NeuroImage 52 (2010) 1390–1400

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/ynimg

A least angle regression method for fMRI activation detection in phase-encoded experimental designs

Xingfeng Li [a,b,*], Damien Coyle [a], Liam Maguire [a], Thomas M McGinnity [a], David R Watson [a], Habib Benali [b]

[a] Intelligent Systems Research Centre, University of Ulster, Magee Campus, Derry, BT487JL, Northern Ireland, UK
[b] Inserm, UPMC University Paris 06, UMR_S 678, Laboratoire d'Imagerie Fonctionnelle, GHU Pitié-Salpêtrière, 91 bd de l'Hôpital, F-75634 Paris Cedex 13, France

ARTICLE INFO

Article history:
Received 12 November 2009
Revised 4 May 2010
Accepted 6 May 2010
Available online 25 May 2010

Keywords:
Least angle regression
Fast orthogonal search
Brain activation detection
fMRI time series
Model selection

ABSTRACT

This paper presents a new regression method for functional magnetic resonance imaging (fMRI) activation detection. Unlike general linear models (GLM), this method is based on selecting models for activation detection adaptively which overcomes the limitation of requiring a predefined design matrix in GLM. This limitation is because GLM designs assume that the response of the neuron populations will be the same for the same stimuli, which is often not the case. In this work, the fMRI hemodynamic response model is selected from a series of models constructed online by the least angle regression (LARS) method. The slow drift terms in the design matrix for the activation detection are determined adaptively according to the fMRI response in order to achieve the best fit for each fMRI response. The LARS method is then applied along with the Moore–Penrose pseudoinverse (PINV) and fast orthogonal search (FOS) algorithm for implementation of the selected model to include the drift effects in the design matrix. Comparisons with GLM were made using 11 normal subjects to test method superiority. This paper found that GLM with fixed design matrix was inferior compared to the described LARS method for fMRI activation detection in a phased-encoded experimental design. In addition, the proposed method has the advantage of increasing the degrees of freedom in the regression analysis. We conclude that the method described provides a new and novel approach to the detection of fMRI activation which is better than GLM based analyses.

Introduction

The general linear models (GLM) method (Nelder and Wedderburn, 1972; Seber and Lee, 2003) has been widely applied for activation detection in functional magnetic resonance imaging (fMRI) data analysis (Friston et al., 1995; Worsley et al., 2002). The basic idea of this method is to construct a design matrix (model) beforehand and then, using a least squares method, estimate the regression parameters of the model. After obtaining the model parameters, a contrast matrix for an effect of interest is often defined to compare different effects. For example, a difference between stimuli is often employed to detect the brain activation in the first level (single run) of the fMRI time series analysis. Then the null hypothesis that the effect difference is zero is tested, and the T statistic for the hypothesis test is the ratio between the effect difference and the estimated standard deviation (Worsley et al., 2002).

When using GLM for fMRI activation detection, the first step is to build a model (design matrix) of how the neuron population responds to an external stimulus. The advantage of defining the model beforehand is that it greatly simplifies the analysis and interpretation of fMRI data if the model or the design matrix is selected correctly. It also simplifies the calculations if only a single fixed model for all neuron population responses in the brain is required. However, these targets are currently not feasible because the exact neuron response in different brain regions is unknown. More importantly, it is not reasonable to assume that all the neurons in the brain respond to the same stimulus in exactly the same way when it is presented on different occasions (Glover, 1999). Also, inaccuracies will occur if one uses the same model (design matrix) for all the neuron populations in the brain. Therefore, it is necessary to build different models for different neuron populations in terms of their fMRI responses. This hypothesis is based on the fact that different neuron populations in the cortex have different responses (shape, magnitude, hemodynamic delay, and slow drift etc.) even for the same stimulus. This is particularly true for retinotopic mapping with phase-encoded design where hemodynamic delay and shape are important factors (Sereno et al., 1995; Engel et al., 1997; Smith et al., 2001; Warnking et al., 2002). Secondly, the order/size of the slow drift for the neuron population is also not known beforehand, particularly in fMRI activation detection. A low order drift model often leads to a bigger T value indicating stronger activation detection, but it be may insufficient to model the fMRI response when more complex dynamics are present. On the other hand, a higher ordered drift model can fit the response well, but has a lower T value with

* Corresponding author. Intelligent Systems Research Centre, University of Ulster, Magee Campus, Derry, BT487JL, Northern Ireland, UK. Fax: +44 28 71 37 52 54.
E-mail address: x.li@ulster.ac.uk (X. Li).

1053-8119/$ – see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.neuroimage.2010.05.017

decreasing of degrees of freedom (df) for the statistical test. To overcome such issues it is necessary to develop a new method to estimate the optimal model order for each voxel in the whole brain. Lastly, the models in the GLM method of fMRI activation detection are usually not parsimonious, and can be computationally expensive to implement for online activation detection. To overcome the limitations of GLM in first level fMRI data analysis, a new method which relies on least angle regression (LARS) (Efron et al., 2004) is proposed.

This paper is organized as follows. The paper begins with the GLM method for fMRI activation detection. Subsequently, details of how the LARS method is derived from the GLM method for the fMRI activation detection are presented. The LARS method for fMRI activation detection is then introduced and two algorithms (Moore–Penrose pseudoinverse (PINV) and fast orthogonal search (FOS)) for implementation of the method are detailed. Finally, the LARS method is applied to study 11 normal subjects with a phase-encoded experimental design. A comparison between the LARS and conventional GLM method is provided and the paper concludes with a discussion of the performance and future application of the proposed method.

Materials and methods

GLM for phase-encoded design

The well-known GLM method for activation detection in the first level (single run) of fMRI data analysis (Worsley et al., 2002) is:

$$Y_t = \underbrace{u_1\beta_0 + u_2\beta_1 + \cdots + u_k\beta_{k-1}}_{\text{input}} + \underbrace{x_{t,1}\beta_k + x_{t,2}\beta_{k+1} + \cdots + x_{t,p}\beta_{p-k+1}}_{\text{drift}} + e_t \quad (1)$$

where Y_t is the fMRI response at time index t;β_p is the coefficient; u_k is the experimental design (or brain system input);$x_{t,p}$ is the constant term to model the slow drift; k is the number of inputs; $p-k+1$ is the number of drift terms in the model; e_t is the error term. For the single input phase-encoded design, we have $k=1$, i.e.:

$$Y_t = \underbrace{u_1\beta_{t,0}}_{\text{input}} + \underbrace{x_{t,1}\beta_1 + x_{t,2}\beta_2 + \cdots + x_{t,p}\beta_p}_{\text{drift}} + e_t \quad (2)$$

Eq. (2) can be written in matrix form as:

$$Y_t = X_t'\beta + e_t \quad (3)$$

where $X_t' = [u_1, x_{t,1}, x_{t,2}, \cdots, x_{t,p}]$, and $\beta = [\beta_0, \beta_1, \cdots, \beta_p]'$. In the GLM for the fMRI activation detection, the solution of Eq. (3) is then:

$$\hat{\beta} = \underset{\beta}{\operatorname{argmin}} \left\| Y_t - \sum_{i=1}^{p} x_{t,1}\beta_i - u_1\beta_0 \right\| \quad (4)$$

where p is the size of the design matrix for slow drift. If the design matrix X_t' and input u_1 are fixed for all neuron populations, this is the conventional GLM method for activation detection. Eq. (3) can be solved by the PINV method i.e., $\hat{\beta} = X_t^+ Y_t$, where X_t^+ is the Moore–Penrose pseudoinverse of matrix X_t'. For large fMRI datasets, the FOS algorithm (see appendix) can be employed to solve the equation. The $x_{t,1}$ constant (zero-th order $x_{t,1}$ component) and higher order polynomial ($x_{t,2}, \cdots, x_{t,p}$) components providing the slow drift term (Bandettini et al., 1993) contributions to response Y_t can also be described by the cosine transform basis functions (Friston et al., 1995) or splines (Smith et al., 1999; Worsley et al., 2002; Tanabe et al., 2002). In this study, we used 0–14th order polynomials to model the slow drift, i.e. , $x_{t,1}=1, x_{t,2}=x, \cdots$, and $x_{t,15}=x^{14}$.

LARS for activation detection

To model the neuron population according to the fMRI response, consider the following optimization problem:

$$\hat{\beta} = \underset{\beta}{\operatorname{argmin}} \left\| Y_t - \sum_{i=1}^{p} X_{t,i}\beta_i - u_1\beta_0 \right\| \text{ subject to } \left(\sum_{i=1}^{p} |\beta_i| + |\beta_0| \right) \le s \quad (5)$$

where $s \ge 0$ is the turning parameter. A large s will result in the ordinary least squares estimation. However, smaller values of s produce shrunken estimates of $\hat{\beta}$, often with many components equal to zero. Eq. (5) can also include other hemodynamic models (inputs) for selection (for the phase-encoded design, we use only one input u_1), so that choosing s becomes a model selection procedure for modelling the hemodynamic response and drift terms. There are several algorithms such as LASSO that can do this (Donoho and Elad, 2003; Huo and Ni, 2007; Hesterberg et al., 2008).

Modelling the hemodynamic response

For the input function u_1, two gamma functions (Glover, 1999), block convolution with Gaussian function (Smith et al., 2001) or a sinusoidal function (Lange and Zeger, 1997) can be employed to model the system input. In the phase-encoded design and standard block design (periodic blocked paradigms), u_1 can be determined by fast Fourier transformation (FFT) of the fMRI time series and the fundamental frequency of the response is used to model the input $u_1$for each voxel to estimate the fMRI response (system input) adaptively. The fundamental frequency of the fMRI response can also be used to estimate the hemodynamic delay (response phase). After the delay has been calculated, the model can be constructed online, and the model is adaptive to the fMRI delay. We include three widely used hemodynamic models for brain system input selection:

1. Because the stimuli are changing periodically, the sinusoidal function can be regarded as a brain system input; i.e. the periodical square block wave (experimental design) can be approximated by the fundamental frequency of the wave. The FFT analysis method is a powerful way to model the hemodynamic response for the periodical change stimuli, the hemodynamic function can be written as (Li et al., 2007a):

$$f_{t,1} = a\cos(\omega t + \theta) \quad (6)$$

where θ is the delay/onset or phase of the response, and can be estimated using the FFT method; ω is the angular frequency, $\omega = 2\pi f$, where f is the frequency of the stimulus/input; a is the magnitude.

2. After the delay/onset is estimated by the FFT analysis, a two-gamma function is built according to the following equation (Glover, 1999):

$$f_{t,2} = \left(\frac{t}{d_1}\right)^{a1} exp\left(-\frac{t-d_1}{b_1}\right) - c\left(\frac{t}{d_2}\right)^{a2} exp\left(-\frac{t-d_2}{b_2}\right) \quad (7)$$

where $a_1=6; b_1=0.9; d_1=a_1\times b_1$; $c=0.35$; $b_2=0.9; a_2=12$; $d_2=a_2\times b_2$, are the typical parameters.

3. We also include the block function convolve with Gaussian function (Smith et al., 2001) to model the shape (i.e. duty cycle changes in the response); the model base function is:

$$f_{t,3} = (block) \otimes \left(\exp\left(-\left(\frac{t}{\sqrt{2}c}\right)^2\right)\right) \quad (8)$$

where block is the block function (for example Figs. 1A and 2A). The full width at half maximum (FWHM) is determined according to: $FWHM = 2\sqrt{2\ln(2)}c$.