

## 基于稀疏表示的帕金森功能连接模式定位

陈勇斌, 李远清<sup>†</sup>

(华南理工大学 自动化科学与工程学院, 广东 广州 510641)

**摘要:** 在脑成像数据分析中, 基于稀疏表示的模式定位算法在群组水平分析中具有非常优秀的性能, 但在单个数据集的情况下结果还不尽如人意. 为此, 文中在先前研究的基础上提出了一种改进算法, 通过基于原始数据集生成多个派生数据集的方法, 来改善算法在单个数据集分析中的不足. 仿真结果表明改进后算法在性能上有显著的提高. 文章随后将该改进算法应用于帕金森病异常功能连接模式定位分析之中, 得到广泛分布于全脑的与该疾病相关的 269 个异常功能连接, 由此对算法的有效性进行了验证, 并可能有助于加强对与该疾病相关的病理生理机制的了解.

**关键词:** 模式定位; 稀疏表示; 多变量模式分析; 功能连接

**中图分类号:** TP391.4      **文献标识码:** A

## Pattern localization of functional connectivity in Parkinson's disease based on sparse representation

CHEN Yong-bin, LI Yuan-qing<sup>†</sup>

(School of Automation Science and Engineering, South China University of Technology, Guangzhou Guangdong 510641, China)

**Abstract:** In the analysis of brain imaging data, the sparse representation-based pattern localization algorithm has a very good performance at the group level data analysis. But at the single level, it's performance is still disappointed. Therefore, in order to compensate for this deficiency, an improved algorithm based on previous research was proposed in this study. By generating multiple derived data sets from the original data and then performing pattern localization procedure, the improved algorithm has better performance compared to the original in simulation. Subsequently, the improved algorithm was applied to the analysis of localizing all abnormal brain functional connections in Parkinson's disease. 269 abnormal connections were obtained and they were widely distributed throughout the entire brain. Thus, the effectiveness of the algorithm was verified and our findings may have the potential to advance the understanding of the neural mechanism of this disease.

**Key words:** pattern localization; sparse representation; multivariate pattern analysis; functional connectivity

### 1 Introduction

Various brain imaging techniques have become available in the recent decades. These techniques make it possible to observe brain structure or brain functional activity noninvasively. As powerful tools to explore and reveal the mechanism of brain activity, they greatly promote the progress of brain science research, and enrich our understanding of the complex structure and function of the brain<sup>[1]</sup>.

In the analysis of brain imaging data, pattern localization has attracted considerable attention. For example, a great deal of researches have been devoted to localization of brain activation induced by different stimuli or localization of abnormal areas of structure in patients with neurological or psychiatric disease based on magnetic resonance imaging (MRI). Traditionally, brain imaging data analysis mainly adopts univariate analy-

sis methods such as general linear model (GLM)<sup>[2-3]</sup>, voxel-based morphometry (VBM)<sup>[4]</sup>. Although these conventional methods have the ability of detecting statistical group differences by analyzing each position of the brain separately and have been tremendously productive, they might ignore potential source information or the spatial correlation in the data, such as the inter-relationship between different brain areas<sup>[5]</sup>. Therefore, they can't provide higher sensitivity and may not obtain satisfactory results.

In recent years, multivariate pattern analysis (MVPA) approaches derived from pattern recognition techniques have been widely implicated in brain imaging data analysis<sup>[6-7]</sup>. Compared with the conventional univariate methods, MVPA presents a higher sensitivity, because MVPA can take into account the pattern of information which may be shown across multiple vari-

Received 22 August 2016; accepted 30 March 2017.

<sup>†</sup>Corresponding author. E-mail: auyqli@scut.edu.cn; Tel.: +86 18620883682.

Recommended by Associate Editor: YU Zhu-liang.

Supported by National Natural Science Foundation of China (91120305) and Guangdong Natural Science Foundation (2014A030312005).

ables<sup>[8-9]</sup>. Several studies have shown that MVPA approaches are capable of extracting stable structural or functional characteristic patterns from brain imaging data<sup>[10-13]</sup>.

However, in most MVPA approaches, not all the informative features are selected because part of them may be sufficient for decoding or classification. In order to achieve the highest classification accuracy, only those features containing great discriminative ability can be extracted. So the rest of the features containing weak discriminative information may be ignored. In many real applications, we want to extract all of the informative features contributing to the discrimination, no matter how small a contribution they provide. For this purpose, Li and his colleagues<sup>[14]</sup> proposed a sparse representation-based pattern localization algorithm and have successfully applied it in localizing brain activation patterns corresponding to different stimulus classes respectively in functional MRI (fMRI) data analysis. In this pattern localization algorithm, feature selection can be modeled as a sparse representation problem<sup>[15]</sup>. And the selected features can be separated into two sets according to the signs of the sparse decomposition weights, corresponding to two brain states, which was demonstrated by data analysis and mathematically proven based on several simplified models in paper [14]. Compared with univariate analysis methods, this algorithm has better performance in localizing all the informative features at the group level (pattern localization across multiple data sets) data analysis, but at the single level (pattern localization based on only one data set), it's performance is still disappointed.

Parkinson's disease (PD) is a common progressive neurodegenerative disorder that manifests principally as resting tremor, rigidity, akinesia and postural instability in people over age 50. Recently, there has been growing interest in exploring the functional connectivity in PD based on fMRI data. However, these studies mainly adopted univariate analysis methods, there has been little focus on examining the functional connectivity<sup>[16]</sup> based on MVPA. MVPA methods in functional connectivity analysis have been used to extract group differences related to psychiatric disorders such as schizophrenia<sup>[17]</sup> and major depression<sup>[18]</sup>. In this present study, we proposed an improved sparse representation-based pattern localization algorithm based on previous study<sup>[14]</sup> to improve the performance at the single level data analysis and employed this improved algorithm to localize all abnormal brain functional connections in PD. Locating all of the abnormal functional connections has the potential to advance the understanding of the neural mechanism of this disease. The results show that the improved algorithm has better performance and functional connectivity abnormalities in PD are widely distributed throughout the entire brain.

### 2 Pattern localization algorithm

The procedure of the sparse representation-based pattern localization algorithm is illustrated in Fig.1. Feature selection was based on the weights of all features, which were determined by sparse representation. We first explain the method for weight determination, and then describe the pattern localization procedure step by step.

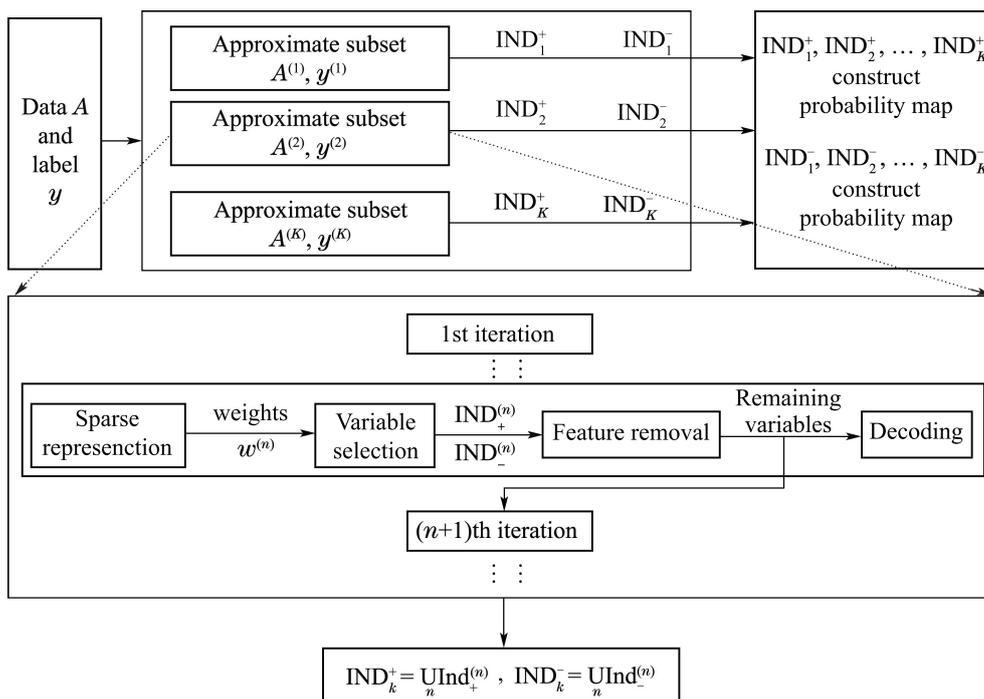


Fig. 1 The algorithm diagram for pattern localization

## 2.1 Feature weight determination

The data were given by a matrix  $\mathbf{A} \in \mathbb{R}^{M \times K_0}$ , where the  $M$  rows and the  $K_0$  columns corresponded to the samples and the features respectively. The column vector  $\mathbf{y} \in \mathbb{R}^M$  was a label vector with 1 representing one class (e.g. patients) and  $-1$  representing another class (e.g. controls). To obtain a weight vector  $\mathbf{w}$  of features, we solved the following optimization problem:

$$\min \|\mathbf{w}\|_1, \text{ s.t. } \mathbf{A}\mathbf{w} = \mathbf{y}, \quad (1)$$

model (1) can be seen as a sparse regression between the data matrix  $\mathbf{A}$  and the label vector  $\mathbf{y}$ . The absolute value of each entry of the optimal solution (denoted as  $\mathbf{w}_0$ ) reflects the contribution of its corresponding feature to the discrimination between two classes.

Setting  $\mathbf{w} = \mathbf{u} - \mathbf{v}$ , where  $\mathbf{u}, \mathbf{v} \in \mathbb{R}^{K_0}$  are non-negative, the optimization problem (1) can be converted to the following equivalent linear programming problem:

$$\begin{aligned} \min \quad & \sum_{i=1}^{K_0} (u_i + v_i), \\ \text{s.t.} \quad & \begin{cases} [\mathbf{A} & -\mathbf{A}][\mathbf{u}^T & \mathbf{v}^T]^T = \mathbf{y}, \\ \mathbf{u}, \mathbf{v} \geq 0, \end{cases} \end{aligned} \quad (2)$$

The solution of (2) can be obtained easily using the MATLAB function "linprog".

## 2.2 Pattern localization procedure

The pattern localization algorithm includes a few steps. In the following, we begin to describe the complete and detailed procedure.

### 2.2.1 Generating multiple approximate subsets

Because the features extracted from brain imaging data included the generalization of "noise", a single pattern localization performed on the total data may pick up irrelevant features. In order to remove these irrelevant features, we generated  $K$  approximate but different subsets from original data and performed pattern localization on each of them. In this way, we can construct two probability maps and test them with a permutation test (as shown later). The data matrix  $\mathbf{A}$  and the column vector  $\mathbf{y}$  were equally partitioned into  $K$  non-overlapping parts according to their rows. Each time, we removed one part and used the rest ( $K-1$ ) parts (denoted as  $\mathbf{A}^{(k)}$  and  $\mathbf{y}^{(k)}$ ,  $k = 1, \dots, K$ ) as an approximate subset and its corresponding label vector. Then we obtained  $K$  pairs of approximate subsets and their corresponding label vectors originated from data matrix  $\mathbf{A}$  and the label vector  $\mathbf{y}$ . We suggest the parameter  $K$  is set larger than 20 since it is related to the calculation of probability maps.

### 2.2.2 Recursive iterative feature search

For each pair of approximate subset  $\mathbf{A}^{(k)}$  and its corresponding label vector  $\mathbf{y}^{(k)}$  ( $k = 1, \dots, K$ ), a recursive iterative feature elimination method relying on the weights obtained by sparse representation was used to pick up as many informative features as possible. In the  $n$ th iteration ( $n = 1, 2, \dots$ ), we performed the following four steps:

**Step 1** Feature weight determination. We applied the sparse representation method mentioned above to the data updated in the previous iteration and obtained a weight vector (denoted as  $\mathbf{w}^{(n)}$ ) of features. In the first iteration, the data to be used were matrix  $\mathbf{A}^{(k)}$ .

**Step 2** Feature selection. We then selected  $N_0$  features (denoted as  $\text{Ind}_+^{(n)}$ ) corresponded to the  $N_0$  largest positive elements and  $N_0$  features (denoted as  $\text{Ind}_-^{(n)}$ ) corresponded to the  $N_0$  smallest negative elements of the weight vector  $\mathbf{w}^{(n)}$ , respectively.

**Step 3** Feature removal. We removed these features in  $\text{Ind}_+^{(n)}$  and  $\text{Ind}_-^{(n)}$  from the data matrix used in the current iteration, and obtained an updated data matrix composed by the remaining features which would be used in the next iteration.

**Step 4** Decoding. We performed a decoding by applying an SVM classifier to the updated data matrix. The prediction accuracy of the labels (denoted as  $r_n$ ) was calculated based on a cross-validation classification procedure. If  $r_n > \alpha$ , where  $\alpha$  is a predefined threshold, go to Step 1. Otherwise terminate above iteration procedure. Theoretically, the best threshold was 50% for the two-class problem, which was adopted in this paper.

After iteration procedure terminated, we obtained two feature sets corresponding to two classes respectively according to the signs of their weights

$$\text{IND}_k^+ = \bigcup_n \text{Ind}_+^{(n)}, \quad \text{IND}_k^- = \bigcup_n \text{Ind}_-^{(n)}. \quad (3)$$

**Remark 1** In the above Step 4, the decoding was based on the updated data matrix composed by the remaining features after feature removal in the  $n$ th iteration.

### 2.2.3 Probability map calculation

As described above, two sets of selected features were obtained using each approximate subset. There were  $K$  approximate subsets in total, so we would finally have  $K$  sets  $\text{IND}_k^+$  and  $K$  sets  $\text{IND}_k^-$  of selected features. Then we constructed a probability map only using the  $K$  sets  $\text{IND}_k^+$ . The probability value of each feature was calculated by counting the number of appearances of this feature in all of the  $K$  sets  $\text{IND}_k^+$  and dividing this number by the total number of se-

lected features in those sets. Similarly, only using the  $K$  sets  $\text{IND}_k^-$ , another probability map was constructed. Thus, two probability maps corresponding to two classes respectively were obtained.

### 2.2.4 Permutation test

In this study, we used a non-parametric permutation test to remove the irrelevant features. In each permutation, we randomly permuted the labels of the data matrix  $\mathbf{A}$  and repeated the above procedure. Each time we obtained two probability maps corresponding to two classes respectively. We performed 100 permutations and obtained 100 probability maps for each class. A null distribution for each class was constructed by pooling all probability values of the 100 probability maps corresponding to this class. Given a significance level  $\alpha_1$ , we found two critical thresholds  $\theta_1$  and  $\theta_2$  corresponding to the percentile  $100(1 - \alpha_1)\%$  of the two null distributions. Then, we applied the two thresholds  $\theta_1$  and  $\theta_2$  to the two probability maps obtained using real labels respectively and obtained two sets of informative features.

In order to improve the performance at the single level data analysis, we still need to present an iteration algorithm for generating multiple derived data sets from the original data matrix  $\mathbf{A}$ . In each iteration, we randomly chose  $L$  rows from matrix  $\mathbf{A}$  to construct a  $L \times K_0$  matrix denoted as  $\mathbf{A}_i$ , and the corresponding  $L$  entries of label vector  $\mathbf{y}$  formed a column vector denoted as  $\mathbf{y}_i \in \mathbb{R}^L$ . After  $N$  iterations, we would get  $N$  pairs of derived data sets  $(\mathbf{A}_1, \mathbf{A}_2, \dots, \mathbf{A}_N)$  and their corresponding label vectors  $(\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_N)$ . For each derived data set  $\mathbf{A}_i$  ( $i = 1, \dots, N$ ), we performed above procedure and obtained two probability maps corresponding to two classes respectively. Then we averaged all probability maps of each class across all derived data sets and obtained two average probability maps also corresponding to two classes respectively. Next, the two average probability maps were tested with a permutation test. In each permutation, we also obtained two average probability maps across all derived data sets. Finally, we obtained two sets of informative features in a similar way as described above based on these average probability maps.

## 3 Simulation experiment

Because the patterns can not be obtained accurately using real brain imaging data due to the influence of noise, so we first design simulation data (preset patterns) and use it to compare the performance of different algorithms.

### 3.1 Experimental design

First, we generated two fixed pattern vectors  $\mathbf{p}_1, \mathbf{p}_2 \in \mathbb{R}^{100}$ , each containing 10 entries took value 1 and 90 entries took value 0 with their positions ran-

domly assigned. Furthermore, the index sets of value 1 of the two patterns were non-overlapped as illustrated in Fig.2.  $\mathbf{P} \in \mathbb{R}^{30 \times 100}$  was a pattern matrix containing 15 rows of pattern  $\mathbf{p}_1$  and 15 rows of pattern  $\mathbf{p}_2$ .  $\mathbf{V} = [v_{jk}] \in \mathbb{R}^{30 \times 100}$  was a noise matrix, of which each column obeys normal distribution, with mean zero and standard deviation  $\sigma$ . We can control the magnitude of the noise via changing the value of parameter  $\sigma$ . The matrix  $\mathbf{P} + \mathbf{V}$  was used to simulate a real data matrix containing noise. The 15 rows of  $\mathbf{P} + \mathbf{V}$  containing the pattern  $\mathbf{p}_1$  were labeled as 1, and the other 15 rows containing the pattern  $\mathbf{p}_2$  were labeled as  $-1$ . Considering the following optimization problem:

$$\min \|\mathbf{w}\|_1, \text{ s.t. } (\mathbf{P} + \mathbf{V})\mathbf{w} = \mathbf{y}. \quad (4)$$

We performed our improved pattern localization algorithm to localize the two patterns  $\mathbf{p}_1$  and  $\mathbf{p}_2$ . For comparison, we also used the original algorithm to localize the two patterns.

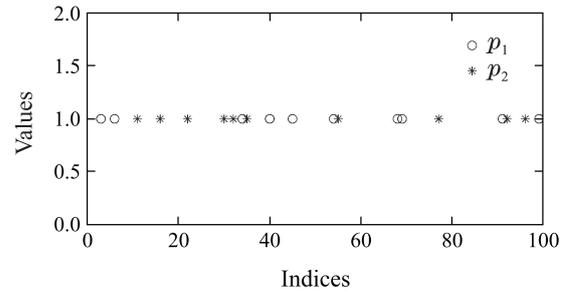


Fig. 2 Two patterns of simulation data

### 3.2 Simulation results

In order to compare the performance of different methods completely, we used a series of noise magnitudes ( $\sigma = 0.1, 0.2, \dots, 1.0$ ) to generate simulation data and localized the two patterns  $\mathbf{p}_1$  and  $\mathbf{p}_2$ . Setting the significance level  $\alpha_1 = 0.001$ , iteration number  $N = 5$ , row number  $L = 20$ , parameter  $K = 20$ , the results averaged across 5 simulations were shown in Fig.3. Compared with the original algorithm, our improved algorithm proposed in this paper has better performance under various noise levels.

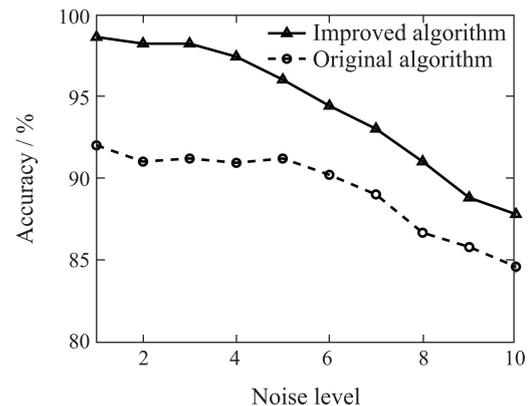


Fig. 3 Localization accuracy comparison

## 4 Pattern localization on real data

In this section, we employed the improved pattern localization algorithm to localize all abnormal brain functional connections in PD.

### 4.1 Participants

The participants included 21 PD patients (10 males and 11 females; mean age, 58.3 years) and 26 healthy controls (10 males and 16 females; mean age 61.3 years). The age and sex differences between the two groups were tested by two-sample  $t$  test and  $\chi^2$  test, respectively. No significances were shown between the two groups ( $p > 0.05$ ).

### 4.2 Data acquisition and preprocessing

All the scans were performed on a GE Signal Excite HD 3.0-T MR scanner equipped with a standard 8-channel head coil in Guangdong General Hospital, China. Foam padding and earplugs were used to limit head movement and reduce scanner noise. Whole-brain coverage resting-state functional images were acquired by using a gradient-echo echo-planar T2\*-weighted imaging (EPI) sequence with the following parameters: 30 slices (in ascending noninterleaved order, parallel to the anterior commissure (AC) - posterior commissure (PC) line), TR = 2000 ms, TE = 30 ms, flip angle =  $80^\circ$ , FOV =  $240 \times 240$  mm, matrix size =  $64 \times 64$ , and slice thickness = 5.0 mm (no gap). During the data acquisition, subjects were instructed to keep their eyes closed and to move as little as possible. The scan time for each participant was 372 s, and 186 volumes were obtained.

We discarded the first 5 volumes of the scanned data for the instability of the initial MRI signal and the participants' adaptation to the circumstance. The following preprocessing consisted of slice timing correction, head motion correction, normalization to an MNI standard brain, detrending, temporal filtering, nuisance covariates regressing out. All fMRI data were preprocessed using the statistical parametric mapping (SPM8) and data processing assistant for resting-state fMRI (DPARSF) programs.

### 4.3 Feature extraction

For each subject, the volumes were partitioned into 116 ROIs according to the automated anatomical labelling atlas<sup>[19]</sup>. Mean time series for each region were extracted by averaging the time series within this region. Then we evaluated the functional connectivity between each pair of regions using the Pearson correlation coefficient and obtained  $(116 \times 115)/2 = 6670$  resting-state functional connections as the features for pattern Localization.

## 4.4 Results

First, we generated 5 derived data sets with the parameter  $L$  set as 40. Then we generated 40 approximate subsets from each derived data set. Given significance level  $\alpha_1 = 0.001$ , we finally obtained 269 abnormal functional connections in PD. To present the results in a clear and concise manner, abnormal functional connections and region weights are displayed in a circle graph (Fig. 4) using a MATLAB tool developed by ourselves. A solid or dashed line indicates that the corresponding functional connectivity exhibited a increase or decrease, respectively, in the patients compared with the healthy controls. Regions are size-coded by weight which was denoted by its occurrence number in the abnormal functional connections. In this investigation, 46% of the abnormal functional connections decreased in the patients compared with the healthy controls. The abnormal functional connections were widely distributed throughout the entire brain and primarily located within or across the default mode, cingulo-opercular and frontal-parietal networks and the cerebellum. For comparison, we also used the original algorithm to localize all abnormal brain functional connections. Given the same significance level  $\alpha_1 = 0.001$  as above, only 127 abnormal connections were extracted. Therefore, the improved algorithm has better sensitivity and can detect more subtle informative features.

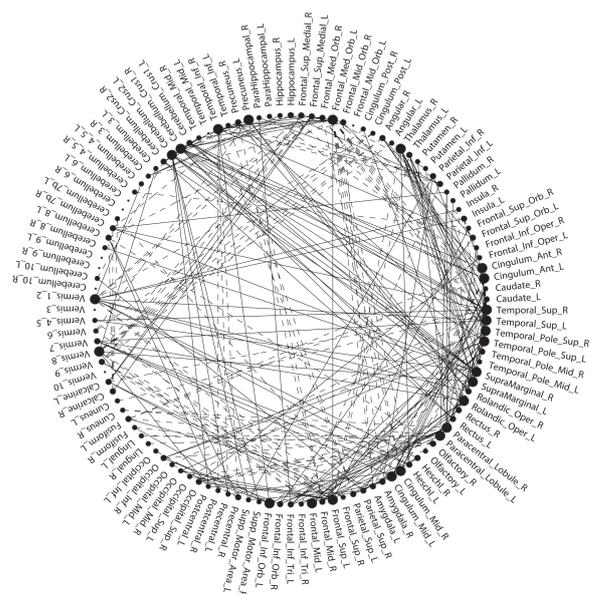


Fig. 4 Region weights and the distribution of the 269 abnormal functional connections in PD

## 5 Conclusions

In this study, we proposed an improved sparse representation-based pattern localization algorithm

based on previous study and employed it to localize all abnormal brain functional connections in PD. 269 abnormal functional connections were obtained, and they were widely distributed throughout the entire brain. These disease-related functional connections might play important roles in the pathophysiology of this disease. Our results have the potential to advance the understanding of the neural mechanisms of PD. However, although our improved algorithm has better performance, its cost also shows a remarkable increase in computational complexity. Therefore, the next step will be to improve the computational speed of this algorithm.

## References:

- [1] LI E, GAO J, LU G, et al. Functional neuroimaging and its applications to critical brain diseases [J]. *Scientia sinica vitae*, 2015, 45(3): 237 – 246.
- [2] FRISTON K J, JEZZARD P, TURNER R. Analysis of functional MRI time-series [J]. *Human Brain Mapping*, 1994, 1(2): 153 – 171.
- [3] PENNY W D, FRISTON K J, ASHBURNER J T, et al. *Statistical Parametric Mapping: The Analysis of Functional Brain Images* [M]. London, UK: Academic Press, 2011.
- [4] ASHBURNER J, FRISTON K J. Voxel-based morphometry—the methods [J]. *NeuroImage*, 2000, 11(6): 805 – 821.
- [5] MOURAO-MIRANDA J, REYNAUD E, MCGLONE F, et al. The impact of temporal compression and space selection on SVM analysis of single-subject and multi-subject fMRI data [J]. *NeuroImage*, 2006, 33(4): 1055 – 1065.
- [6] O' TOOLE A, JIANG F, ABDI H, et al. Theoretical, statistical, and practical perspectives on pattern-based classification approaches to the analysis of functional neuroimaging data [J]. *Journal of Cognitive Neuroscience*, 2007, 19(11): 1735 – 1752.
- [7] KRIEGESKORTE N. Pattern-information analysis: from stimulus decoding to computational-model testing [J]. *NeuroImage*, 2011, 56(2): 411 – 421.
- [8] YANG Z, FANG F, WENG X. Recent developments in multivariate pattern analysis for functional MRI [J]. *Neuroscience Bulletin*, 2012, 28(4): 399 – 408.
- [9] PEREIRA F, MITCHELL T, BOTVINICK M. Machine learning classifiers and fMRI: a tutorial overview [J]. *NeuroImage*, 2009, 45(1): S199 – S209.
- [10] FAN Y, BATMANGHELICH N, CLARK C M, et al. Spatial patterns of brain atrophy in MCI patients, identified via high-dimensional pattern classification, predict subsequent cognitive decline [J]. *NeuroImage*, 2008, 39(4): 1731 – 1743.
- [11] ZHU C, ZANG Y, CAO Q, et al. Fisher discriminative analysis of resting-state brain function for attention-deficit/hyperactivity disorder [J]. *NeuroImage*, 2008, 40(1): 110 – 120.
- [12] WEYGANDT M, BLECKER C R, SCHÄFER A, et al. fMRI pattern recognition in obsessive-compulsive disorder [J]. *NeuroImage*, 2012, 60(2): 1186 – 1193.
- [13] ECKER C, ROCHA-REGO V, JOHNSTON P, et al. Investigating the predictive value of whole-brain structural MR scans in autism: a pattern classification approach [J]. *NeuroImage*, 2010, 49(1): 44 – 56.
- [14] LI Y, LONG J, HE L, et al. A sparse representation-based algorithm for pattern localization in brain imaging data analysis [J]. *PLoS One*, 2012, 7(12): e50332.
- [15] LI Y, YU Z, BI N, et al. Sparse representation for brain signal processing: a tutorial on methods and applications [J]. *IEEE Signal Processing Magazine*, 2014, 31(3): 96 – 106.
- [16] BARDIN J. Making connections [J]. *Nature*, 2012, 483(7390): 394 – 396.
- [17] SHEN H, WANG L, LIU Y, et al. Discriminative analysis of resting-state functional connectivity patterns of schizophrenia using low dimensional embedding of fMRI [J]. *NeuroImage*, 2010, 49(4): 3110 – 3121.
- [18] ZENG L, SHEN H, LIU L, et al. Identifying major depression using whole-brain functional connectivity: a multivariate pattern analysis [J]. *Brain*, 2012, 135(5): 1498 – 1507.
- [19] SCHMAHMANN J D, DOYON J, MCDONALD D, et al. Three-dimensional MRI atlas of the human cerebellum in proportional stereotaxic space [J]. *NeuroImage*, 1999, 10(3): 233 – 260.

## 作者简介:

**陈勇斌** (1977–), 男, 博士研究生, 目前研究方向为脑信息处理, E-mail: mailcyb@163.com;  
**李远清** (1966–), 男, 博士, 教授, 博士生导师, 目前研究方向为脑机接口与脑信息处理, E-mail: auyqli@scut.edu.cn.