Semi-supervised Kernel Canonical Correlation Analysis of Human Functional Magnetic Resonance Imaging Data

Jacquelyn A. Shelton

Max Planck Institute for Biological Cybernetics and Universität Tübingen, Tübingen, Germany

Women in Machine Learning Workshop
December 7th, 2009

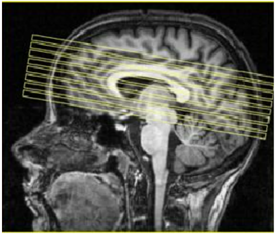
MAX-PLANCK-GESELLSCHAFT

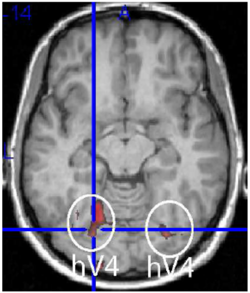
BIOLOGISCHE KYBERNETIK

Introduction

Motivation

- **Neuroscience: assess natural processing**, i.e. fMRI – reduce dimensions to main activity during shown stimulus



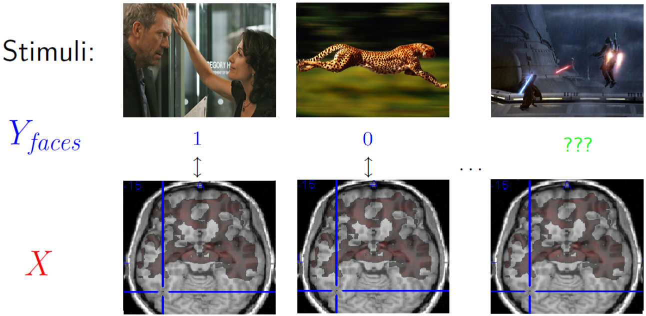


- **Problems:** high-dimensional data, expensive labels
- **Goal:** Canonical Correlation Analysis in semi-supervised learning framework

Paired Data

- Samples in **2 modalities**: representations of 1 process,
  $\longrightarrow$ labeled video shown during fMRI acquisition

Illustration:

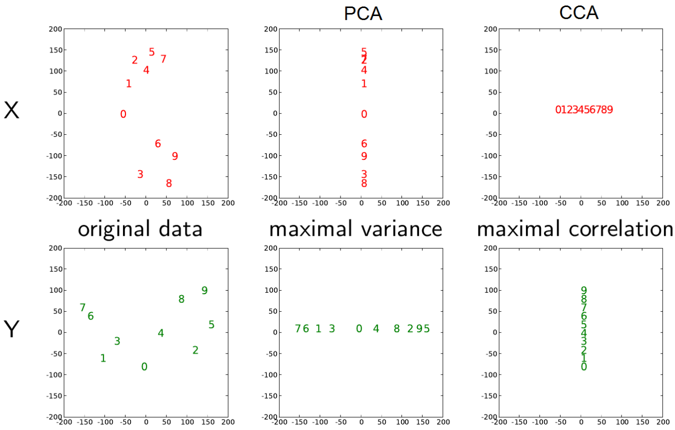


fMRI data: (labeled) $X = \{x_1, x_2, \ldots, x_n\}$, (unlabeled) $\{x_{n+1}, \ldots, x_p\}$
Corresponding labels: $Y = \{y_1 = 1, y_2 = 0, \ldots, y_n\}$

$\longrightarrow$ Paired data (fMRI with labels): $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$

Canonical Correlation Analysis (CCA)

- Finds projection directions in each modality's subspace that **maximize correlation** between the projected data
  $\longrightarrow$ Not directions of (potentially noisy) maximal variance



Kernel Canonical Correlation Analysis

- **CCA:** maximize correlation between X and Y projections

Optimize CCA e.g. as a generalized eigenvalue problem:

$$\max_{w_x, w_y} \quad \frac{w_x^T C_{xy} w_y}{\sqrt{(w_x^T C_{xx} w_x)(w_y^T C_{yy} w_y)}} \tag{1}$$

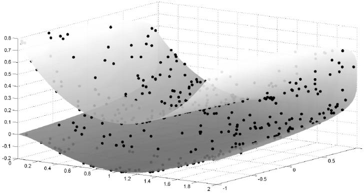
- **Kernelized CCA (KCCA)**: general, optimization easier
- **Regularized KCCA**: avoid degenerate solutions

Optimize Tikhonov regularized KCCA:

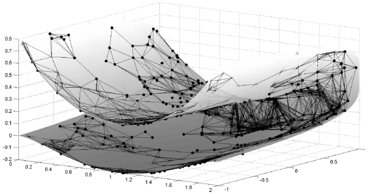
$$\max_{\alpha, \beta} \frac{\alpha^T K_x K_y \beta}{\sqrt{\alpha^T \left(K_x^2 + \varepsilon_x K_x\right) \alpha \beta^T \left(K_y^2 + \varepsilon_y K_y\right) \beta}} \tag{2}$$

Manifold assumption

- **Manifold assumption**: high-dimensional data lie on a low-dimensional manifold $\mathcal{M}$ (Belkin et al., 2006)
- Functions should vary smoothly along $\mathcal{M}$ – small gradient
- Estimate the gradient $\nabla_{\mathcal{M}}$ by constructing a graph along the manifold $\mathcal{M}$:



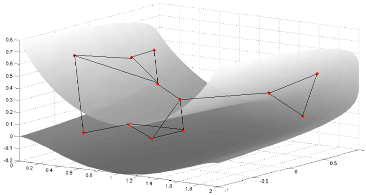
Samples of manifold



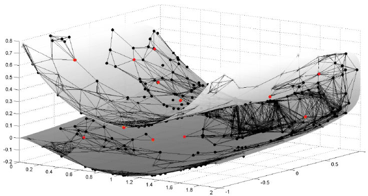
Graph estimate of manifold

Laplacian Regularization

- Gradient estimate $\nabla_{\mathcal{M}}$ of functions along $\mathcal{M}$ leads to **Laplacian regularization** – adding term $\mathcal{L}$ to optimization enforces smoothness along the manifold
- Optionally unlabeled data can be included to improve estimate of manifold $\rightarrow$ **semi-supervised**



Poor estimate: Graph with few data points



Better estimate: Graph with more data points

Laplacian regularized SSKCCA:

$$\max_{\alpha,\beta} \frac{\alpha^T K_{\hat{x}x} K_{yy} \beta}{\sqrt{\alpha^T \left(K_{\hat{x}x} K_{x\hat{x}} + R_{\hat{x}}\right) \alpha \beta^T \left(K_y^2\right) \beta}}, \tag{3}$$

with regularizers

$$R_{\hat{x}} = \underbrace{\varepsilon_x K_{\hat{x}\hat{x}}}_{\text{Tikhonov}} + \underbrace{\frac{\gamma_x}{m_x^2} K_{\hat{x}\hat{x}} \mathcal{L}_{\hat{x}} K_{\hat{x}\hat{x}}}_{\text{Laplacian}}$$

- SSKCCA will favor directions α and β whose projections are **smooth along the manifold** (Blaschko et al., 2008)

- **fMRI data (X):** human volunteer during viewing of 2 movies
  - 350 time slices of 3D fMRI brain volumes per movie
- **Labels (Y):** Continuous labels, 1 movie – 5 observers' scores:
  Faces - Color - Bodies - Language - Motion (Bartels and Zeki 2004)
- **Linear kernel** in all experiments

Experiments

(a) KCCA with Tikhonov regularization
$\longrightarrow$ labeled data only

(b) KCCA with Tikhonov and Laplacian regularization
$\longrightarrow$ labeled data only

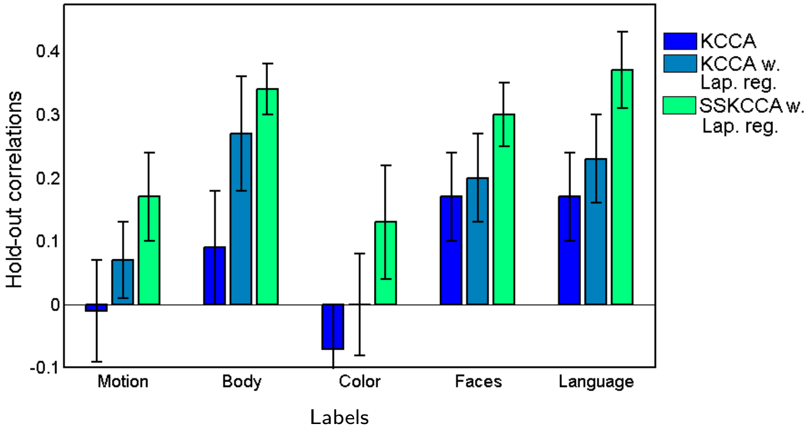
(c) SSKCCA with Tikhonov and Laplacian regularization
$\longrightarrow$ labeled and unlabeled data

- **Model Selection**: criterion from (Hardoon et al., 2004) to optimize over the regularization parameters (ε_x and γ_x)

Experiments

Results – Quantitative

Mean holdout correlations from five-fold cross validation across [each of the five] variables in all experiments.

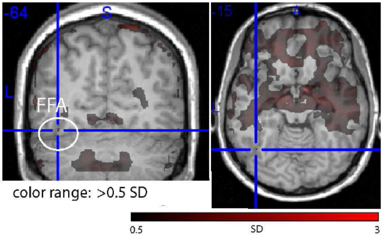


→ SSKCCA generalizes better than KCCA

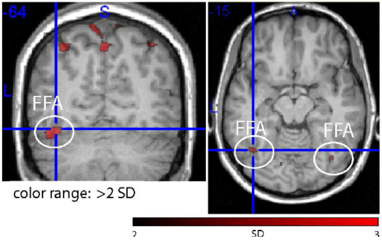
Experiments

Results – Qualitative

Visualization of learned weight vectors for faces



KCCA, Tikhonov regularization



SSKCCA, Tikhonov and Laplacian regularization

$\longrightarrow$ SSKCCA localizes regions of brain activity, following (Bartels and Zeki, 2004)

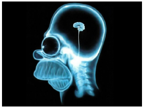
Summary

- SSKCCA learned expected regions of brain activity corresponding to input stimuli (Bartels and Zeki, 2004)
- KCCA with Laplacian regularization improves correlation by enforcing smoothness of projections along the manifold
- SSKCCA with use of unlabeled data further improves performance
- Check out **poster M26** for our extension of this work using resting state fMRI data as an unlabeled data source

Summary

- SSKCCA learned expected regions of brain activity corresponding to input stimuli (Bartels and Zeki, 2004)
- KCCA with Laplacian regularization improves correlation by enforcing smoothness of projections along the manifold
- SSKCCA with use of unlabeled data further improves performance
- Check out **poster M26** for our extension of this work using resting state fMRI data as an unlabeled data source

Thanks.



Appendix – References

1. Bartels, A., Zeki, S., and Logothetis, N. K. (2008). Natural vision reveals regional specialization to local motion and to contrast-invariant, global flow in the human brain. Cereb Cortex 18:705-717.
2. Bartels, A., Zeki, S. (2004). The chronoarchitecture of the human brain - natural viewing conditions reveal a time-based anatomy of the brain. NeuroImage 22:419-433.
3. Belkin, M., Niyogi, P., Sindhwani, V.: Manifold Regularization: A Geometric Framework for Learning from Labeled and Unlabeled Examples. JMLR (2006)
4. Blaschko, M.B., Lampert, C.H., Gretton, A. (2008). Semi-supervised Laplacian Regularization of Kernel Canonical Correlation Analysis. ECML
5. Hardoon, D. R., S. Szedmak and J. Shawe-Taylor. (2004). "Canonical Correlation Analysis: An Overview with Application to Learning Methods," Neural Computation, 16, (12), 2639-2664.
6. Friston, K., Ashburner, J., Kiebel, S., Nichols, T., Penny, W. (Eds.) Statistical Parametric Mapping: The Analysis of Functional Brain Images, Academic Press (2007)
7. Shelton, J., Blaschko, M., and Bartels, A. (05 2009). Semi-supervised subspace analysis of human functional magnetic resonance imaging data, Max Planck Institute Tech Report, (185) (05 2009)
8. Blaschko, M., Shelton, J., and Bartels, A. (12 2009) Augmenting Feature-driven fMRI Analyses: Semi-supervised learning and resting state activity. NIPS.

Appendix

Kernelization

$$\max_{w_x, w_y} \frac{w_x^T C_{xy} w_y}{\sqrt{w_x^T C_{xx} w_x \; w_y^T C_{yy} w_y}}. \tag{4}$$

We denote $\mathcal{H}_x$ the reproducing kernel Hilbert space (RKHS) associated with k_x, and denote the associated feature map $\phi_x : \mathcal{X} \to \mathcal{H}$, i.e. $k_x(x_i, x_j) = \langle \phi_x(x_i), \phi_x(x_j) \rangle$.

$$\max_{f_x, f_y} \frac{f_x^T \hat{C}_{xy} f_y}{\sqrt{f_x^T \hat{C}_{xx} f_x \; f_y^T \hat{C}_{yy} f_y}} = \max_{\alpha, \beta} \frac{\alpha^T K_x K_y \beta}{\sqrt{\alpha^T K_x^2 \alpha \; \beta^T K_y^2 \beta}}, \tag{5}$$

$$\max_{\alpha, \beta} \frac{\alpha^T K_x K_y \beta}{\sqrt{\alpha^T \left(K_x^2 + \varepsilon_x K_x\right) \alpha \beta^T \left(K_y^2 + \varepsilon_y K_y\right) \beta}}, \tag{6}$$

Denoting the kernel matrix computed using the data in X as $K_{xx} \in \mathbb{R}^{n \times n}$, the matrix computed using $\hat{X}$ and X as $K_{\hat{x}x} \in \mathbb{R}^{m_x \times n}$, the matrix computed using $\hat{X}$ with itself as $K_{\hat{x}\hat{x}} \in \mathbb{R}^{m_x \times m_x}$, etc. Kernel matrices for $\mathcal{Y}$ can be defined analogously.
Semi-supervised Laplacian regularized generalization of above equation:

$$\max_{\alpha, \beta} \frac{\alpha^T K_{\hat{x}x} K_{y\hat{y}} \beta}{\sqrt{\alpha^T \left(K_{\hat{x}x} K_{x\hat{x}} + R_{\hat{x}}\right) \alpha \beta^T \left(K_{\hat{y}y} K_{y\hat{y}} + R_{\hat{y}}\right) \beta}}, \tag{7}$$

Graph Laplacian term $\mathcal{L}$:

$$\mathcal{L} = D^{-1/2}(D - W)D^{-1/2}$$

where W is the matrix of similarities between data points
and D is the diagonal matrix with entries of W's row sums

for similarity kernel $(K)_{ij} = \exp\left(\frac{-\|x_i - x_j\|^2}{\sigma^2}\right)$
and diagonal of row sums $(D_{\hat{x}\hat{x}})_{ii} = \sum_{j=1}^{n+p_x}(K_{\hat{x}\hat{x}})_{ij}$.

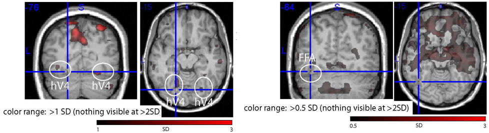
Appendix

Data and Acquisition

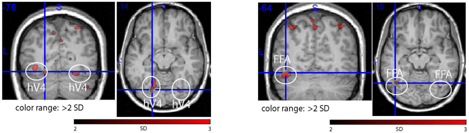
- fMRI data of one human volunteer during viewing of 2 movies.
- 350 time slices of 3-dimensional fMRI brain volumes acquired with Siemens 3T TIM scanner, separated by 3.2 s (TR), with a spatial resolution of 3x3x3 mm.
- Pre-processed according to standard procedures using the Statistical Parametric Mapping (SPM) toolbox [6].

Visualization of learned weight vectors (w_x) for color and face stimuli, following [2].

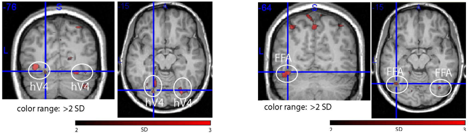
Color Faces



(a) CCA, Tikhonov regularization

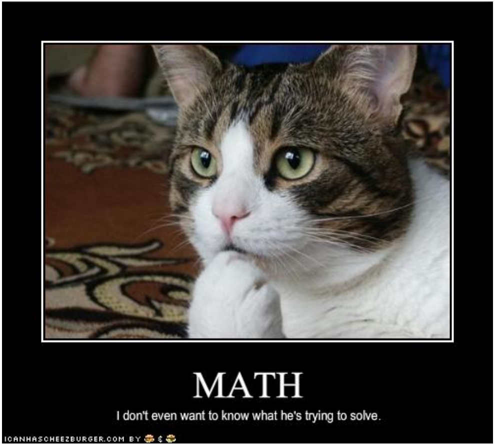


(b) CCA, Tikhonov and Laplacian regularization



(c) Semi-supervised CCA, Tikhonov and Laplacian regularization

Just a kitty


MATH
I don't even want to know what he's trying to solve.
ICANHASCHEEZBURGER.COM BY