Incorporating Spatial Priors into an Information Theoretic Approach for fMRI Data Analysis

Junmo Kim[1], John Fisher[1,2], Andy Tsai[1], Cindy Wible[3], Alan Willsky[1], William Wells[2,3]

[1]Laboratory for Information and Decision Systems
Massachusetts Institute of Technology
Cambridge, MA, USA

[2]Artificial Intelligence Laboratory
Massachusetts Institute of Technology
Cambridge, MA, USA

[3]Department of Radiology
Brigham and Women's Hospital
Harvard Medical School
Boston, MA, USA

Abstract

In previous work a novel information-theoretic approach was introduced for calculating the activation map for fMRI analysis [Tsai et al, 1999]. In that work the use of mutual information as a measure of activation resulted in a nonparametric calculation of the activation map. Nonparametric approaches are attractive as the implicit assumptions are milder than the strong assumptions of popular approaches based on the general linear model popularized by Friston et al [1994]. Here we show that, in addition to the intuitive information-theoretic appeal, such an application of mutual information is equivalent to a hypothesis test when the underlying densities are unknown. Furthermore we incorporate local spatial priors using the well-known Ising model thereby dropping the implicit assumption that neighboring voxel time-series are independent. As a consequence of the hypothesis testing equivalence, calculation of the activation map with local spatial priors can be formulated as mincut/maxflow graph-cutting problem. Such problems can be solved in polynomial time by the Ford and Fulkerson method. Empirical results are presented on three fMRI datasets measuring motor, auditory, and visual cortex activation. Comparisons are made illustrating the differences between the proposed technique and one based on the general linear model.

Motivation

- The exact nature of the relationship between the experimental protocol and the measured fMRI signal are not well known. An information-theoretic approach is appealing in that it is principled while making few assumptions about the time structure of the fMRI signal as it relates to the protocol.
- In practice, activation decisions at each voxel are assumed to be independent of neighboring voxels. Spurious responses are then removed by *ad hoc* techniques (e.g. morphological operators). We describe an automatic maximum a posteriori (MAP) detection method where the well-known Ising model is used to model neighboring voxel dependencies.

Brief Description of Previous Approach

- We compute nonparametric mutual information (MI) between the protocol and voxel time series to compute the fMRI activation map.
- Mutual information, a commonly used concept in communications theory, measures the degree to which one signal explains another (in our case the degree to which the protocol explains the observed fMRI data). It is a measure of whether signals (or random variables) are statistically dependent (as opposed to linearly dependent).

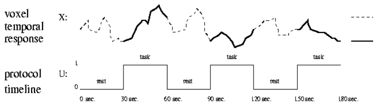


Fig. 1. Illustration of the protocol time-line, $S_{X|U=0}$, and $S_{X|U=1}$.

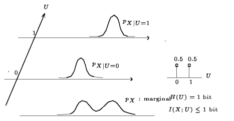


Figure 2.3: Illustration of $p_{X|U=0}, p_{X|U=1}, p_X$, and p_U

$S_{X|U=0}$ denotes the set of X_i's where the protocol is 0 while $S_{X|U=1}$ denotes the set of X_i's where the protocol is 1.
We assume that each voxel has its own two pdfs, $p_{X|U=0}(x), p_{X|U=1}(x)$ such that $S_{X|U=[0,1]}$ are i.i.d. according to $p_{X|U=[0,1]}(x)$.

- The MI between X and U is as follows:

$$I(X;U) = H(U) - H(U \mid X) = h(X) - h(X \mid U) = h(X) - \frac{1}{2}h(X \mid U=0) - \frac{1}{2}h(X \mid U=1)$$

- All terms are functions of the unknown underlying density. These are estimated non-parametrically (using the Parzen window density estimate) as:

$$h(X) = -\int_S p_X(x)\log p_X(x)dx \approx -\frac{1}{n}\sum_{i=1}^{n}\log\hat{p}_X(x_i) \qquad \hat{p}_X(x_i) = \frac{1}{n\sigma}\sum_j k\left(\frac{x_i - x_j}{\sigma}\right) \qquad k(x) \text{ is a unimodal density}$$

MI as a Nonparametric Hypothesis Test

- Consider the following hypothesis test (where x_t is the voxel time-series and u_t is the protocol encoding).

$$\begin{aligned} H_0 &: x_t, u_t \sim p_X(x)p_U(u) \quad \text{, i.e. } X, U \text{ are statistically independent} \\ H_1 &: x_t, u_t \sim p_{X,U}(x,u) \quad \text{, i.e. } X, U \text{ are statistically dependent} \end{aligned}$$

- The log-likelihood ratio, which is the optimal test statistic by the Neyman-Pearson lemma is equivalent to an estimate of mutual information.

$$T_n = \sum_{t=1}^{n}\log\left(\frac{p_{X,U}(x_t,u_t)}{p_X(x_t)p_U(u_t)}\right) \approx nE\left[\log\left(\frac{p_{X,U}(X,U)}{p_X(X)p_U(U)}\right)\right] = nI(X;U)$$

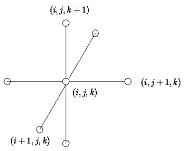
- More precisely, the likelihood ratio can be approximated as follows:

$$\frac{p_{\underline{X}|H_1}(X_1,\ldots,X_n \mid H_1)}{p_{\underline{X}|H_0}(X_1,\ldots,X_n \mid H_0)} \approx e^{n(\hat{I}(X;U)-\gamma)}$$

$$\text{where } \gamma = \frac{1}{2}\left[D(\hat{p}_{X|U=0} \| p_{X|U=0}) + D(\hat{p}_{X|U=1} \| p_{X|U=1})\right] - D(\hat{p}_X \| p_X).$$

Incorporating Spatial Priors (via the Ising Model)

- Let $\Omega = \{\omega : \omega \in \{0,1\}^{N_1 \times N_2 \times N_3}\}$ be the set of all possible [0-1] configurations and let $\omega(i,j,k)$ be a component of any one sample configuration $\omega(\cdot,\cdot,\cdot)$.
- The prior probability of a 3D activation map $Y(\cdot,\cdot,\cdot)$ is given as follows:



$$P(Y(\cdot,\cdot,\cdot) = \omega) = \frac{1}{Z}e^{-U(\omega)} \qquad Z = \sum_{\omega\in\Omega} e^{-U(\omega)}$$

$$U(\omega) = \beta\sum_{i,j,k}\left(\omega(i,j,k)\oplus\omega(i+1,j,k) + \omega(i,j,k)\oplus\omega(i,j+1,k) + \omega(i,j,k)\oplus\omega(i,j,k+1)\right)$$

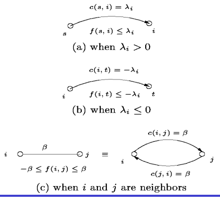
- Formulation: The MAP problem is as follows after a little algebraic manipulation:

$$\hat{Y}(\cdot,\cdot,\cdot) = \arg\max_{y(\cdot,\cdot,\cdot)} \sum_{i,j,k} n(\hat{I}_{i,j,k}(X;U) - \gamma)\,y(i,j,k) - \beta\sum_{i,j,k}\left(y(i,j,k)\oplus y(i+1,j,k) + y(i,j,k)\oplus y(i,j+1,k) + y(i,j,k)\oplus y(i,j,k+1)\right)$$

- Intuitive understanding of the relationship of γ and β. Suppose the 6 neighbors of a voxel are declared to be active. The length of the time series is n=60. Consquently, β=1, corresponds to an s = 6 β/n = 0.1 nats = 0.14 bits lowering of the effective threshold.

$$n(\hat{I} - \gamma) - 6\beta > 0 \qquad n\left(\hat{I} - (\gamma + \frac{6\beta}{n})\right) > 0$$

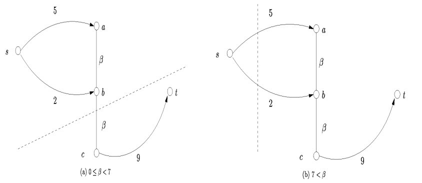
Exact Solution to the MAP Problem (due to Greig et al.)



(a) when $\lambda_i > 0$

(b) when $\lambda_i \leq 0$

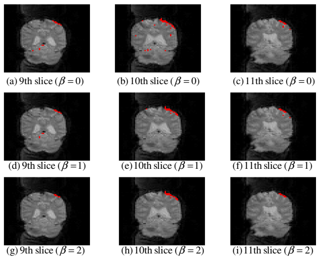
(c) when i and j are neighbors



(a) $0 \leq \beta < 7$

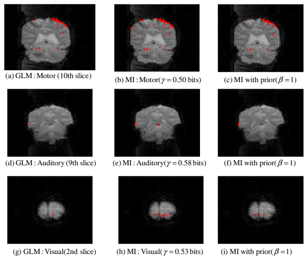
(b) $7 < \beta$

fMRI Comparisons and Results



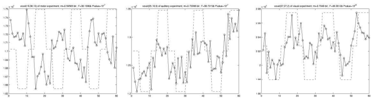
(a) 9th slice ($\beta = 0$) (b) 10th slice ($\beta = 0$) (c) 11th slice ($\beta = 0$)
(d) 9th slice ($\beta = 1$) (e) 10th slice ($\beta = 1$) (f) 11th slice ($\beta = 1$)
(g) 9th slice ($\beta = 2$) (h) 10th slice ($\beta = 2$) (i) 11th slice ($\beta = 2$)

Fig. 9th, 10th, and 11th slices of the motor cortex experiments for different values of β and $\gamma = 0.6$ bits. Detections are denoted as red pixels.



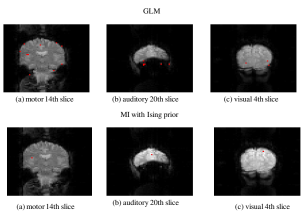
(a) GLM : Motor (10th slice) (b) MI : Motor($\gamma = 0.50$ bits) (c) MI with prior($\beta = 1$)
(d) GLM : Auditory (9th slice) (e) MI : Auditory($\gamma = 0.58$ bits) (f) MI with prior($\beta = 1$)
(g) GLM : Visual(2nd slice) (h) MI : Visual($\gamma = 0.53$ bits) (i) MI with prior($\beta = 1$)

Fig. 2. Comparison of fMRI Analysis results from motor, auditory, and visual experiments



(a) motor (b) auditory (c) visual

Fig. 3. Temporal responses of voxels newly detected by the MI using Ising prior



(a) motor 14th slice (b) auditory 20th slice (c) visual 4th slice
(a) motor 14th slice (b) auditory 20th slice (c) visual 4th slice

Fig. 4. Comparison of fMRI Analysis results from motor, auditory, and visual experiments with lowered GLM threshold

Conclusions

- The previously presented information theoretic approach, can be formulated in a hypothesis testing framework.
- The consequence of the hypothesis testing interpretation is that spatial priors can be incorporated via the Ising model resulting in a MAP estimation problem. The specific form is reducible to a minimum cut problem which can be solved exactly in polynomial time.
- MI approach with spatial priors was able to detect "true" activations with a smaller number of spurious responses.

Future Work

- The approach can be extended to any graph composed of doubleton potentials (beyond local neighborhoods).
- Validation over a larger set of data.

References

D. M. Greig, B. T. Porteous, and A. H. Seheult. Exact maximum a posteriori estimation for binary images. *Journal of the Royal Statistical Society. Series B(Methodological)*, 51(2):271-279, 1989.

A. Tsai, J. Fisher III, C. Wible, W. I. Wells, J. Kim, and A. Willsky. Analysis of fMRI data using mutual information. In C. Taylor and A. Colchester, editors, *Second International Conference on Medical Image Computing and Computer-Assisted Intervention*, volume 1679 of *Lecture Notes in Computer Science*, pages 473-480. Springer, Sep 1999.