Support Vector Clustering for Brain Activation Detection

Defeng Wang[1], Lin Shi[2], Daniel S. Yeung[1], Pheng-Ann Heng[2], Tien-Tsin Wong[2], and Eric C.C. Tsang[1]

[1] Department of computing, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong, China
{csdfwang, csdaniel, csetsang}@comp.polyu.edu.hk
[2] Department of Computer Science and Engineering, The Chinese University of Hong Kong, Shatin, N.T., Hong Kong, China
{lshi, pheng, ttwong}@cse.cuhk.edu.hk

Abstract. In this paper, we propose a new approach to detect activated time series in functional MRI using *support vector clustering (SVC)*. We extract Fourier coefficients as the features of fMRI time series and cluster these features by SVC. In SVC, these features are mapped from their original *feature space* to a very high dimensional *kernel space*. By finding a compact sphere that encloses the mapped features in the kernel space, one achieves a set of cluster boundaries in the feature space. The SVC is an effective and robust fMRI activation detection method because of its advantages in (1) better discovery of real data structure since there is no cluster shape restriction, (2) high quality detection results without explicitly specifying the number of clusters, (3) the stronger robustness due to the mechanism in outlier elimination. Experimental results on simulated and real fMRI data demonstrate the effectiveness of SVC.

1 Introduction

Functional magnetic resonance imaging (fMRI) is a non-invasive tool to observe the brain neural activities when the subject is undertaking cognitive or motor tasks. Most activation detection techniques can be categorized as *model-driven* and *data-driven* approaches. Apart from principle component analysis (PCA) and independent component analysis (ICA), clustering is a family of effective data-driven approaches to identify unknown responses in fMRI data. The aim of clustering approach is to separate the time series into clusters — each contains voxels with similar activation patterns. Existing clustering approaches to fMRI data analysis include but not limited to crisp C-means [1], Kohonen clustering neural network [2], fuzzy C-means (FCM) [3], hierarchical clustering [1], etc.

However, most existing methods are suffering from the problem of choosing the number of clusters. The detected set of voxels varies significantly when the number of clusters is set to different values. For example, in crisp C-means and FCM, the user has to define a suitable value for the number of clusters C; in hierarchical clustering, there is similar dilemma in determining a proper cut for the
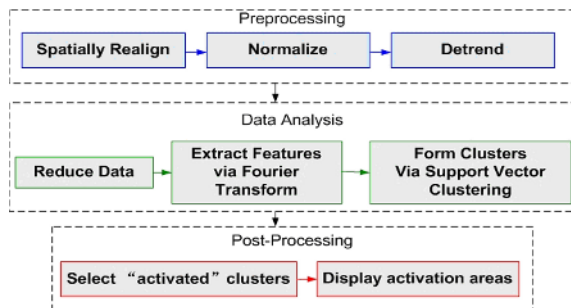
Fig. 1. fMRI Data Processing Scheme

dendrogram. Moreover, the existing fMRI clustering methods implicitly assume the potential clusters are scattered in certain shapes, such as hyper-spherical (when using Euclidean distance) and hyper-ellipsoidal (when using Mahalanobis distance). This oversimplified assumption leads to partitions against the natural data structure and thus achieves results with high false alarm rate.

Support vector machine (SVM) [4] is a preferred classifier in many theoretical and empirical areas, due to its salient properties such as margin maximization and kernel substitution for classifying data in a high dimensional kernel space. Inspired by SVM, the support vector clustering (SVC) algorithm is proposed [5].

In this paper, we explore the application of SVC to solve the fMRI activation detection problem. The whole scheme is illustrated in Fig. 1. We first extract features from fMRI time series using the Fourier transform, and then apply SVC to cluster these Fourier coefficients. They are mapped to a high dimensional kernel space using a kernel function, e.g. the Gaussian kernel. In the kernel space, we calculate the smallest sphere enclosing these mapped features, which appears in the feature space as a set of contours enclosing the original features. These contours are interpreted as cluster boundaries. The points within the same continuous region encompassed by the contour are considered as a cluster. SVC has been shown to be able to generate clusters with arbitrary shapes and eliminate the outlier by using a regularization constant that controls the compactness of the sphere in the kernel space – points out of that sphere is considered as outliers.

2 Materials and Methods

2.1 Dataset

Simulated Dataset Generation. According to a new BOLD response modeling technique [6], the BOLD response $s(t)$ is modeled as the combination of two gamma functions $g_a(t)$ and $g_b(t)$ convoluted with stimulus $c(t)$:

$$g_a(t) = \left(1 - e^{-1/d_a}\right)^2 (t+1)e^{-t/d_a}, \quad g_b(t) = \left(1 - e^{-1/d_b}\right)^2 e^{-t/d_b},$$

$$s(t) = f_a(g_a * c)(t - d_0) + f_b(g_b * c)(t - d_0) + f_c(g_a * c)(t - d_0)(g_b * c)(t - d_0)$$

Table 1. Parameters for the simulated data

Region	f_a	f_b	f_c	d_a	d_b	d_0
A	0.7	0.04	0.6	2	12	3
B (left and right)	0.5	0.3	0.5	3	6	2
C	0.35	0.1	1	5	5	4

To simulate the real-world fMRI data and to get concrete performance measurements for comparison, we construct four artificial activation regions, i.e. A, B(left and right), and C regions in Fig. 2(b), with three sets of values for the parameters (see Table 1) in the above-mentioned BOLD response model. The baseline image is constructed by averaging time courses on the 29th slice in the processed real auditory fMRI dataset (cf. Fig. 2(a)) from Wellcome Department of Cognitive Neurology at University College London. The size of simulated data is $79 \times 95 \times 1$, and the length for each time series is 200. Then we add onto all the voxels the additive white Gaussian noise with intensities proportional to the baseline voxel intensities. In the experiment, we simulate various contrast-to-noise ratio (CNR).

In Vivo **fMRI Dataset.** We also use the auditory fMRI data acquired by Wellcome Department of Cognitive Neurology at University College London to validate the effectiveness of our method. We use it with the permission from the Functional Imaging Laboratory. The experiment was conducted to measure the
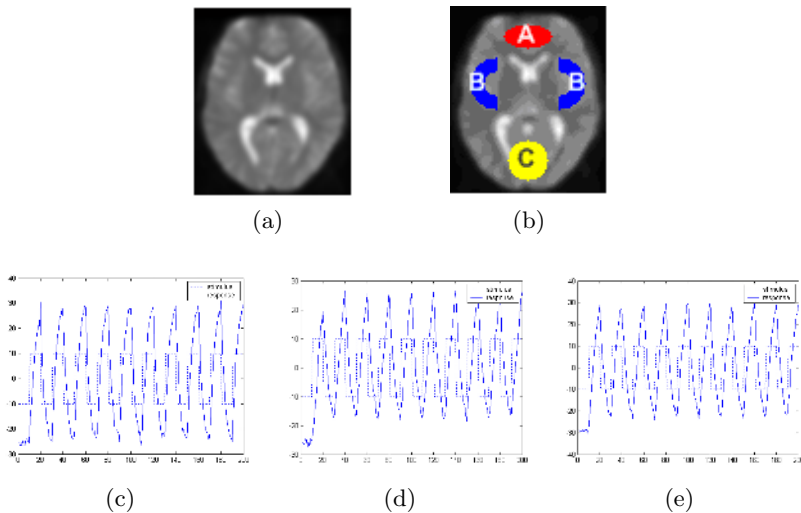

(a)   (b)

(c)   (d)   (e)

Fig. 2. Simulated Dataset (a) baseline slice generated from the 29th slice of real data; (b) spatial layout of simulated BOLD activation; (c) average activation time series in A; (d) average activation time series in B; (e) average activation time series in C

activation in the subjects brain when given an auditory bi-syllabic stimulation. This whole brain BOLD/EPI data was acquired with a modified 2-T SIEMENS scanner and the TR is set to 7 sec. The spatial resolution of the volume data is $64 \times 64 \times 64$, and the size of each voxel is $3mm \times 3mm \times 3mm$. In this experiment, the subject is given a boxcar stimulation: first begins with rest, and then switches between rest (6 scans, last for 42 sec) and stimulation (6 scans, last for 42 sec). By discarding the first few scans, the acquired 96 scans are reduced to 84 scans. The structural dataset that has been aligned with the function data is also provided for reference.

2.2   Methods

Preprocessing. To remove the head motion effect in raw fMRI data, we first spatially realign the slices with the first slice. Then the dataset is normalized by subtracting the mean amplitude from each time series and adjust their variances to one in order to emphasize their variation patterns. The last preprocessing step is first-order polynomial detrending [7] that removes the undesired drift.

Data Analysis. *(1)Data Reduction:* To accelerate feature extraction and data clustering, also to avoid ill-balance clustering result, we use t-test to remove the data points that will definitely not be considered as the activated voxels. All the voxels that are likely to be activated will be acquired by setting a low threshold in the t-test. *(2)Feature Extraction:* BOLD response to the periodic stimulus can be well characterized by the Fourier coefficients [8]. As the paradigm in our experiment is typical box-car, which is periodic, we use the Fourier transform to acquire the features for clustering. Another merit of using Fourier transform to extract features is to avoid explicitly modeling the delay from the stimulation, as the delay changes in various brain locations and for different subjects. Suppose the fMRI dataset is $\mathbf{Y}$, a $T \times N$ matrix, where N is the number of time series, each with a length T. Columns in $\mathbf{Y}$ are time series and rows in $\mathbf{Y}$ correspond to scans. Since we have performed the normalization and detrending in the preprocessing stage, the columns of $\mathbf{Y}$ are zero mean, zero drift time series. The harmonic components are calculated as

$$\alpha_\ell(t) = \sin\left(\frac{\ell}{2}\omega t\right) \cdot (\ell\ mod\ 2 + 1) + \cos\left(\frac{\ell+1}{2}\omega t\right) \cdot (\ell\ mod\ 2),\ \ell = 1 \ldots L, \quad (1)$$

where ω is the fundamental frequency of the experimental paradigm, $L \leq (2\pi)/(\triangle t \cdot \omega)$ is the number of harmonic components, and $\triangle t$ is the temporal sampling period. We then form a $T \times L$ design matrix $\mathbf{A} = [\mathbf{a}_1 \ldots \mathbf{a}_L]$. So far, the data can be expressed as $\mathbf{Y} = \mathbf{A} \cdot \mathbf{X} + \boldsymbol{\varepsilon}$ [9], where $\mathbf{X}$ is an $L \times N$ harmonic image matrix containing the linear coefficients, and $\boldsymbol{\varepsilon}$ is the noise matrix. An estimate of $\mathbf{X}$ can be computed using a least squares fit, i.e. $\hat{\mathbf{X}} = (\mathbf{A}^t\mathbf{A})^{-1}\mathbf{A}^t\mathbf{Y}$. We treat $\hat{\mathbf{X}}$ as the input data in the next SVC step.

Support Vector Clustering. There are two major steps in support vector clustering: a training step to train a kernel radius function and a cluster labeling step to assign a cluster index to each data point.

(1) Training: Assume $\mathbf{X} = \{\mathbf{x}_i\}, i = 1 \ldots N$ is the input data. Using a non-linear transformation, $\mathbf{X}$ can be transformed to a high-dimensional kernel space, where we can find the smallest hyper-sphere that encloses the maps of $\{\mathbf{x}_i\}$, i.e. $\{\phi(\mathbf{x}_i)\}$ with the radius R:

$$\|\phi(\mathbf{x}_j) - \mathbf{a}\|^2 \leq R^2 + \xi_j, \tag{2}$$

where $\mathbf{a}$ is the center and $\xi \geq 0$ are the slack variables that enables soft boundaries calculation. Using the Lagrangian with a regularization constant C in its penalty term, $L = R^2 - \sum_j \left(R^2 + \xi_j - \|\phi(\mathbf{x}_j) - \mathbf{a}\|^2 \right) \beta_j - \sum_j \xi_j \mu_j + C \sum_j \xi_j$, problem (2) can be solved by dealing with its dual problem:

$$\max\ W = \sum_j \phi(\mathbf{x}_j)^2 \beta_j - \sum_{i,j} \beta_i \beta_j \phi(\mathbf{x}_i)\phi(\mathbf{x}_j) \tag{3}$$

$$s.t.\ \ 0 \leq \beta_j \leq C,\ \sum_j \beta_j = 1,\ j = 1 \ldots N. \tag{4}$$

Support vectors are the points on the boundary of the sphere with $0 < \beta_j < C$. The trained kernel radius function is defined as $f(\mathbf{x}) := R^2(\mathbf{x}) = \|\phi(\mathbf{x}) - \mathbf{a}\|^2 = K(\mathbf{x}, \mathbf{x}) - 2 \sum_j \beta_j K(\mathbf{x}_j, \mathbf{x}) + \sum_{i,j} \beta_i \beta_j K(\mathbf{x}_i, \mathbf{x}_j)$, where the kernel function $K(\mathbf{x}_i, \mathbf{x}_j)$ substitutes the inner products $\phi(\mathbf{x}_i) \cdot \phi(\mathbf{x}_j)$. In this work, we use the most commonly used kernel function, Gaussian kernel, which is defined as $K(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\|\mathbf{x}_i - \mathbf{x}_j\|^2/2\delta^2\right)$.

(2) Cluster labeling: As the trained kernel radius function is topologically invariant, the level set of $f(\cdot)$ can be decomposed into several disjoint sets [10], $L_f(\hat{R}^2) := \{\mathbf{x} : f(\mathbf{x}) \leq \hat{R}^2\} = C_1 \cup \ldots \cup C_p$, where $\hat{R} = R(\mathbf{x}_i), i = 1 \ldots N$ is the radius in kernel space, p is the number of clusters determined by $f(\cdot)$, and C_i, $i = 1 \ldots p$ are different clusters.

3   Experiment and Results

3.1   On Simulated Dataset

For the simulated dataset, the data processing scheme (cf. Fig. 1) is performed. After preprocessing, the t-test was used to reduce the dataset, which generated the result as shown in Fig. 3(a). As there are three different activation patterns in the simulated dataset, our target is to find four clusters: three contain three types of activated voxels, and the fourth is for the non-activated voxels. The performance of SVC is compared with those of crisp C-means, ICA [11], and FCM. For crisp C-means and FCM, we set the number of clusters to be four. ICA was constrained to generate three independent components. For the SVC, we set the kernel width σ to 0.2 and the regularization parameter C to 10^3. Four clusters are detected by SVC. In FCM, the cluster label for each voxel is determined by
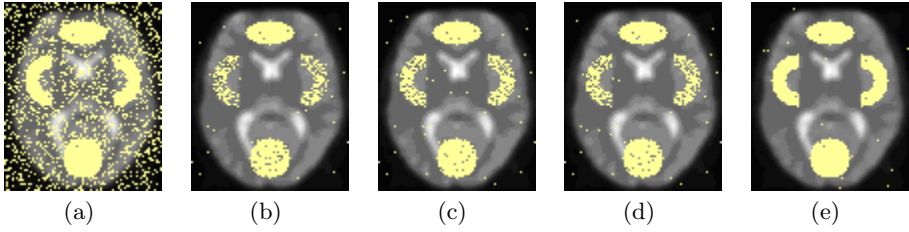
Fig. 3. Detection results from (a) after t-test (b) C-means (c) ICA (d) FCM (e) SVC
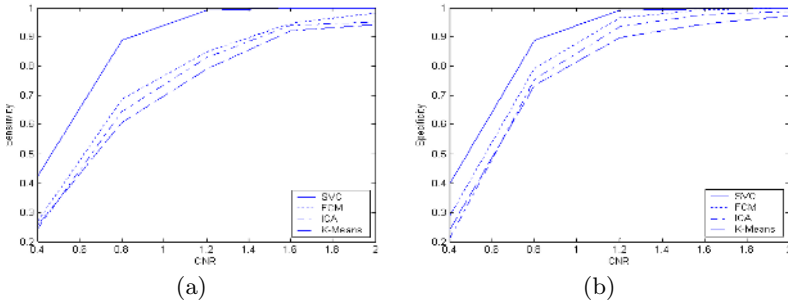

Fig. 4. Performance of C-Means, ICA, FCM, and SVC (a) Sensitivity (b) Specificity

finding the largest membership value; and in ICA, each voxel is assigned to the largest component. For the detected clusters, we find the "activated" clusters whose centroids are the most correlated ones to the stimulus paradigm.

Fig. 3 shows the detection results from C-means, ICA, FCM, and SVC, when CNR is set to 1.2. When CNR increases from 0.4 to 2, the sensitivity (the percentage of voxels correctly detected as activated) and specificity (the percentage of voxels correctly detected as non-activated) of the four methods are plotted in Fig. 4. One can find that the proposed SVC achieves both higher sensitivity and specificity than other methods under different CNR's.

3.2   On *In Vivo* Dataset

For the *in vivo* dataset, as we do not know the desired number of clusters, in order to achieve a set of finer clusters, we set the width parameter σ to 0.15 in SVC; while for FCM and C-means, the number of clusters is set to 30. Other settings are similar as discussed in 3.1. Fig. 5 shows that activated voxels on the 29th slice detected by C-means, FCM, ICA, and SVC cover the Brodmanns area (BA) 42 (primary auditory cortex) and BA 22 (auditory associated area) when they are superimposed on the accompanied structural data. But SVC detected more continuous activated regions, which are demonstrated to have stronger relationship with the stimulus (see Fig. 6(a)); however the voxels detected by C-means, but not by SVC, are shown to be almost irrelevant (see Fig. 6(b)).
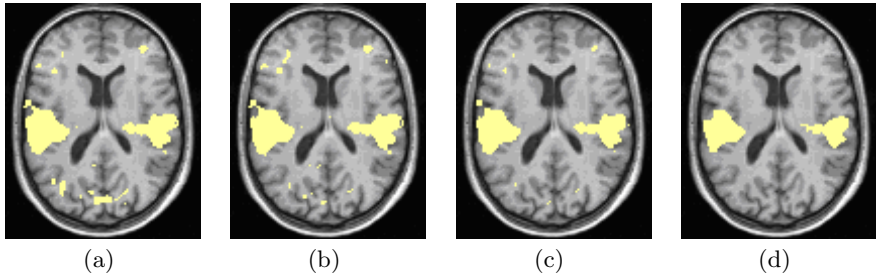
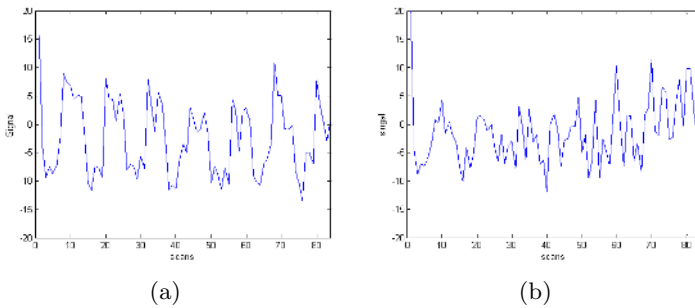Fig. 5. The activated clusters detected by (a) C-means, (b) ICA, (c) FCM, (d) SVC

Fig. 6. The *in vivo* activation detection results (a) the averaged signal detected by SVC (b) The averaged signal detected by C-means but not by SVC

4   Discussions and Conclusion

The main contribution of this paper is finding an effective and reliable activation detection method, support vector clustering (SVC), which is free of the cluster shape assumption existing in most fMRI clustering approaches, and is able to remove the outlier points. The experimental results show the effectiveness of SVC in comparison with commonly used crisp C-means, FCM and ICA.

In the SVC algorithm that uses Gaussian kernel, the width parameter σ determines how fine the samples are clustered. Users can have a control over the clustering result by tuning this parameter: increasing σ will lead to a coarse clustering result, and vice versa. Empirical value for σ is in the range from 0.1 to 3. The regularization constant C affects the amount of outlier points, and hence influences the compactness of the generated clusters. Further research may be on how to automatically determine the values for the two parameters in fMRI activation detection by incorporating the field expert knowledge.

Acknowledgment

References

1. Goutte, C., Toft, P., Rostrup, E., Nielsen, F.A., Hansen, L.K.: On clustering fmri time series. NeuroImage **9** (1999) 298–310
2. Chuang, K., Chiu, M., Lin, C.C., Chen, J.: Model-free functional mri analysis using kohonen clustering neural network and fuzzy c-means. IEEE Trans. Medical Imaging **18** (1999) 1117–1128
3. Baumgartner, R., Scarth, G., Teichtmeister, C., Somorjai, R., Moser, E.: Fuzzy clustering of gradient-echo functional mri in the human visual cortex. Magnetic Resonance Imaging **7** (1997) 1094–1101
4. Vapnik, V.: The Nature of Statistical Learning Theory. Springer-Verlag, New York (1999)
5. Ben-Hur, A., Horn, D., Siegelmann, H., Vapnik, V.: Support vector clustering. Machine Learning Research **2** (2001) 125–137
6. Purdon, P., Solo, V., Brown, E.M., Weisskoff, R.: Functional mri signal modeling with spatial and temporal correlations. NeuroImage **14** (2001) 912C923
7. Bandettini, P., Jesmanowicz, A., Wong, E., Hyde, J.: Processing strategies for time-course data sets in functional mri of the human brain. Magnetic Resonance in Medicine **30** (1993) 161–173
8. Lange, N., Zeger, S.: Non-linear fourier time series analysis for human brain mapping by functional magnetic resonance imaging. Applied Statistics **46** (1997) 1–29
9. Ffriston, K., Jezzard, P., Turner, R.: Analysis of functional mri time-series. Human Brain Mapping **2** (1994) 69–78
10. Lee, J., Lee, D.: An improved cluster labeling method for support vector clustering. IEEE Trans. Pattern Aanalysis Machine Intelligence **27** (2005) 461–464
11. Makeig, S.: Eeg/ica toolbox for matlab [online]. Available: http://www.sccn.ucsd. edu/ scott/ica.html (2005, April)