Model Estimation and Selection for Representing Group fMRI Activations

Karin Engel
karin.engel@ovgu.de
Klaus D. Toennies

Department of Simulation and Graphics
Otto von Guericke University
Magdeburg, Germany

Abstract

Identifying functional brain regions from fMRI data involves the comparison of individual activations and the inference of group activation model. To overcome the shortcomings of voxel-based analyses, which model the data as a smooth random field, we use a method that directly compares the individual activation patterns. In our work, an optimal, generative model of the activation foci of interest is computed by employing factor analysis and model selection techniques. We show the advantages of our approach to functional localisation on synthetic data and data from an auditory fMRI experiment.

1 Introduction

Identifying brain regions of interest across subjects using functional magnetic resonance imaging (fMRI) is difficult due to the considerable degree of inter-subject variability in shape, location and configuration of these regions [4]. The correspondence problem can be solved by constraining group analysis by functional labels, or macro-anatomical landmarks [7]. The standard procedure consists in registering the different brains, performing a voxel-wise, multivariate analysis (*e.g.* random effects analysis, RFX) and comparing functionally specific effects in the group activation map to atlases containing architectonic information from post mortem brains [5, 9, 13]. Beyond that, structural matching techniques have been recently employed for comparing the functional activation patterns across subjects [6, 8, 11, 12]. We use the surface-based, structural analysis method ISA by Engel, *et al.* [6], which allows investigating thoroughly the large number of small but separate activation foci in the auditory cortex (AC) in relation to regional, macro-anatomical landmarks [1, 13]. Our work focusses on the model estimation and selection problem in order to extend their approach: Based on the activation mapping, a group activation model is obtained by applying standard manifold learning methods. A model that generalises to a wider population (*i.e.* unseen data) without over-fitting is selected by comparing the performance of different models. The quality of the functional localisation results can then be evaluated statistically.

2 Structural Multi-subject Analysis

Prior to the group analysis, each individual fMRI data set is transformed to a surface-based, sparse description $Y(s) = (\mathbf{y}_1, \ldots, \mathbf{y}_{N(s)}), s = 1, \ldots, K$, in terms of the spatial coordinates of

local maxima $\mathbf{y}_j$ in the K activation maps [6]. Therefore, the functional volumes are pre-processed, projected onto the individual cortical surfaces, and analysed in the general linear model framework. The individual activation patterns $Y(s)$ are assumed to be instantiated from a group model $X = (\mathbf{x}_1, \ldots, \mathbf{x}_N)$, which represents the activation foci of specific regions of interest (ROI). Each pattern may be subject to random and structural error (*e.g.*, due to measurement and detection error, or inter-individual differences). In order to separate the ROI from noise, our method uses a parametric model $\mathscr{G}(\mathbf{p}, X)$ of the group activation pattern, where the parameters $\mathbf{p}$ define constraints on the functional variability across subjects.

Operto, *et al.* [8] use a Markov process for inferring the group model by comparing the relative positions of activations in a global reference space. The BFL detection by Thirion, *et al.* [11, 12] relies on the leave-one-out validation of individual activations being observed from average (RFX) activation maps X. In their approaches, $\mathbf{p}$ would comprise global pose, spatial relaxation and smoothing parameters (to account for misregistration), as well as a reproducibility criterion. In contrast, the ISA method of Engel, *et al.* [6] simultaneously estimates the group activation model and recovers correspondences between the activation foci of specific functional fields by matching in an embedded, *i.e.* intrinsic, pattern space. Here, X is a reference pattern, and $\mathbf{p}$ contains local pose and variation parameters.

2.1 Activation Mapping

Let the mapping of activation patterns $Y \in \{Y(s), s = 1, \ldots, K\}$ be represented by a function

$$\ell(L^*) \colon X \mapsto Y, \quad \text{where} \quad L^* = \arg\max_L \mathscr{C}(L, P, \tau), \tag{1}$$

such that $\ell(\mathbf{x}_i)$ is the observed activation focus $\mathbf{y}_j \in Y$ that best corresponds to model point $\mathbf{x}_i \in X$ (and vice versa). L^* defines the optimal pairwise assignments w.r.t. the functional $\mathscr{C}(L, P, \tau) = \sum_i \sum_j P(i, j) L(i, j)$, where P is a correspondence probability matrix with elements from $[0, 1]$, and

$$L(i, j) = \begin{cases} 1, & \text{if } P(i, j) = \max_i P(i, j) = \max_j P(i, j) \wedge P(i, j) > \tau \in [0, 1) \\ 0, & \text{otherwise.} \end{cases} \tag{2}$$

Following [6] for determining true correspondences (1), an iterative scheme estimates at each discrete time step $t > 0$ the correspondence probability P, which depends on the feature affinity in embedded space, and computes a non-rigid, geometry-preserving transformation to align X and Y w.r.t. the matching pairs i, j. This transformation is described as a smooth, time-varying displacement field $u(\mathbf{x}, t), \mathbf{x} \in X$, which is expressed in terms of a weight vector $\mathbf{q}$ as $u(\mathbf{x}, t) = \Phi \mathbf{q}(t)$, *i.e.* $X(t) = X + u(\mathbf{x}, t)$. The orthonormal vectors in Φ span the pattern space according to the chosen Gaussian kernel embedding of the activation foci.

2.2 Inferring a Graphical Model of the Functional ROI

Finally, the group activation pattern is defined as

$$\bar{X} = \{\mathbf{x}_i \colon p(\mathbf{x}_i | Y) \geq \vartheta\}, \tag{3}$$

where $\vartheta \in [0, 1]$ is a threshold on the reproducibility of activations in the group. Based on the correspondences $L^*_{s,\varsigma}$ with a reference pattern $X = Y(\varsigma), \varsigma = 1, \ldots, K$, we can directly study the properties of the point distribution that results from the Gaussian kernel embedding [6].

More specifically, standard methods from statistics, *i.e.* (kernel) PCA [10, 14], can be applied for robustly building a statistically representative model of the group activation,

$$X^* = \bar{X}^* + \Psi \mathbf{b}. \tag{4}$$

Here, $\bar{X}^*$ denotes the observation mean, and the functions $\psi_k \in \Psi$ span the generative pattern space according to the $d \times d$-empirical covariance matrix $C = \Psi \Delta^2 \Psi^\top$ of the centred random variables in embedded space. The weight vector $\mathbf{b}$ comprises the latent variables, which follow a $\mathcal{N}(0, \mathbf{I})$ distribution with d degrees of freedom (DOF). The posterior $p(\mathbf{b}|Y)$ may then be used instead of the energy functional $\mathscr{C}$ for assessing the matching confidence.

2.3 Model Selection

The observable variables Y are aggregated in a model (4) representing the underlying structural organisation of the data. Latent variables, as inferred by factor analysis (Sect. 2.2), represent shared variance, *i.e.* variations in the spatial coordinates of ROI, expanded along the (ordered) principal components ψ_k. Each observation Y deviates from the maximum a-posteriori reconstruction $\bar{X}^* + \bar{\Psi}\mathbf{b}^*$ by the residual ρ, for which $p(\rho) = \mathcal{N}(0, \sigma \mathbf{I})$. In our case, the reconstruction error depends on the complexity m of the generative model, as well as on the reliability of the underlying correspondence sets (*i.e.* quality of the "training data").

By model selection one wants to find the $m < d$-dimensional basis expansion that minimises the empirical risk $R(\sigma, \mathbf{b}, m)$ of the regression function. R is a function of the measurement error, of the estimation error, *i.e.* distance between the model parameters in the full (Eq. 4) and truncated model space $\bar{\Psi} \in \mathbb{R}^{d \times m}$, and of the approximation error $r(m) = \sum_{l=m+1}^{d} \mathbf{b}(l)^2$ (cf. [2]). The model with the smallest number of DOF m is selected, such that no more complex model gives a significantly lower risk. We use the method of Cootes *et al.* [3] for comparing the empirical distribution $\hat{p}(\rho)$ with a theoretical distribution $p(\rho)$ using error propagation and the Bhattacharya metric, $\mathscr{B}$, for hypothesis testing.

Unmatched features with $\sum_i L(i,j) = 0, \forall i$, are assigned a null label $\ell(\mathbf{y}_j) = \emptyset$, *i.e.* considered "noise". As a result, each subject may or may not show a region associated with a focus of activation defined at the group level, and the particular measurement may or may not be included in the correspondence sets used for learning. Our solution to this "chicken and egg" problem is to select as reference X a representative pattern $Y(\varsigma)$ from the pairwise correspondences $L^*_{s,n}, s \neq n$. A naïve choice is the sample with the largest number of features, *i.e.* $\varsigma_1 = \arg\max_s |Y(s)|$. Since selecting an individual observation involves the risk of introducing a bias in the results, we propose the following alternative strategies for model selection. The second method uses cross-validation, and selects

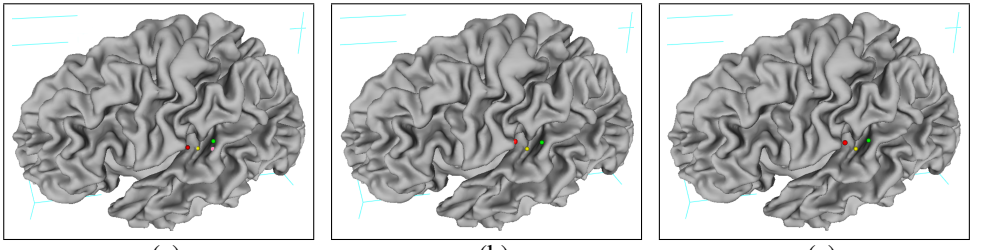
$$\varsigma_2 = \arg\max_s \ \mathscr{C}(L^*_{s,n}, P, \tau), \ n \in \{1, \ldots, K\} \setminus s, \tag{5}$$

based on the inter-pattern similarity function $\mathscr{C}$ (cf. Sect. 2.1). The third method uses

$$\varsigma_3 = \arg\max_s \ \delta(L^*_{s,n}, P, \tau), \ n \in \{1, \ldots, K\} \setminus s, \tag{6}$$

where δ averages the (robust) Mahalanobis distance between the embedded features, which are observed in a fraction of ϑ subjects (cf. Sect. 2.2). Fourthly, we can select from the sample the pattern $Y(\varsigma)$ as reference that gives rise to the graphical model, of which the distrbution of residuals $\hat{p}(\rho_\varsigma)$ best matches the distribution $p(\rho)$ of the observation noise, *i.e.*

$$\varsigma_4 = \arg\min_s \ \mathscr{B}(\hat{p}(\rho_\varsigma), p(\rho)). \tag{7}$$



(a) (b) (c)

Figure 1: Predicted location of auditory ROI within the temporal region of a reference cortex. The presented group activation foci were identified in at least 7 of 9 subjects using the proposed multi-subject analysis (a) and related techniques (see Section 3).

3 Experimental Results

For a quantitative analysis, we synthesised a ground truth pattern X of $N = 10$ activation foci with a minimum inter-focus spatial distance of 5 mm on a reference cortical surface from our database. $K = 100$ disturbed instances Y of this pattern were generated by duplicating the original pattern, and introducing random position error $e(X) \sim \mathcal{N}(0, \sigma), \sigma = 2$mm, and structural error $e(N) \sim \mathcal{N}(0, \varepsilon), \varepsilon = 0.1$. In our experiments, the data set was randomly split into a training and test set, and then each of the methods (Sect. 2.3) was run on this split. We computed the error rates using $\vartheta = 0.5$ in a repeated random sub-sampling validation. The model X that was selected using minimisation of R using Equations 6 and 7, performed best and provided a good reconstruction of the ground truth. Except from the largest sample-based model (*i.e.* using ς_1), we obtained superior results over ISA [6]. The difference in the performance was statistically significant ($p < 0.01$, one-sided t-test).

We further compared our results on real data from an auditory fMRI study with the ROI identified using RFX analysis, ISA [6], and a clustering method (referred to as CVC) that employs principles of [11]. For the sake of fairness, all analyses were constrained to the local, surface-based reference spaces of the auditory territories described in [1]. In brief, for $\vartheta = \frac{5}{9}$, the proposed method extracted 9 group activation foci compared with 7 regions from ISA and CVC. Four regions were identified in the RFX group map, in which small but separate regions were fused into larger clusters or "averaged out". In comparison with our method, both ISA and CVC computed suboptimal assignments, most probably due to the inferior reliability of the underlying models. The reference pattern was in our case chosen according to Equation 6. As shown in Figure 1(a), the four foci with highest reproducibility in at least 7 of 9 subjects identified one activation in the primary AC (red label), one ROI in the secondary AC (yellow) and two regions (green and pink) on planum temporale, *i.e.* association cortex. Each of the auditory territories can be further subdivided into at least three areas and thus the number of regions found corresponds well with this expectation. Parts of these regions were also detected by ISA and CVC (Figs. 1b and c).

4 Conclusions and Future Work

This paper presents a novel method for inferring a generative model of the positions of ROI from fMRI data of multiple subjects. Effectively, the method relaxes the common, over-simplifying assumption of activated regions being clustered in a common spatial reference.

Correspondences are found across subjects by a topology-preserving registration of the activation patterns in an embedded space [6], and then used for learning a generative group activation model. In combination with the evaluated strategies for model selection, our approach improves previous work on the structural analysis of group functional data [6, 8, 11], and allows a statistical assessment of the individual observations and predictive performance of the group activation model. The identification of outlying observations is a difficult problem in the given high dimension (*i.e*. whole brain)–small sample–"several sources of valid variation" scenario. Therefore, an important direction for future research is to improve the inter- and *intra*-subject modelling, such that to each region a probability can be assigned of being a ROI given its relative position and specific signal characteristics.

References

[1] A. Brechmann, F. Baumgart, and H. Scheich. Sound-level dependent representation of frequency modulations in human auditory cortex. J Neurophys, 87:423-433, 2002.

[2] O. Chapelle, V. Vapnik, and Y. Bengio. Model selection for small sample regression. Machine Learning, 48(1):9-23, 2002.

[3] T. F. Cootes, N. A. Thacker, and C. J. Taylor. Automatic Model Selection by Modelling the Distribution of Residuals. In *Proc. ECCV (4)*, pages 621-635, 2002.

[4] J. T. Devlin and R. A. Poldrack. In praise of tedious anatomy. Neuroimage, 37(4):1033-1058, 2007.

[5] S. Eickhoff, *et al*. Testing anatomically specified hypotheses in functional imaging using cytoarchitectonic maps. Neuroimage, 32(2):570-582, 2006.

[6] K. Engel, K. Toennies, and A. Brechmann. Surface-based anatomo-functional parcellation of the auditory cortex. In *Proc. IEEE ISPA*, pages 602-609, 2009.

[7] B. Fischl, *et al*. Cortical folding patterns and predicting cytoarchitecture. Cerebral Cortex, 18(8):1973-1980, 2007.

[8] G. Operto, *et al*. Surface-based structural group analysis of fMRI data. In *Proc. MICCAI (1)*, pages 959-966, 2008.

[9] P. Rasser, *et al*. A deformable brodmann area atlas. In *Proc. IEEE ISBI*, pages 400-403, 2004.

[10] S. T. Roweis. EM algorithms for PCA and SPCA. In *Proc. NIPS*, pages 626-632, 1998.

[11] B. Thirion, P. Pinel, and J. B. Poline. Finding landmarks in the functional brain: Detection and use for group characterization. In *Proc. MICCAI (2)*, pages 476-483, 2005.

[12] A. Tucholka, *et al*. Triangulating cortical functional networks with anatomical landmarks. In *Proc. IEEE ISBI*, pages 612-615, 2008.

[13] D. Viceic, *et al*. Local landmark-based registration for fMRI group studies of nonprimary auditory cortex. Neuroimage, 44(1):145-153, 2009.

[14] C. K. I. Williams. On a connection between kernel PCA and metric multidimensional scaling. Machine Learning, 46(1-3):11-19, 2002.