

Multi-subject brain decoding with multi-task feature selection

Liye Wang^a, Xiaoying Tang^a, Weifeng Liu^a, Yuhua Peng^a, Tianxin Gao^{a,*} and Yong Xu^{b,*}

^a*School of Life Science, Beijing Institute of Technology, Beijing 100081, China*

^b*Chinese PLA General Hospital (301 Hospital), Beijing 100853, China*

Abstract. In the neural science society, multi-subject brain decoding is of great interest. However, due to the variability of activation patterns across brains, it is difficult to build an effective decoder using fMRI samples pooled from different subjects. In this paper, a hierarchical model is proposed to extract robust features for decoding. With feature selection for each subject treated as a separate task, a novel multi-task feature selection method is introduced. This method utilizes both complementary information among subjects and local correlation between brain areas within a subject. Finally, using fMRI samples pooled from all subjects, a linear support vector machine (SVM) classifier is trained to predict 2-D stimuli-related images or 3-D stimuli-related images. The experimental results demonstrated the effectiveness of the proposed method.

Keywords: Multi-subject brain decoding, fMRI, hierarchical model, multi-task feature selection, classification

1. Introduction

In the last decade, brain decoding has become a common approach employed in fMRI data analysis [1]. Generally, brain decoding aims at predicting whether a cognitive, perceptual or behavioral stimulus is related to a sample of fMRI data, and the prediction can be used to further investigate the basic relationship between brain areas and stimuli, i.e. higher prediction accuracy usually indicates greater relevance with the brain areas. Typically, fMRI data is represented as volumes of four-dimensional samples, in which each sample corresponds to a category. Therefore, supervised machine learning can be used for brain decoding.

Multi-subject brain decoding is of great interest to researchers studying common neural characteristics among different human brains, and a robust decoding framework for all subjects is desirable. Conventional decoding methods usually treat brain voxels as features. However, the variability of activation patterns across brains at the voxel level makes it difficult to build a satisfactory decoder using samples pooled from different subjects [2]. Thus, decoders are usually built for individual subjects separately. Moreover, comparing with the high feature dimensionality, the number of fMRI samples is very limited. Thus, feature selection plays an important role in fMRI decoding. A recently developed

*Corresponding authors: Tianxin Gao, School of Life Science, Beijing Institute of Technology, Beijing 100081, China. Tel.: 86-10-68912154; Fax: 86-10-68915956; E-mail: gtx@bit.edu.cn.

Yong Xu, Chinese PLA General Hospital (301 Hospital), Beijing 100853, China. Tel.: 86-10-55499109; Fax: 86-10-68247198; E-mail: xuyong@301.zh.mil.

approach of multi-task feature selection, which utilizes complementary information among tasks, has shown to be a powerful feature selection model for multi-subject analysis [3]. However, the conventional multi-task feature selection method, group lasso, assumes that merely shared features exist among tasks, which often contradicts the actual situation that particular features for each task may also exist.

In this study, a novel brain decoding framework was designed to predict 2-D stimuli-related images or 3-D stimuli-related images. First, a hierarchical model was built for the extraction of features from fMRI samples [4,5]. Specifically, BOLD responses of brain voxels in each region of interest (ROI) were used as low-layer features to train an SVM classifier for the corresponding ROI [6,7]. The outputs of the trained SVMs are then used as high-layer features, thus reducing the variability of activation patterns across different brains. Although the number of high-layer features is considerably small, it still exceeds the number of training samples. Furthermore, a new multi-task feature selection method that is more flexible compared to existing methods was proposed. It is worth noting that, here, a task is interpreted as feature selection for one subject; thus, multi-task feature selection means feature selection tasks across multiple subjects are grouped together. The advantage of the proposed feature selection method is that it considers both shared features across tasks and particular features of each individual task. Once feature selection was performed on the high-layer features obtained from the ROI-specific SVMs, a final SVM classifier was trained with samples pooled from all subjects.

The contribution of this paper is two-fold: (1) A hierarchical model is proposed to extract robust features across brains; and (2) A novel multi-task feature selection method is developed to more effectively select high-layer features, enabling the establishment of a more powerful multi-subject decoding framework.

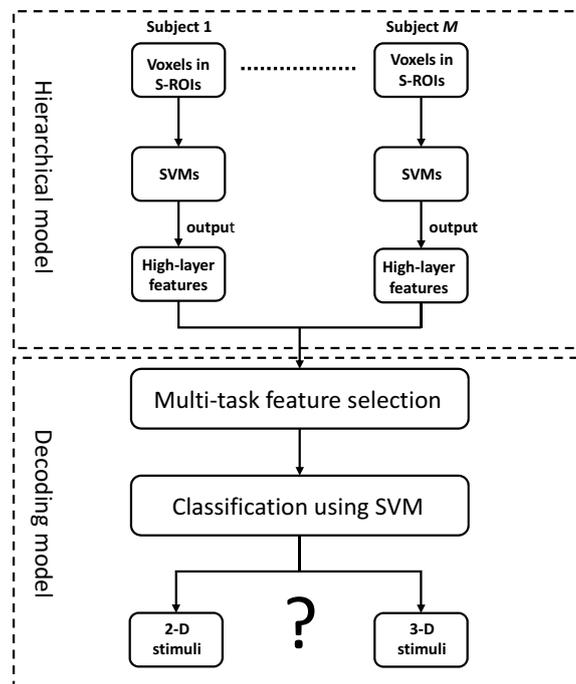


Fig. 1. Flowchart of the proposed framework. S-ROIs denotes the term 'selected ROIs'.

The remainder of this article is organized as follows. Section 2 introduces the proposed framework. The experiment and corresponding settings are described in Section 3, which also provides a validation of the proposed method. Section 4 briefly concludes the study.

2. Proposed method

The proposed framework is mainly composed of two parts, a hierarchical model and a decoding model, which are illustrated as a flowchart in Figure 1. The hierarchical model extracts high-layer features using voxels in selected ROIs, and its output is then used by the decoding model for multi-task feature selection and SVM classification. During multi-task feature selection, relevant high-layer features for each subject are identified. Then, an SVM classifier performs prediction with the selected high-layer features as input.

2.1. Select candidate ROIs

The proposed method parcels the whole brain, which consists of over 300,000 individual voxels, into many ROIs of the same dimension. In this study, different scales were used for the parcellation, including $2 \times 2 \times 2$, $4 \times 4 \times 4$, $6 \times 6 \times 6$, $8 \times 8 \times 8$, and $10 \times 10 \times 10$ voxels, with a less than 50% overlapping between ROIs. First, a T-value corresponding to neural activation was computed for each individual voxel using SPM5 (<http://www.fil.ion.ucl.ac.uk/spm/>), a free toolbox for neuroimage analysis. After assessing each voxel, T-maps can be drawn for each subject independently, as shown in Figure 2. Ordinarily, voxels with high T-values are strongly related to the stimuli; thus, T-values can be used to evaluate whether a region of the brain is activated during the given activity. T-values based on 3-D stimuli were used to select ROIs, because when a subject is presented with the same scene, the brain areas activated by 3-D stimuli usually subsume those activated by 2-D stimuli. To measure the activation within an ROI, the T-mean for the ROI, the averaged T-values for all subjects and all voxels in that ROI, was used. The ROIs whose T-means exceed a predefined threshold were selected as candidate ROIs. Figure 3 shows the variation of the number of selected ROIs (candidate ROIs) with the scale of ROIs.

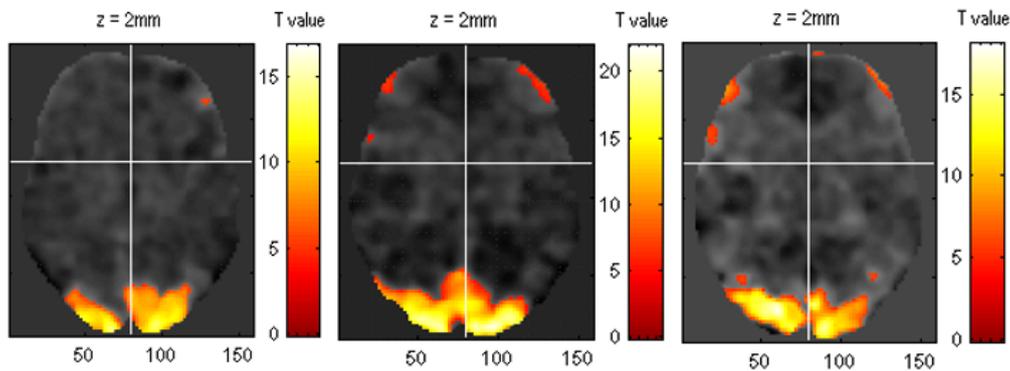


Fig. 2. T-maps of 3-D visual stimuli for each subject.

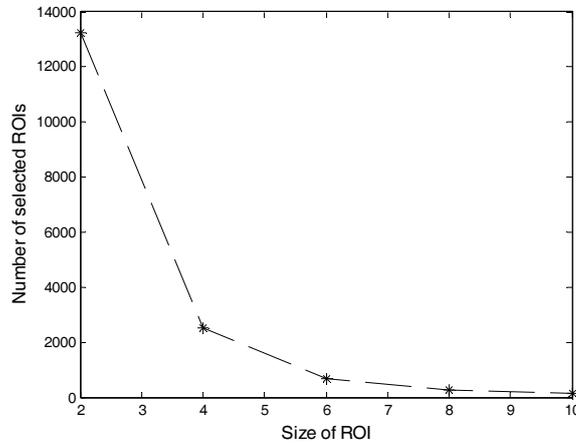


Fig. 3. Number of selected ROIs at different scales.

2.2. Extract high-layer features

In the proposed decoding scheme, a snapshot of BOLD activity is recorded at each time point during the completion of a given task, such as the viewing of a 2-D or 3-D image. The BOLD response of each individual voxel is used as a low-layer feature, and by concatenating all features within a selected ROI, a low-layer feature vector for the ROI is obtained. Then, for each subject, a linear SVM is trained for each selected ROI using the fMRI samples from that subject. Specifically, if the number of selected ROIs is D , a total of D SVMs will be trained for each subject. Signed distance of a low-layer feature vector \mathbf{x} from the corresponding SVM decision boundary is calculated as $d = \frac{\mathbf{w}^T \mathbf{x} + b}{\|\mathbf{w}\|_2}$ and subsequently used as a high-layer feature [5], where \mathbf{w} denotes a weight coefficient vector determined by SVM, $\|\cdot\|_2$ is L_2 norm, superscript T denotes a vector/matrix transpose and b is a bias term in the linear model. Thus, each sample can be represented by a D -dimensional high-layer feature vector through the concatenation of the signed distances of all selected ROIs. Throughout the paper, matrices, vectors and scalars are denoted by boldface uppercase, boldface lowercase and normal italic letters, respectively.

2.3. Proposed multi-task feature selection

Considering the small amount of samples and the high dimensionality of feature vectors, dimension reduction is a common task in neuroimaging-based pattern analysis. Among various dimension-reducing methods, in this study, focus was put on using the sparse least squared regression method, which has been successfully applied to a variety of applications [8]. For a binary classification (where 2-D stimuli and 3-D stimuli are considered as two categories), labels of categories $\mathbf{y}^{(i)} \in \{+1, -1\}$ for the i -th subject can be represented by a linear combination of high-layer features as follows:

$$\mathbf{y}^{(i)} = \mathbf{X}^{(i)T} \mathbf{a}^{(i)} + \mathbf{e}^{(i)}, \quad (1)$$

across subjects while allowing specific selections to be tailored on a subject-level basis. Another benefit of using the proposed method is that the fourth term of Eq. (3) encourages pair-wise highly correlated features to be selected [10], which conforms to the fact that brain areas are usually highly correlated to each other. To find the optimal solution to Eq. (3), an accelerated proximal gradient method was employed [11,12]. Finally, a simple approach was used to unify the structure of feature vectors for all subjects, i.e. only features with a greater-than 50% probability to be selected across all subjects were preserved.

2.4. SVM classifier

Let $\mathbf{x}^{(i)}$ be the feature vector of the i -th training sample and $y^{(i)} \in \{+1, -1\}$ be the corresponding class label, the primary optimization problem of SVM is given as:

$$\begin{aligned} \min_{\mathbf{w}, b, \xi_i} & \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^N \xi_i, \\ \text{s. t. } & y^{(i)} (\mathbf{w}^T \phi(\mathbf{x}^{(i)}) + b) \geq 1 - \xi_i, \\ & \xi_i \geq 0, i = 1, \dots, N, \end{aligned} \quad (4)$$

where ξ_i is the non-negative slack variable, C is the penalty parameter, ϕ is the kernel induced mapping function, and b is a bias. For a given test sample \mathbf{x} , the decision function of SVM for the predicted label is defined as

$$F(\mathbf{x}) = \text{sign}(\sum_{i=1}^N y^{(i)} \alpha^{(i)} k(\mathbf{x}^{(i)}, \mathbf{x}) + b), \quad (5)$$

where $\alpha^{(i)}$ is the Lagrange multiplier and $k(\mathbf{x}^{(i)}, \mathbf{x}) = \phi^T(\mathbf{x}^{(i)})\phi(\mathbf{x})$ is the kernel function for $\mathbf{x}^{(i)}$ and \mathbf{x} .

3. Experimental results

3.1. Materials and experimental settings

Using a 3.0T Philips scanner (TR=2 s, TE=30 ms, FA=90 degrees, FOV=230.4×230.4 mm, 34 slices, voxel size of 2.88×2.88×3.5 mm), 210 EPI images of the whole brain were acquired from three healthy participants, from each 70 images were collected. Each image was labeled by a 2-D or 3-D visual stimulus of the same scene. For each subject, the 70 collected images include 35 2-D stimuli labeled images and 35 3-D stimuli labeled images. All images were preprocessed, i.e. realigned, normalized, and smoothed, using SPM5.

To evaluate the performance of the proposed decoding framework, a 7-fold cross-validation approach was employed. To validate the superiority of the proposed feature selection method, group lasso was used for comparison. Two additional experiments were performed. In one of them, all high-layer features were used to train an SVM classifier without feature selection, denoted as H-SVM, while in the other one, low-layer features corresponding to the voxels whose T-values exceeded the

Table 1

The classification accuracy (%) of three competing methods at different ROI scales (voxel)

| Size of ROI | $2 \times 2 \times 2$ | $4 \times 4 \times 4$ | $6 \times 6 \times 6$ | $8 \times 8 \times 8$ | $10 \times 10 \times 10$ |
|------------------|-----------------------|-----------------------|-----------------------|-----------------------|--------------------------|
| Proposed method | 84.1 ± 2.6 | 87.4 ± 1.8 | 88.5 ± 3.3 | 89.4 ± 2.5 | 86.5 ± 2.3 |
| Group lasso [11] | 81.5 ± 3.7 | 84.8 ± 1.7 | 80.7 ± 2.8 | 80.1 ± 2.3 | 79.3 ± 3.2 |
| H-SVM [7] | 77.8 ± 1.9 | 77.4 ± 4.6 | 77.0 ± 6.1 | 77.8 ± 4.9 | 76.3 ± 5.7 |

predefined threshold were used to train an SVM classifier for decoding, denoted as L-SVM. It should be noted that the same hierarchical models and linear SVM classifiers were used in all experiments¹, and SVM classifiers of all competing methods were trained using the fMRI samples pooled from all subjects.

3.2. Experimental results and evaluation

The proposed method, and two other competing methods were evaluated by their classification accuracies. Table 1 shows the mean classification accuracies of the three competing methods using features extracted with the proposed hierarchical model, with the standard deviations among the 7 folds also reported. It can be observed from Table 1 that the proposed method outperformed the other competing methods at all ROI scales and reached the best performance (89.4%) at the ROI dimension of $8 \times 8 \times 8$ voxels. In another experiment, L-SVM reached an average accuracy of 73.4%, with a standard deviation of 3.9%, a lower performance compared to all other competing methods.

4. Conclusion

In this paper, a novel framework for multi-subject brain decoding was presented. Specifically, a hierarchical model was defined to address the problem that it is difficult to extract robust features across brains. This was validated by comparing the results of H-SVM and L-SVM. Furthermore, a novel multi-task feature selection method was proposed. The method selects features more flexibly by utilizing both shared information among all subjects and particular information from each subject. The experimental results validated the superiority of the proposed framework over all other competing ones. That is to say, the proposed framework provides a more powerful tool for studying the basic relationship between brain areas and external stimuli.

Acknowledgement

The work was supported in part by the National Natural Science Foundation of China (Nos. 81271568 and 81371647). Liye Wang was also supported by the China Scholarship Council (No. 201306030036).

¹The hierarchical model was not used when evaluating L-SVM.

References

- [1] Y. Kamitani and F. Tong, Decoding the visual and subjective contents of the human brain, *Nature Neuroscience* **8** (2005), 679–685.
- [2] J.V. Haxby, Multivariate pattern analysis of fMRI: The early beginnings, *NeuroImage* **62** (2012), 852–855.
- [3] A. Evgeniou and M. Pontil, Multi-task feature learning, *Advances in Neural Information Processing Systems* **19** (2007), 41.
- [4] Y. Xu, D. Zhang and J. Y. Yang, A feature extraction method for use with bimodal biometrics, *Pattern recognition* **43**(2010), 1106-1115.
- [5] M. Liu, D. Zhang and D. Shen, Hierarchical fusion of features and classifier decisions for Alzheimer's disease diagnosis, *Human Brain Mapping* **35** (2013), 1305–1319.
- [6] T. B. Meier et al., Support vector machine classification and characterization of age-related reorganization of functional brain networks, *NeuroImage* **60**(2012), 601-613.
- [7] C. Cortes and V. Vapnik, Support-vector networks, *Machine Learning* **20** (1995), 273–297.
- [8] D. Zhang and D. Shen, Multi-modal multi-task learning for joint prediction of multiple regression and classification variables in alzheimer's disease, *NeuroImage* **59** (2012), 895–907.
- [9] A. Jalali et al., A dirty model for multi-task learning, *Advances in Neural Information Processing Systems* **3** (2010), 964–972.
- [10] H. Zou and T. Hastie, Regularization and variable selection via the elastic net, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **67** (2005), 301–320.
- [11] J. Liu and J. Ye, Efficient $l_{1/lq}$ norm regularization, arXiv preprint arXiv:1009.4766, 2010.
- [12] X. Chen et al., Accelerated gradient method for multi-task sparse learning problem, *Ninth IEEE International Conference on Data Mining (ICDM)*, 2009, 746–751.