

The most common fMRI statistical analysis approach is by using voxel-wise general linear models (GLM)[2], to see how well observed BOLD response associates with expected BOLD response from stimuli. Voxels that are activated by a task are identified through conducting statistical inference on a task-related regression parameter. For a given voxel, the GLM is:

Y=XB+E

where $\boldsymbol{Y}$ is a t × 1 vector of measured BOLD signal intensities of the voxel over time, and $\boldsymbol{E}$, a t × 1 vector, represents the error components. $\boldsymbol{X}$ is a t × K design matrix including the expected BOLD signal change generated by convolving hemodynamic response function (HRF) and implemented tasks stimuli function. $\boldsymbol{B}$ equals $[b_1\ b_2 \ldots\ b_K]^T$, a K × 1 regression coefficients vector that includes those parameters that are task-related. The t × 1 error matrix, $\boldsymbol{E}$, is assumed to be normally distributed with mean zero and variance $\sigma^2\boldsymbol{V}$, where σ^2 is amount of temporal variance and $\boldsymbol{V}$, a t × t matrix, represents temporal autocorrelation structure. $\boldsymbol{Y}$ is assumed with multivariate normal probability distribution as follows:

$$f\left(\boldsymbol{Y},\boldsymbol{B},\sigma^2\boldsymbol{V}\right)=\frac{1}{\left(2\pi\right)^{t/2}\left|\sigma^2\boldsymbol{V}\right|}\exp\left(-\frac{1}{2}\left(\boldsymbol{Y}-\boldsymbol{XB}\right)'\left(\sigma^2\boldsymbol{V}\right)^{-1}\left(\boldsymbol{Y}-\boldsymbol{XB}\right)\right)$$

where $|\sigma^2\boldsymbol{V}|$ is the determinant of $\sigma^2\boldsymbol{V}$. Using generalized least square (GLS) estimation of regression parameters and corresponding variance values, one is able to be made inferences about a voxel's active status through testing of hypotheses, as will be described next.

One sided voxel-wise SPRT

We consider SPRT for one sided hypothesis test for contrast of voxel-specific regression parameters from GLM models. Two-sided analogues are similar. The general form of one sided hypotheses is

$$H_0\text{: } \boldsymbol{cB}-\theta_0=0 \text{ versus } H_a\text{:}\boldsymbol{cB}-\theta_0\geq\delta, \qquad (1)$$

where $\boldsymbol{c}$ equals $[c_1\ c_2\ \ldots c_k\ \ldots c_K]$, a K × 1 contrast vector. $\boldsymbol{cB}$ is a linear combination of corresponding coefficients. δ is considered with practical important difference from θ_0. For instance, the hypothesis test in (1) can be of the form $H_0: b_k = 0$ against $H_a: b_k \geq 1$, and hence represents the test of whether or not a voxel is activating in association with task k when the corresponding regressors of b_k representing expected HRF activated by task k and 1 is assumed with practical difference from 0. Other comparisons of multiple task activations can be represented by linear combinations of regression parameters associated across several tasks.

$\boldsymbol{cB}$ is distributed normally with mean $E(\boldsymbol{cB})$ and variance $Var(\boldsymbol{cB})$. The statistics for one sided SPRT is a likelihood of $E(\boldsymbol{cB})$ given $\boldsymbol{cB} = \theta_1$ divided by likelihood of $E(\boldsymbol{cB})$ given $\boldsymbol{cB} = \theta_0$. The formula of likelihood ratio is:

$$\begin{aligned}\Lambda_t &= \log\left(\frac{f\left(\boldsymbol{cB}\,|\,\theta_1\right)}{f\left(\boldsymbol{cB}\,|\,\theta_0\right)}\right)\\ &= \log\left(\frac{\frac{1}{\left(2\pi\right)^{1/2}\left|Var\left(\boldsymbol{cB}\right)\right|^{1/2}}\exp\left(\frac{-1}{2}\left(\left(\boldsymbol{cB}-\theta_1\right)'Var\left(\boldsymbol{cB}\right)^{-1}\left(\boldsymbol{cB}-\theta_1\right)\right)\right)}{\frac{1}{\left(2\pi\right)^{1/2}\left|Var\left(\boldsymbol{cB}\right)\right|^{1/2}}\exp\left(\frac{-1}{2}\left(\left(\boldsymbol{cB}-\theta_0\right)'Var\left(\boldsymbol{cB}\right)^{-1}\left(\boldsymbol{cB}-\theta_0\right)\right)\right)}\right)\\ &= \frac{1}{2}\left\{\left(\left(\boldsymbol{cB}-\theta_0\right)'Var\left(\boldsymbol{cB}_n\right)^{-1}\left(\boldsymbol{cB}-\theta_0\right)\right)-\left(\left(\boldsymbol{cB}-\theta_1\right)'Var\left(\boldsymbol{cB}\right)^{-1}\left(\boldsymbol{cB}-\theta_1\right)\right)\right\}\end{aligned} \qquad (2)$$

These unknown parameters vary in some practical situations. For example, $E(\boldsymbol{cB})$ and variance $Var(\boldsymbol{cB})$ differ from person to person and change between various fMRI experimental designs. Therefore, according to Cox's work[3], unknown parameters $E(\boldsymbol{cB})$ and $Var(\boldsymbol{cB})$ are replaced by corresponding maximum likelihood estimators (MLEs) computed by the following equations:

$$\boldsymbol{c}\hat{\boldsymbol{B}}_t=\boldsymbol{c}\left(\left(\boldsymbol{X}'\boldsymbol{V}\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{V}\boldsymbol{Y}_{t\times1}\right) \qquad (3)$$

$$Var\left(\boldsymbol{c}\hat{\boldsymbol{B}}_t\right)=Var\left(\boldsymbol{Y}_{t\times1}\right)\times\boldsymbol{c}\left(\boldsymbol{X}'\boldsymbol{V}\boldsymbol{X}\right)^{-1}\boldsymbol{c}' \qquad (4)$$

The temporal autocorrelation structure, $\boldsymbol{V}$, is assumed known.

Consider how stopping occurs for one voxel. Horizontal stopping boundaries are employed. Based on user defined type I error α and type II error β, the decision is made by following rejection/acceptance rules after collecting one scan image at scan time point t. These rules are defined as:

1. Continue sampling when $B < \Lambda_t < A$
2. Stop sampling and accept H_0 when $\Lambda_t < B$
3. Stop sampling and accept H_a when $A < \Lambda_t$

where stopping boundaries (A,B) = ($\log((1-\beta)/\alpha)$, $\log(\beta/(1-\alpha))$). (5)

Multiple comparison correction[4]

Bonferroni correction is used in this study, as these corrections are easily reflected in the stopping boundary specifications. The Type I and Type II errors for a single voxel's hypothesis are modified as $\alpha_n = \alpha/N$ and $\beta_n = \beta/N$, where N is the total number of hypotheses (voxels) being considered at once.

Global stopping rule

Since we are possibly testing a large number of voxels at once, a "global" decision is needed on when to stop, based on aggregate performance of SPRTs across the tests. For efficient stopping, we suggest a rule that stops when a pre-determined, user-defined percentage of tests satisfy the stopping criteria of (5). Note that when a parameter value lies in a range between the null and alternative hypotheses, it is harder to decide between them, so that longer scan times are needed. Ironically, it is for these values that there is the greatest indifference as to whether the null or alternative is decided upon. A key to success for this approach is thus to choose a pre-determined percentage that is reflective of the number of voxels with activation levels that don't lie in the mid-range. The process of global stopping rules is:

1. Predetermine a targeted level of G% for acceptable percentage of voxels with decisive detection on activation
2. Stop fMRI scanning if at least G% of voxels satisfies (5). Otherwise, continue scanning

The voxel-wise SPRT procedures

The procedures of voxel-wise SPRT are recursively employed until the global stopping point as following steps 1 to 6:

Step 1: Collect one new fMRI image.

Step 2: Apply real time pre-processing procedures, such as normalized drift correction.

Step 3: Compute MLEs of $E(\boldsymbol{cB})$ and $Var(\boldsymbol{cB})$ based on equations (3) and (4).

Step 4: Compute SPRT statistic Λ_t, based on equation (2)

Step 5: Determine if stopping would be invoked for the voxel-level test based on rejection/acceptance rules, in (5).

Step 6: Apply global stopping criterion given the pre-determined target G %, to determine if fMRI–wide stopping should be invoked. If not, repeat from Step 1.

Final Step: If the specified global stopping rule is satisfied, at each voxel all fMRI signal data that has been collected up to the stopping is used to make a final determination as to activation status. The likelihood ratio as in (2) for one sided hypothesis will be computed, and rule for deciding between hypotheses is to select the associated hypothesized parameter value with the largest corresponding likelihood value.

SIMULATIONS and RESULTS

R package "nruRosim"[5] was used to generate a simulated fMRI image and the simulated dataset was analyzed within the Matlab environment (64-bit version R2012a The Mathworks, Natick, MA).

In simulations, the block paradigm is given as the order of following sequence:

R|A|R|B|R|A|R|B|R|…

where R represents a rest block and A and B respectively represent task A block and task B block. This alternating cycle is repeated for 60 times. Each voxel has its own assigned simulated fMRI signal. The simulated fMRI signal was generated by combining activation term and noise term which includes white noise, low frequency drift, physiological noise (heart beat and respiratory rate), temporal and spatial autocorrelation noise.

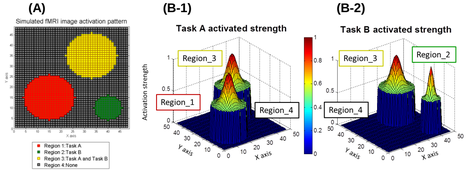

Figure 1— Simulation structure:
(A)The simulated image has a size of 48 × 48 voxels. Three circular shaped regions 1, 2 and 3 with exponential decay activation responses are separately activated by task A only (β_1=β, β_2=0), task B only (β_1=0, β_2=β) and both task A and task B (β_1=β, β_2=β). Two fMRI images were separately generated by defining β_1=β_2=1and β_1=β_2=3. The signal-to-noise ratio (SNR) equals 0.1 when the task activation strength equals 1 and the SNR equals 0.3 when the task activation strength equals 3. (B-1) and (B-2) separately show the task A and task B activation strength structure.

Efficiency of one sided voxel-wise SPRT on activation detection

For one voxel with t time series data, the linear model, including intercept term and two task related regression parameters, is described as follows:

$$\boldsymbol{Y}=\boldsymbol{GB}+\boldsymbol{E};\boldsymbol{E}\sim N(0,\sigma^2\boldsymbol{I})$$

$$\begin{bmatrix}y_1\\ \vdots\\ y_n\\ \vdots\\ y_t\end{bmatrix}=\begin{bmatrix}1 & g_1(1) & g_2(1)\\ \vdots & \vdots & \vdots\\ 1 & g_1(n) & g_2(n)\\ \vdots & \vdots & \vdots\\ 1 & g_1(t) & g_2(t)\end{bmatrix}\begin{bmatrix}b_0\\ b_1\\ b_2\end{bmatrix}+\begin{bmatrix}e_1\\ \vdots\\ e_n\\ \vdots\\ e_t\end{bmatrix}$$

where $\boldsymbol{Y}$ includes the observed fMRI signal intensities from time point 1 to t (1≤n≤t.) $g_1(.)$ and $g_2(.)$ separately represent the expected task A and task B BOLD signals which are generated by convoluting double-gamma HRF with corresponding experimental stimuli function. Type I error and Type II error are separately defined as 0.01 and 0.1. The stopping boundaries in the SRRT method are corrected by a Bonferroni approach. In GLM analysis, the multiple comparison issue is corrected by controlling the false discovery rate (FDR), q=0.01.

Simulation study result

The hypothesis of task A activation is H_0: $\boldsymbol{cB}$ = 0 against H_a: $\boldsymbol{cB}$ > 1 where c equals [0 1 0] and the hypothesis of task B is H_0: $\boldsymbol{cB}$ = 0 against H_a: $\boldsymbol{cB}$ > 1 where c equals [0 0 1].

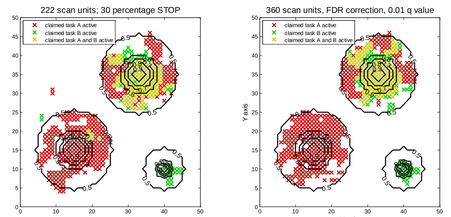

Figure 2 — Activity detection image results:
*Results with around 5% error percentage of claimed activation results by one sided voxel-wise SPRT and GLM. Two pictures are the images with claimed activation by one sided voxel-wise SPRT (left) and one sided voxel-wise GLM (right). Red **x** signs, green **x** and yellow **x** are separately claimed task A activation status, claimed task B activation status and claimed both tasks activation status. Activation strength contours are also labeled in these two images.*

A. Detection accuracies in one sided voxel-wise SPRT approach (30%global stopping, 212 scan units)

True activation strength	Region 1	Region 2	Region 3	Region 4
b=0.0				99.37% (=1428/1437)
1.0 > b ≥ 0.8	97.30% (= 36/37)	88.89% (= 8/9)	94.59 % (= 35/37)	

B. Detection accuracies in one sided voxel-wise GLM approach (360 scan units)

True activation strength	Region 1	Region 2	Region 3	Region 4
b=0.0				99.93% (=1436/1437)
1.0 > b ≥ 0.8	94.59 % (= 35/37)	100% (= 9/9)	100% (= 37/37)	

p.s. Detection accuracies = number of voxels is correctly classified to specific classification / number of voxels truly belong to specific classification

Real fMRI study and RESULTS

To demonstrate the potential impact of our proposed approaches with actual fMRI data, we consider a well-known and repeatedly described distinct region in the human ventral visual pathway, the fusiform face area (FFA), that is associated with showing faces, a distinct visual stimulus. The objective was to identify specific regions related with seeing adults face but not house image. Subjects performed 600 scan units experiment including 8 different stimuli: adult face, computer GUI, computer robot, computer humanoid, juvenile animals, houses, kid face and adult animals. These 8 stimuli are presented in the screen in random order. Among the stimuli, 24 adult face stimuli and 24 house image are presented. Recorded data were performed spatial smoothing in pre-processing steps and then 717 voxels located in FFA were selected to be analyzed by voxel-wise SPRT and GLM.

fMRI image result

The hypothesis of differential activation between adult face and house can be represented as a linear contrast of the associated task parameters, H_0: b_{adult_face}-b_{house} = 0 against H_a: b_{adult_face}-b_{house} > 1. 200 scan units per session over 3 sessions were used in practice (600 total), versus a reduction to 209 scan units, including 3 times adult face and 5 house stimuli, required through our proposed sequential methods.

Results from fixed GLM (600 time units) and voxel-wise SPRT (60 percent STOP, 209 time units)

Figure 3 — Adult face Activity detection image results:
The complete data results determined that 88 voxels were active for the differential activation, while our dynamic, SPRT-based approach found 123 such active voxels. There was an overlap of 68 voxels being found active with both approaches. Overall stopping of administration was determined when more than 60% of the voxel-level SPRTs called for stopping individually.

CONCLUSIONS

In simulations, voxel-wise SPRT is shown to be able to detect the voxels with relatively high activation levels with high accuracy, while reducing scan times substantially. In a real-data example, an objective is to identify regions that activate when shown a face stimulus but not when shown a house stimulus. This involved analysis of a contrast of regression parameters. Again, large saving in scan times of over 65% were observed using the sequential approach as compared with the length of the original design. These findings indicate that there are potentially great savings in scan time by individualizing fMRI experimental designs through dynamic adjustment of stimuli based on real-time fMRI. Practical implications include reduced costs for fMRI sessions while insuring acceptable levels of statistical accuracy in activation determinations. This work also serves as a basis for development of more complex and dynamic fMRI experimentation.

REFERENCES

1. *Wald A (1947) Sequential Analysis. New York: Wiley.*
2. *K. J. Friston APH, K. J. Worsley, J.-P. Poline, C. D. Frith, R. S. J. Frackowiak (1995) Statistical parametric maps in functional imaging: A general linear approach. Human Brain Mapping 2: 189-210.*
3. *Cox DR (1963) Large Sample Sequential Tests for Composite Hypotheses. Sankhyā: The Indian Journal of Statistics, Series A: 5-12.*
4. *Shyamal K. De MB (2012) Step-up and step-down methods for testing multiple hypotheses in sequential experiments. Journal of Statistical Planning and Inference 142: 2059–2070.*
5. *Marijke Welvaert JD, Beatrijs Moerkerke, Geert Verdoolaege, Yves Rosseel (2011) neuRosim: An R Package for Generating fMRI Data. Journal of Statistical Software 44: 1-18.*