

KCCA for fMRI Analysis

David R. Hardoon^{a*}, John Shawe-Taylor^a and Ola Friman^b

^aSouthampton University, Southampton, UK., ^bHarvard Medical School, Boston, USA[†]

Abstract. We use Kernel Canonical Correlation Analysis (KCCA) to infer brain activity in functional MRI by learning a semantic representation of fMRI brain scans and their associated activity signal. The semantic space provides a common representation and enables a comparison between the fMRI and the activity signal. We compare the approach against Canonical Correlation Analysis (CCA) by localising “activity” on a simulated null data set. Finally we present an approach to reconstruct an activity signal from a testing-set fMRI scans (both simulated and real), a method which allows us to validate our initial analysis.

1 Introduction

Understanding the functional processes of the brain is still a new and difficult task. Functional Magnetic Resonance Imaging (fMRI) is a relatively new tool with the purpose of mapping the sensor, motor and cognitive tasks to specific regions in the brain. The underlying mechanics of this technique is in the regulation of the blood flow as an excess of oxygen is supplied to active neurones causing an increase in oxygenated blood surrounding the tissue of the active brain region. This effected is referred to as BOLD (Blood Oxygenation Level Dependent) signal.

We present a Kernel CCA (KCCA) approach to measure the active regions of the brain using fMRI scans and their activity signal. Friman et. al [1] have shown that CCA can gives us the ability to introduce several time-courses as the BOLD response has been shown to vary both between people and brain regions. In previous work [2] we have shown that applying kernel methods [3] can increase the performance of CCA. Finally we show that due to the properties of KCCA [4] we are able to use this approach to reconstruct the activity signal from an “unknown” testing-set fMRI scans a process that allows us to validated our prior analysis.

The paper is divided as follows, Section 2 gives a brief introduction the method used and the baseline comparison. Section 3 describes the experiments taken place and Section 4 brings forward final discussion.

2 Method

Proposed by H. Hotelling in 1936 [5], Canonical correlation analysis seeks a pair of linear transformations one for each of the sets of variables such that when the set of variables are transformed the corresponding co-ordinates are maximally correlated [6]. Let

$$\rho = \max_{\mathbf{w}_x, \mathbf{w}_y} \frac{\mathbf{w}_x' C_{xy} \mathbf{w}_y}{\sqrt{\mathbf{w}_x' C_{xx} \mathbf{w}_x \mathbf{w}_y' C_{yy} \mathbf{w}_y}} \quad (1)$$

the maximum canonical correlation is the maximum of ρ with respect to \mathbf{w}_x and \mathbf{w}_y . We represent the two time-courses as a linear combination of pixel time-course

$$\mathbf{x}(\mathbf{t})\mathbf{w}_x = x_1(t)w_{x_1} + \dots + x_m(t)w_{x_m}$$

and any chosen time sequence to represent the fMRI modal

$$\mathbf{y}(\mathbf{t})\mathbf{w}_y = y_1(t)w_{y_1} + \dots + y_m(t)w_{y_m}.$$

CCA may not extract useful descriptors of the data because of its linearity. Kernel CCA offers an alternative solution by first projecting the data into a higher dimensional feature space (where $n < N$)

$$\phi : \mathbf{x} = (\mathbf{x}_1, \dots, \mathbf{x}_n) \mapsto \phi(\mathbf{x}) = (\phi_1(\mathbf{x}), \dots, \phi_N(\mathbf{x}))$$

before performing CCA in the new feature space, essentially moving from the primal to the dual representation approach.

*Correspondence author (Hardoon) at drh@ecs.soton.ac.uk.

[†]The work was done while Ola Friman was at Linkoping University, Sweden

Definition 1. $\langle \cdot, \cdot \rangle$ denotes the Euclidean inner product of the vectors \mathbf{x}, \mathbf{y} and is equal to $\mathbf{x}'\mathbf{y}$. Where we use A' to denote the transpose of a vector or matrix A .

Kernels are methods of implicitly mapping data into a higher dimensional feature space, a method known as the “kernel trick”. A kernel is a function K , such that for all $x, z \in \mathcal{X}$

$$K(x, z) = \langle \phi(x), \phi(z) \rangle \quad (2)$$

where ϕ is a mapping from \mathcal{X} to a feature space F . The weights can be written as a linear combination of the training examples, let

$$\mathbf{w}_x = \phi(\mathbf{x})\alpha. \quad (3)$$

Hence we obtain from CCA (for details [4, 6]) the dual representation

$$\rho = \max_{\alpha, \beta} \frac{\alpha' K_x K_y \beta}{\sqrt{\alpha' K_x^2 \alpha \beta' K_y^2 \beta}}. \quad (4)$$

As KCCA requires two views we provide K_x as a kernel from the fMRI brain scan and K_y as a kernel from the chosen time-sequence. In [6] we observe that with full rank kernel matrices maximal correlation can be obtained, suggesting that learning is trivial. To force non-trivial learning we introduce a control on the flexibility of the projection mappings by penalising the norms of the associated weights (detailed description of CCA and KCCA can be found in [4]), we obtain

$$\rho = \max_{\alpha, \beta} \frac{\alpha' K_x K_y \beta}{\sqrt{(\alpha' K_x^2 \alpha + \kappa \alpha' K_x \alpha)(\beta' K_y^2 \beta + \kappa \beta' K_y \beta)}}. \quad (5)$$

3 Experiments

3.1 Activity Localisation

We compare KCCA to the baseline CCA as presented in [1, 7, 8, for brevity we refer the reader to the papers for details on the CCA method]. The regularisation parameter κ from equation (5) is computed a priori as described in [2, 6]. In the following experiment we use the correlation values computed by CCA though in KCCA we prefer to compute the weights associated to the pixels, as this can give us more information on the activity of each pixel, this step can also be done with CCA although we do not compute it as this approach is not as intuitive as with KCCA.

As it is impossible to tell which method is better when real data is used, we experiment with controlled simulated data. We embed square-wave “activity” in a nulldata set (no brain activity). The paradigm of the applied activity is 10 images rest, 10 images activity and so forth. Resulting with 200 time points. As we know the activation period we use for our time sequence a square-wave representation of activity (1) and rest (−1) over the 200 time-course. We use a linear kernel for both the fMRI data and the square-wave “activity”.

In figure 1 the found true-positive and false-positive pixels using CCA are plotted we are able to observe that although there is a decline in the number of false-positive pixels located, the number of false-positive outnumbers the number of true-positive pixels. At a threshold of 0.74 we are able to observe in figure 1(right plot) that the rate of false-positive drop below that of the true-positive but not in a significant measure. In figure 2 the found true-positive and false positive pixels using KCCA are plotted. As expected the number of false-positive pixels start at a much higher rate then that found with CCA as with KCCA we take into account all the pixels in the image. Although this worse start we are able to observe a sharp drop in the number of false positive pixels accompanied with a stedy and slower drop in the number of true-positive pixels. We also find that the number of true-positive pixels located surpasses the number of false-positive, with a relatively low threshold. This suggests that KCCA is able to extract a better ratio of true to false positive pixels.

3.2 Statistical Reconstruction

In the following section we present an approach of statistically reconstructing a signal from the fMRI scans. This reconstruction approach will allow us to determine the validity of our prior analysis for if we have learnt the

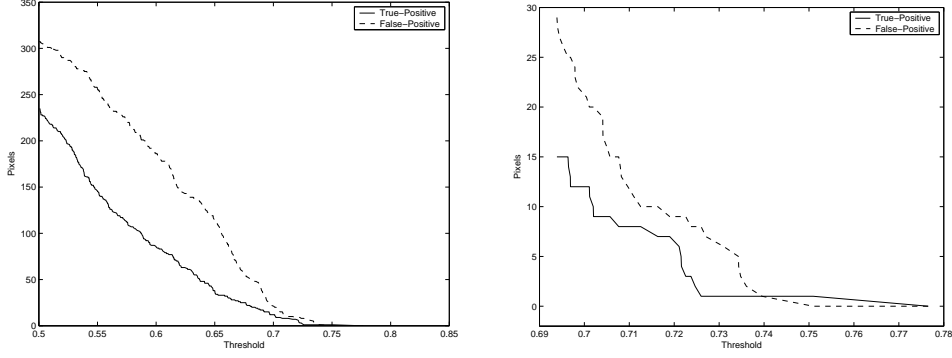


Figure 1. Left: CCA Found True/False-Positives. Right: Same Plot, Zoomed-In.

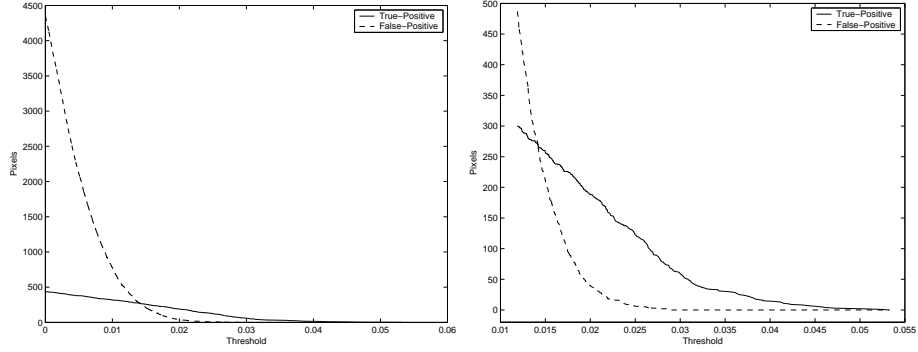


Figure 2. Left: KCCA Found True/False-Positives. Right: Same Plot, Zoomed-In.

appropriate function we will be able to reconstruct it. Let X be the fMRI training examples and X_t the fMRI testing examples, Y be the training activity time sequence and Y_t the testing activity signal we want to reconstruct. Let

$$g_{w_x, w_y} = \|Xw_x - Yw_y\|^2 \quad (6)$$

where $g_{w_x, w_y} \approx 0$ as we want the feature Xw_x from one view of the data to be identical to the feature Yw_y obtained from the second view of the data, this will be true on the training data if there is a high correlation between the two views. Therefore we can rewrite equation (6) as

$$\|Xw_x - Yw_y\|^2 \approx 0 \quad (7)$$

$$Xw_x \approx Yw_y \quad (8)$$

We divide the fMRI data into a training and testing set. On the training data we compute the KCCA coefficients α, β . Let $K_{x_t} = \langle X_t, X \rangle$ be the fMRI testing kernel and $K_{y_t} = \langle Y_t, Y \rangle$ be the time sequence testing kernel. [4] have shown that this equivalence can be held true also for the testing data using efficient regularisation. Hence we justify the usage of equation (8) and equation (3) to define

$$\begin{aligned} X_t w_x &\approx Y_t w_y \\ K_{x_t} \alpha &\approx Y_t' Y \beta. \end{aligned}$$

As we are interested in finding the testing-set unknown activity time sequence we can rearrange the equation to

$$Y_t \approx (K_{x_t} \alpha \cdot (Y \beta)^{-1})'. \quad (9)$$

As we are no longer interested in the weight vector but in the reconstruction of the signal, we are not confined to the usage of inner product kernels, in the following experiment we compare the success rate between the linear kernel as used, to the Gaussian kernel (defined in equation (10))

$$K(x_i, x_j) = \exp \frac{-\|x_i - x_j\|^2}{\sigma^2} \quad (10)$$

using σ as the minimum distance between the different labelled images.

We test our approach using the square-wave time sequence of 1 representing activity and -1 representing rest for both the simulated and real data experiments. The real data is comprised of mental calculation, the adding of two numbers, and right hand index finger flexing. For the simulated data and mental calculation we use the first 160 scans for training and the remaining 20 for testing, while with the finger flexing we use the first 180 for training and the remaining 20 for testing. We randomise the examples prior to the training and testing separation. Once we obtain the reconstructed Y_t we threshold it by $T = 0$, i.e. Let \hat{Y} be the thresholded reconstructed signal

$$\hat{Y} = \begin{cases} 1 & \text{if } Y_t \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

Table 1 shows the average overall results of successfully reconstructing the activity time sequence for the testing fMRI data over 100 repeats using both a linear & Gaussian kernels. We are able to see that the linear kernel performs better than the Gaussian. It is important to state at this stage the difference between CCA and kernel CCA (KCCA) with a linear kernel. The former uses a larger number of features than in CCA, which are computed implicitly in the kernel.

Table 1. Success rate in reconstructing the test activity time sequence

Data-Set	Linear	Gaussian
Simulated data	97.4%	95.8%
Finger flexing	75.25%	69.25%
Mental calculation	48.6%	41.55%

We provide an initial experiment in attempt to learn the mental process prior to the finger flexing by setting the square-wave sequence such that the three images before the actual finger flexing were considered as active and all the remaining images were considered as inactive. We have trained as before using a linear kernel and attempted to reconstruct this new square-wave sequence over an average of 100 random repeats. We find that we can successfully reconstruct the signal with an success average of 83.3%.

4 Discussion

For future work we would like to try more elaborate time basis functions and to experiment on different data types (emotional, mental and other motor data) and tailored kernels for better extracting the activity/signal. A further interesting avenue would be to observe the performance of applying our KCCA approach to other techniques of brain analysis and also to more complex tasks. We also speculate that KCCA would be able to handle a multiple task fMRI scenario.

Acknowledgements

The first and second authors would like to acknowledge the financial support of EU Project LAVA, No. IST-2001-34405 and thank Craig Saunders for many useful discussions.

References

1. O. Friman, J. Carlsson, P. Lundberg et al. "Detection of neural activity in functional MRI using canonical correlation analysis." *Magnetic Resonance in Medicine* **45(2)**, pp. 323–330, February 2001.
2. D. R. Hardoon & J. Shawe-Taylor. "KCCA for different level precision in content-based image retrieval." In *Proceedings of Third International Workshop on Content-Based Multimedia Indexing*. IRISA, Rennes, France, 2003.
3. N. Cristianini & J. Shawe-Taylor. *An Introduction to Support Vector Machines and other kernel-based learning methods*. Cambridge University Press, 2000.
4. J. Shawe-Taylor & N. Cristianini. *Kernel Methods for Pattern Analysis*. Cambridge University Press, 2004.
5. H. Hotelling. "Relations between two sets of variates." *Biometrika* **28**, pp. 312–377, 1936.
6. D. R. Hardoon, S. Szedmak & J. Shawe-Taylor. "Canonical correlation analysis; an overview with application to learning methods." Technical Report CSD-TR-03-02, Royal Holloway University of London, 2003.
7. O. Friman, M. Borga, P. Lundberg et al. "A correlation framework for functional MRI data analysis." In *Proceedings of the 12th Scandinavian Conference on Image Analysis*. SCIA, Bergen, Norway, June 2001.
8. O. Friman. *Adaptive Analysis of Functional MRI Data*. Ph.D. thesis, Linköping Studies in Science and Technology, 2003.