A Self-Supervised Voxel Shuffling Framework for Kernel-Based fMRI activation detection

Chendi Han[1], Zhengshi Yang[1], Xiaowei Zhuang[1], and Dietmar Cordes[1]
[1]*Cleveland Clinic, Las Vegas, NV, United States*

Synopsis

Motivation: Kernel-based methods are efficient for global feature extraction and activation detection in fMRI analysis. Properly defining the kernel mapping function requires supervised training, which can easily overfit due to high dimensionality, even for linear kernels.

Goal(s): Design a data augmentation framework to generate supervisory examples without requiring ground truth or new datasets. The augmented data should preserve the original information, and by maximizing the similarity, we should get more accurate results.

Approach: By shuffling the voxels based on their activation properties.

Results: In two simulated and two real fMRI datasets, our method can select the best hyperparameters and generate more accurate activation patterns.

Impact: The proposed augmentation method effectively finds the optimal kernel mapping and mitigates overfitting in activation detection without relying on spatial ground truth information. This could be used in real fMRI data and broaden the potential for modeling more complex relationships.

Introduction

Kernel-based methods are powerful for dimension reduction in fMRI analysis. Previous methods used supervised optimization[1] or an additional dataset serving as negative pairs[2] to determine the unknown kernel mapping, which is less efficient and hard to expand to real datasets when there is no ground truth. Inspired by self-supervised learning in image processing[3], we propose a data augmentation method that could generate supervisory examples by voxel shuffling. By maximizing the similarity before and after augmentation, we show our method could reduce overfitting and increase accuracy even for complicated kernels. Results are proven by both simulated and real fMRI datasets in the activation detection problem.

Methods

Suppose fMRI data with $\mathbf{Y} \in \mathbb{R}^{T\times Q}$, where T indicates the time and Q is the total number of voxels, given a certain design signal $X_{\text{eff}} \in \mathbb{R}^{T\times 1}$. Kernel canonical correlation analysis (KCCA) solves this problem by mapping the original data to kernel space, maximizing the correlation and get activation patterns $\alpha \in \mathbb{R}^{1\times Q_4}$. This process could involve unknown functions or regularization parameters. For example, the linear kernel contains one additional regularization parameter γ to reduce overfitting[2]. Our goal is to find the augmented data $\mathbf{Y}'$, which could serve to determine the unknowns. Figure 2 shows our proposed algorithm. Starting from arbitrary kernel mapping and performing KCCA, we can divide the activation α into activated and non-activated clusters. We use $Q_+ = 0.1Q$ and $Q_{\text{Non}} = 0.9Q$ to indicate the number of voxels in each cluster. Our shuffling algorithm involves switching the voxel locations inside each cluster along the decision boundary. Specifically, we select Q_1 voxels in the non-activated cluster and Q_2 voxels in the activated cluster. Within the Q_1 voxels, the voxel with the highest α is changed to the lowest α, and the same for the second highest/lowest voxels, and so on. The same shuffling is repeated for Q_2. The detailed choice for Q_1 and Q_2 is shown in Figure 2. Practically, we use two shufflings and take the average. The data after shuffling is indicated by $\mathbf{Y}'$, which is then input to the KCCA to get another activation pattern α'. The similarity is measured by treating α as ground truth, then using different thresholds to compute the Receiver Operating Characteristic (ROC) curve, with area under the curve (AUC) with False Positive Rate (FPR) smaller than 0.1 defined as similarity. We implement the surrogate optimization algorithm in MATLAB to maximize it.

Results

We use two datasets to test the method: the HCP dataset with 87 subjects performing a working memory task[5], and our in-house scans with 16 subjects performing both faces and pictures episodic memory tasks[6]. In the HCP dataset, the design signal uses contrasts targets minus non-targets[7]. For our in-house scans for each subject and task, the contrasts encoding minus control and recognition minus control effectively give us 64 different combinations. Figure 3 shows the 5 different kernels considered in this paper. All the data are minimally preprocessed using the SPM12 package[8], normalized to the MNI atlas[9], with Gaussian smoothing FWHM=4 mm and 7 spatial orientation filters for kernel methods with the same smoothing level[2]. The simulation is performed based on 20 subjects in the HCP dataset, with the resting fMRI treated as noise. The first simulation added the design signal with contrast encoding minus control to 6 regions in the brain[10]. The second simulation added task fMRI signals from voxels with the top 1% correlation to the design signal, which were randomly shuffled and added to the locations of voxels with the top 10% highest correlations[11]. Both simulations adjust the signal level to match the AUC for Single Voxel Smoothing (SV)[9]. Results for the two datasets are shown in Figure 4 (a) and (b). We find a strong relationship between shuffling robustness and accuracy, proving our data augmentation algorithm can select the best kernel mapping by maximizing the similarities. Next, we perform the same computation on real datasets, with the criterion set to be the Gray Matter (GM) overlapping. Using gray matter as ground truth, different thresholds for α are used to evaluate the ROC curve. The results for two different datasets are shown in Figure 5 (a) and (b). A similar relationship between shuffling robustness and gray matter overlapping is observed.

Discussion

The main finding of the study is that we have proposed a robust way to do data augmentation. We find that by maximizing the similarities before and after augmentation, we can determine the best kernel mapping without using the spatial ground truth or additional datasets.

Acknowledgements

This study was funded by NIH-R01AG071566 and NIH-P20GM109025.

References

[1] Xifra-Porxas, Alba, et al. Estimating brain age from structural MRI and MEG data: Insights from dimensionality reduction techniques. NeuroImage, 2021. 231: p. 117822.

[2] Yang, Z., et al., 3D spatially-adaptive canonical correlation analysis: Local and global methods. Neuroimage, 2018. 169: p. 240-255.

[3] Chen, Ting, et al. "A simple framework for contrastive learning of visual representations." International conference on machine learning. PMLR, 2020: p. 1597-1607.

[4] Hardoon, D.R., et al., Canonical correlation analysis: An overview with application to learning methods. Neural computation 2004. 16: p. 2639-2664.

[5] Barch, D.M., et al., Function in the human connectome: task-fMRI and individual differences in behavior. Neuroimage, 2013. 80: p. 169-189.

[6] Jin, Mingwu, et al. "A preliminary study of functional abnormalities in aMCI subjects during different episodic memory tasks." Magnetic Resonance Imaging. 2012. 30.4: p. 459-470.

[7] Cordes, Dietmar, et al. "Optimizing the performance of local canonical correlation analysis in fMRI using spatial constraints." Human brain mapping. 2012. 33.11 : p. 2611-2626.

[8] J. Ashburner, Computational anatomy with the spm software, Magnetic resonance imaging. 2008, 27: p. 1163–1174

[9] M. F. Glasser et al., The minimal preprocessing pipelines for the human connectome project. Neuroimage, 2013. 80: p. 105–124

[10] Yang, Zhengshi, et al. A robust deep neural network for denoising task-based fMRI data: An application to working memory and episodic memory. Medical Image Analysis. 2020. 60: p. 101622.

[11] Cordes, Dietmar, et al. "Optimizing the performance of local canonical correlation analysis in fMRI using spatial constraints." Human brain mapping. 2012. 33.11: p. 2611-2626.

Figures

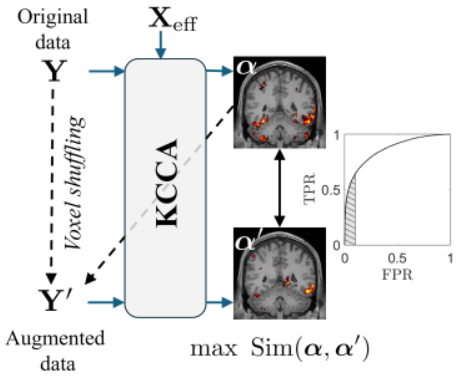
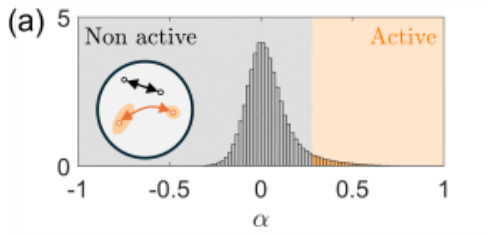
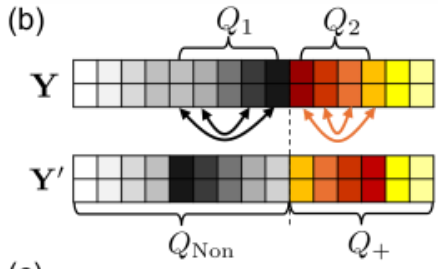


Figure 1: Illustration of the data augmentation method used in the paper. Original fMRI data $\mathbf{Y}$ is input to KCCA to get the activation α. After voxel shuffling, the augmented data is labeled as $\mathbf{Y}'$ with the corresponding activation α'. The similarity is characterized by the ROC curve, by treating the top 10% of α as truth activation and implementing different thresholds to α'.





(c)

Shuffling 1	$Q_1 = Q_2 = 0.5Q_+$
Shuffling 2	$Q_1 = Q_{\text{Non}}, Q_2 = Q_+$

Figure 2: Illustration of the voxel shuffling method. (a) Distribution of the activation α. Results do not change significantly for different data or kernels. Based on the output α, the top 10% of voxels are set to be activated. (b) Two clusters with numbers of voxels Q_1 and Q_2 are chosen around the boundary. Voxels inside each cluster are shuffled based on their α value. (c) The choice for the Q_1 and Q_2 used in the paper.

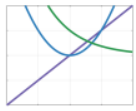
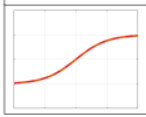
	Kernel name	Expression
	Linear	$K_{\mathbf{Y}} = \mathbf{Y}\mathbf{Y}^T$
	Parabolic	$K_{\mathbf{Y}} = (\mathbf{Y}\mathbf{Y}^T + b^2)^2$
	Gaussian	$K_{\mathbf{Y}} = \exp\left(-\lvert\lvert\mathbf{Y} - \mathbf{Y}^T\rvert\rvert^2/\sigma^2\right)$
	Tanh	$K_{\mathbf{Y}} = \tanh(b\mathbf{Y}\mathbf{Y}^T + c)$
	Mixed Tanh	$K_{\mathbf{Y}} = \tanh(b_1\mathbf{Y}\mathbf{Y}^T + b_2\lvert\lvert\mathbf{Y} - \mathbf{Y}^T\rvert\rvert^2 + c)$

Figure 3: Five kernels and their convergence properties used in this paper. Based on their convergence properties, the linear, Gaussian, and parabolic are the unbounded kernels, while the hyperbolic tangent and mixed hyperbolic tangent are the bounded kernels.

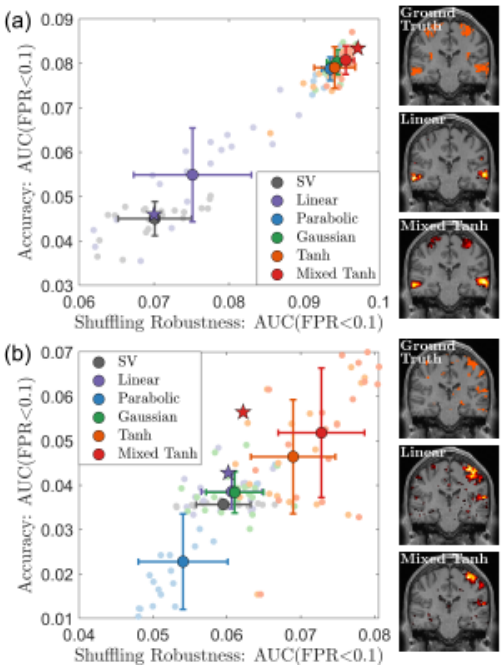


Figure 4: The relationship between shuffling robustness and accuracy for simulated data. Each small dot represents one subject analyzed using the specified methods, and the bigger one indicates the group-level average and error bar. Panels (a) and (b) correspond to the first and second simulations, respectively. In each simulation, one subject was selected (indicated by a star), and the corresponding activation patterns are displayed on the right. Compared with SV, the mixed hyperbolic tangent kernel increased the AUC by 80.58% and 45.19% in simulations one and two, respectively.

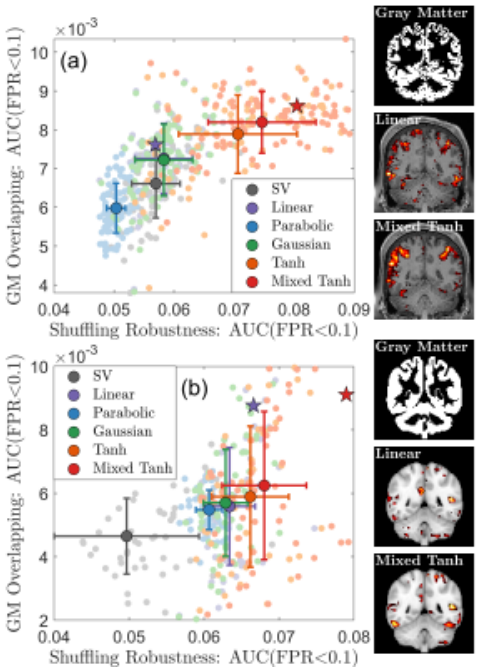


Figure 5: The relationship between shuffling robustness and gray matter overlapping for real data. Each small dot represents one subject analyzed using the specified methods, and the bigger one indicates the group-level average and error bar. Panels (a) and (b) correspond to the HCP and in-house scans, respectively. In each dataset, one subject was selected (indicated by a star), and the corresponding activation patterns are displayed on the right. Compared with SV, the mixed hyperbolic tangent kernel increased the AUC by 25.74% and 37.14% for HCP and in-house scans, respectively.