

AN INFORMATION-BASED CLUSTERING APPROACH FOR FMRI ACTIVATION DETECTION

Lijun Bai^a, Wei Qin^a, Jimin Liang^a, Jie Tian^{b*}, IEEE Senior Member

^aLife Science Research Center, XiDian University

^{b*}Medical Image Processing Group, Institute of Automation, Chinese Academy of Sciences

ABSTRACT

Most clustering algorithms in fMRI analysis implicitly require some nontrivial assumption on data structure. Due to arbitrary distribution of fMRI time series in the temporal domain, such analysis may mislead and limit the detector's performance. In this work, the authors exploited the application of an information-based clustering algorithm (Iclust) which could avoid these assumptions and provide many other benefits, such as no cluster shape restriction, no need of a prior definition about similarity measure, and the ability of capturing both linear and nonlinear dependence. Results from both artificial and real fMRI data indicated that the proposed framework could achieve better spatio-temporal accuracy, and enabled the exploration of fine functional distinction of the human visual system in accordance with its well-known anatomy organizations.

Index Terms— Pattern clustering methods, Magnetic resonance imaging

1. INTRODUCTION

Functional magnetic resonance imaging with high temporal and spatial resolution represents a powerful technique for visualizing rapid and fine activation patterns of the human brain. However, the low signal-to-noise ratio (SNR) and confounding sources of artifacts in the fMRI time sequences, make the detection of brain activation a challenging task. Recently, clustering methods are gaining increasing attentions as they aim at separating time series into several patterns according to their similarities. Well known members of this category are Fuzzy C-means (FCM) [1], hierarchical clustering [2], Kohonen clustering neural network [3], etc.

Indeed, several clustering methods are based upon the controversial postulate that data series conform to certain shapes (structure assumption problem). For example, single-linkage hierarchical clustering tends to produce “long chained” clusters [2]. The FCM groups data into scatter with a ‘hyper-spherical’ shape if adopting the Euclidean distance as similarity measure [1]. These methods only provide high-quality results when the data scatter in the

assumed shapes; while in a general case of more complex clusters with arbitrary shapes, their performance degenerates markedly. In addition, the central question in clustering is whether an essentially unsupervised analysis of datasets can recover categories that have ‘meaning’ [4]. Therefore, the question of how to construct an accurate model obviously is quite involved, raising further issues that are often addressed arbitrarily before the cluster analysis begins [4].

As mentioned above, the literature on clustering in fMRI data analysis is vast. Therefore, the primary aim in this paper was not limited to suggest yet another clustering algorithm, but rather to focus on questions about the formulation of clustering problems in fMRI data analysis which may provide plausible methods to these concerns (such as cluster validity and parameterization selection). To address this question, a novel information-based clustering method originally proposed by Slonim et al. [4] was introduced. In this framework, the formulation of clustering solution falls on the search of a tradeoff between maximizing the mean similarity of elements within a cluster and minimizing the complexity of the description provided by cluster membership. In a systematic manner, the results obtained by this new approach were compared and evaluated with classical general linear model analysis (GLM) and model-free schemes such as FCM and independent component analysis (ICA).

2. MATERIALS AND METHODS

2.1. Stimulated Dataset Generation

To simulate the real-world fMRI data and to get concrete comparison of performance, the baseline image (29th slice) was selected from the real auditory fMRI dataset from Wellcome Cognitive Neurology at University College London, and simulated activation time-series were added to the baseline voxels in three artificial activation subregions (spatial pattern) shown in Fig. 1(a). The simulated activation responses were obtained by convolving a stimulation pattern with the HRF as shown in Fig. 1(b). Then an additive white Gaussian noise was added to all voxels with intensities proportional to the baseline voxel.

SNR was defined as the ratio of mean signal over the noise standard deviation.

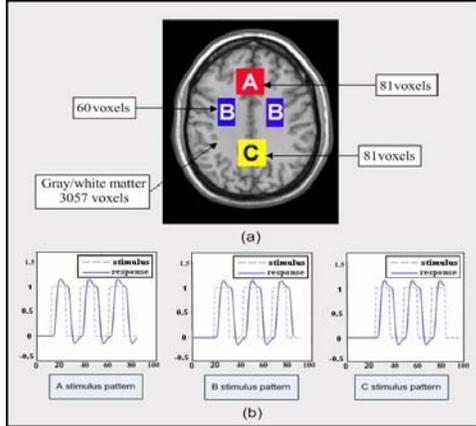


Fig. 1 Spatial pattern of activation in the simulated data. (a) fMRI image with three ROIs used as the basis for generating assumed active voxels; (b) Artificially temporal response and stimulation reference of three subregions.

2.2. In Vivo fMRI Dataset

In vivo experiments were carried out with a 3.0 Tesla Signa (GE) MR whole body Scanner. Nine subjects took part in this study after they provided informed consent approved by a local human study review board. A checkerboard pattern flashing for one second at 8Hz was used as the visual stimulus. As baseline condition, a central red dot in front of the same isoluminant background was presented for 30s between stimuli. During the task, participants were instructed to focus their gaze on the center of the screen and to minimize blinking whilst the checkerboard was presented. Twelve stimuli were presented per run, and took approximately 6 min.

Functional images of 32 slices (spanning the cerebral cortex) were collected in a sagittal orientation parallel to the AC-PC plane with 4 mm slice thickness (no gaps) using a single-shot gradient-recalled echo planar imaging (EPI) sequence (TE = 27 ms, TR = 2000ms, flip angle = 85°, matrix size = 64 x 64, FOV: 256mm x 256mm, giving an in-plane resolution: 4 x 4 mm).

2.3. Procedure of Data Analysis

Data preprocessing were performed with the Statistical Parametric Mapping SPM2 software package (Wellcome Department of Imaging Neuroscience, London UK). Prior to the time series analysis, the data were processed to remove low-frequency signal changes and motion-related artifacts. Data preprocessing was spatially realigned, un-warped, slice-timed, normalized to the MNI space, and smoothed with an isotropic 6-mm FWHM Gaussian kernel, with resulting voxels size of $3mm \times 3mm \times 3mm$. Because the only interest laid on the signal variation of functional MR images

rather than their baseline signal intensities, the mean value of each voxel signal was subtracted by its mean value over the entire image sequence to eliminate the dependency of baseline intensity level.

3. MATHEMATICAL FRAMEWORK

3.1. Mutual information as similarity measure

Mutual information is a well-established concept to measure relatedness between a pair of variables independent of assumptions about the form of the underlying probability distributions. In this work, we are trying to measure the relation between activation patterns of voxels across a variety of time points $t = 1, 2, \dots, N$. For voxel i and j , we can define the joint probability density of their pattern levels

$$P_{i,j}(y_i, y_j) = 1/N \sum_{t=1}^N \delta(y_i - y_i(t)) \delta(y_j - y_j(t)) \quad (1)$$

becoming smooth as $N \rightarrow \infty$. Given the above joint distributions, their relations measured by mutual information come from the following

$$s(i, j) = I_{i,j} = \int dy_i \int dy_j P_{i,j}(y_i, y_j) \log_2 [P_{i,j}(y_i, y_j) / P_i(y_i) P_j(y_j)] \text{bits} \quad (2)$$

Although (2) become smooth as $N \rightarrow \infty$, one can still obtain reliable estimates by applying the ‘direct’ estimation method originally developed in the neural coding analysis [5].

3.2. Optimal clustering solution

Clust attempts to formulate the clustering problem as a tradeoff between maximizing the mean similarity of elements within a cluster and minimizing the complexity of the description provided by cluster membership. Given the similarity measure $s(i, j)$ expressed by mutual information, optimal clustering is a probabilistic assignment to clusters C according to $P(C/i)$ to maximize the object function

$$F = \langle s \rangle - TI(C; i) \quad (3)$$

Where $\langle s \rangle$ defines the mean similarity of elements chosen at random out of each cluster,

$$\langle s \rangle = \sum_C P(C) \sum_{i,j} P(i/C) P(j/C) s(i, j) \quad (4)$$

and $I(C; i)$ embodies the identity information of elements in certain cluster,

$$I(C; i) = \sum_C \sum_i P(C/i) P(i) \log [P(C/i) / P(C)] \quad (5)$$

the Lagrange multiplier T enforces the constraint on $I(C; i)$.

In general, it is not possible to find an explicit solution for $P(C/i)$ that maximizes F . However, if we assume that F is differentiable with respect to the variable $P(C/i)$, self-consistent equations that any optimal solution must obey:

$$P(C/i) = P(C) / Z(i; T) \exp \{1/T [2s(C; i) - s(C)]\} \quad (6)$$

Where $Z(i;T)$ is a normalization constant and $s(C;i)$ is the expected similarity between i and a member of cluster C ,

$$s(C;i) = \sum_{i_1=1}^N P(i_1/C) s(i_1,i) \quad (7)$$

and $s(C)$ is the average similarity among pairs chosen independently out of the cluster C indicated in (4).

Although (6) is an implicit set of equations, we can turn this self-consistency condition into an iterative algorithm to search an explicit numerical solution for $P(C/i)$, yielding a (perhaps local) maximum of F . If the solution of algorithm at the m_{th} iteration can be denoted as $P^{(m)}(C/i)$, at the $m+1_{st}$ iteration, the update step should be followed by:

$$P^{(m+1)}(C/i) \leftarrow P^{(m)}(C) \exp\{1/T[2s^{(m)}(C;i) - s^{(m)}(C)]\} \quad (8)$$

In principle we repeat this procedure for different initializations and choose the preferred solution that maximizes $F = \langle s \rangle - TI(C;i)$. In some extreme cases, the above algorithm might produce a non-monotonic behavior in F . However, for the regime $T \geq \max_{i_1, i_2} 2s(i_1, i_2)$, it is possible to prove this convergence analytically [6].

4. RESULTS

4.1. On Simulated Dataset

In order to alleviate the computation burden in clustering, the researchers firstly reduced the number of voxels using t-test between response and stimulus. In this work, we targeted to find four clusters: three clusters covering three types of activated voxels respectively, and the fourth one corresponding to non-activated voxels. For Iclust, intensive analysis emphasized the parameters pair $\{N_c = 4, T = 1/25\}$ for which the saturation of $\langle s \rangle$ was relatively clear. For GLM, the significance threshold was set to $P < 0.0001$. For FCM, the number of clusters was set to four, and ICA was constrained to generate three independent components. Fig. 2(a)-(d) show the detection results from Iclust, GLM, ICA, and FCM, when SNR was set to 1.1. We could see that activated components detected by proposed Iclust were intactly identified and sufficiently separated from strong noise with accurate spatial distribution, while other competing methods detected the activation regions with some misidentifications and contaminating noise.

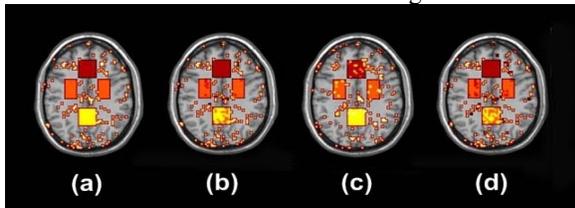


Fig. 2 Detection results of phantom data from (a) Iclust (b) GLM (c) ICA (d) FCM.

To quantitatively assess the performance of these methods and distinguish between effect-region and pure-noise voxels, the authors estimated the sensitivity and specificity when SNR increased from 0.4 to 2.0 (Fig. 3). According to Fig. 3, Iclust achieved both highest sensitivity and specificity at different SNR levels with a significant margin, compared to other alternatives. As SNR increased, Iclust boosted quite quickly and greatly surpassed other competing methods.

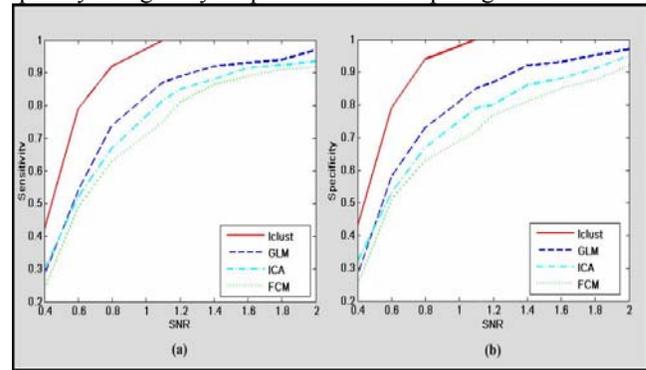


Fig. 3 Illustration of sensitivity and specificity performance of each method for different SNRs ranging from 0.4 to 2. (a) Sensitivity; (b) Specificity.

4.2. In Vivo Dataset

In this study, we analyzed the optimal solution with $\{N_c = 20, 1/T = 25\}$ for Iclust that obtained the highest value $\langle s \rangle$, where 89% voxels had nearly deterministic assignments to one of the clusters. After 50 different random initializations, the local maxima of F was determined. As to ICA and FCM, the number of clusters was 30, and the significance threshold in the GLM was $P < 0.001$. Fig. 4 (a)-(d) show that activated voxels, detected by different methods on the 19th slice of visual fMRI datasets, cover the primary visual cortex when they were superimposed on the accompanied structural data thresholded at $z = 3$ ($P < 0.001$, uncorrected). Specifically, four component maps of Iclust, had highly correlated time course with the experimental paradigm, though presented spatially different distributions. For convenient comparison with other methods, deliberate combination of these four maps into one was carried out, results shown in Fig. 4(a) (the four component maps would be discussed respectively in later paragraph). The “interesting” clusters by Iclust were highly evaluated by both robustness and stronger correlation with the stimulus (averaged $r = 0.76$). Though anatomically connected with visual cortical areas, spatial activation patterns detected by other methods, failed to present any significant activation in relevance with the stimulation paradigm, which pushed itself into hard interpretation.

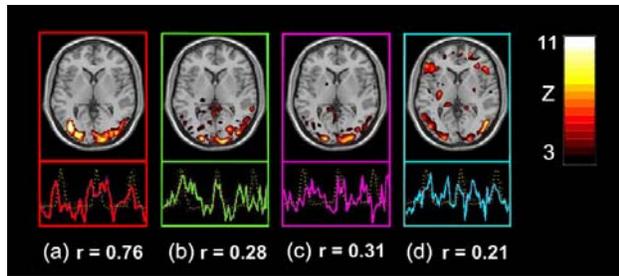


Fig.4 Cluster plot obtained from a representative subject detected by (a) Iclust (b) GLM (c) ICA (d) FCM from a visual task. Activation detection was done at the specificity level of 99.9% ($P < 0.001$, uncorrected). The average (normalized) time-courses were also reported with colors (temporal time series shown in solid line and reference function in dashed line).

For Iclust, four clusters had highly correlated time courses with the experimental paradigm. For clear presentation, consolidated map was color-coded in Fig. 5 (a) according to their different spatial layout, with respective clusters and corresponding correlation coefficients shown in Fig. 5 (b)-(e). These evidence demonstrated that the visual cortex actually consisted of subregions behaving differently, but other methods limited the extent, to which significantly activated voxels could be interpreted in terms of their specific functional involvement. Such subdivision presented in this study could be interpreted by both anatomical and functional studies: V1 is mainly involved in the initial encoding of a visual stimulus; V5/MT is known as motion-sensitive regions; V2/3 represent a mediating area transmitting information between V1 and V5/MT consistent with neuroanatomical models of visual motion perception [7]. The anatomical layout and temporal profile indicated that this further separation was neurophysiologically meaningful, not an artificial splitting.

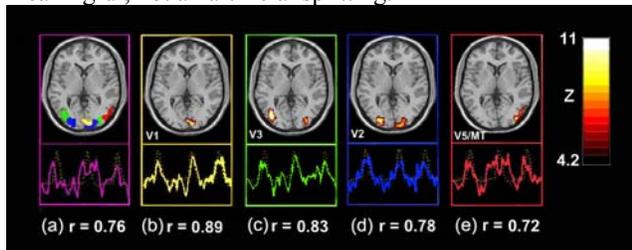


Fig.5 Task-related clusters detected by the proposed method for a representative subject ($P < 0.0005$, uncorrected). Averaged time course (solid line) and reference function (dashed line) were also presented. (a) Composite map of the four clusters with color coding: yellow, the primary visual cortex V1 (BA 17); blue, the extrastriate visual cortex V2 (BA 18); green, located in V3 (BA 19); red, located in V5/MT (BA 19 and BA 37). (b) - (e) activation location corresponding to the color-coded clusters.

5. DISCUSSION AND CONCLUSION

In summary, this study demonstrated the feasibility of a new attempt, and addressed limitation of potential assumptions of data structures involved in previous fMRI clustering analysis by means of the information-based clustering method. Two aspects made this application attractive: the reliable, tight clusters emerged both algorithmically and statistically by capturing any type of dependencies, strongly nonlinear structures included; the intrinsic data substructure could be automatically discovered and presented in a fine separation of the neurocognitive processes involved in the visual task. Our results significantly demonstrated that not all voxels detected by conventional algorithms ought to be functionally homogeneous, and could be further divided into several units with neurophysiologically functional involvement, regarded as preliminary in nature and needed to be replicated in prospective research.

This basic analytical framework was applied in demonstrating the functional subdivision of the human visual system, which generally displayed a high signal-to-noise ratio (SNR) compared to other systems, such as neurocognitive system. In reference to our attempt, prospective research remains to be illustrated whether Iclust can produce equally encouraging performance in other cortical networks with lower SNRs. Apparently, the author would not argue the optimum of our choice of clustering pattern, regarding some well-defined criteria. Therefore, seeking the answers of how to define a relevant criterion and developing an optimal segmentation posed another challenge for further researches.

6. REFERENCES

- [1] E. Moser, M. Diemling, R. Baumgartner, "Fuzzy clustering of gradientecho functional MRI in the human visual cortex: Part II. Quantification," *J Magn Reson Imaging*, vol. 7 (6), pp. 1102– 8, 1997.
- [2] Dietmar Cordes, Vic Haughton, John D. Carew, Konstantinos Arfanakis and Ken Maravilla, "Hierarchical clustering to measure connectivity in fMRI resting-state data," *J Magnetic Resonance Imaging*, vol. 20, Issue 4, pp. 305-317, 2002.
- [3] K.H. Chuang, M.J. Chiu, C.C. Lin, J.H. Chen, "Model-free functional MRI analysis using Kohonen clustering neural network and fuzzy C-means," *IEEE Trans Med Imaging*, vol. 18 (12), pp. 1117-1128, 1999.
- [4] Noam Slonim, Gurinder Singh Atwal, Gas'per Tkac'ik, and William Bialek, "Information-based clustering," *Proc. Natl. Acad. Sci. USA*, vol. 102 (51), pp. 18297-18302, 2005.
- [5] T. M. Cover, & J. A. Thomas, *Elements of Information Theory*, Wiley, New York, 1991.
- [6] N. Slonim, G. S. Atwal, G. Tka'cik, & W. Bialek, "Estimating mutual information and multi-information in large networks," *cs.IT/0502017*, 2005.
- [7] LC. Sincich, JC. Horton, "The circuitry of V1 and V2: integration of color, form, and motion," *Annu Rev Neurosci*, 2004.