Biomedical Signal Processing and Control 3 (2008) 327–333

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

journal homepage: www.elsevier.com/locate/bspc

An MRF spatial fuzzy clustering method for fMRI SPMs

Lili He *, Ian R. Greenshields

University of Connecticut, Computer Science and Engineering. 371 Fairfield Road, Unit 2155, Storrs, CT 06269, United States

ARTICLE INFO

Article history:
Received 22 August 2007
Received in revised form 13 June 2008
Accepted 24 June 2008
Available online 19 August 2008

Keywords:
fMRI
Fuzzy c-means
Spatial fuzzy clustering
Markov Random Field
Statistical parametric maps

ABSTRACT

The paper presents a method for spatial fuzzy clustering (SFC) via Markov Random Fields (MRF) for the detection of brain activation regions in Functional Magnetic Resonance Imaging (fMRI) statistical parametric maps (SPMs) to improve the accuracy of the detection of such regions. The fMRI SPM is assumed to be an MRF and we define a fuzzy neighborhood energy function to describe the interaction between neighboring voxels. The final labeling is determined by a joint fuzzy membership. We compare the proposed spatial fuzzy clustering technique with the usual voxel-wise thresholding, traditional fuzzy clustering and Contextual Clustering (CC) [E. Salli, H.J. Aronen, S. Savolainen, A. Korvenoja, A. Visa, Contextual clustering for analysis of functional MRI data, IEEE Transactions on Medical Imaging 20 (2001) 403–414]. Experiments based on synthetic and real fMRI data demonstrate that the clustering performance of our method is significantly better than both simple thresholding and conventional non-spatial fuzzy clustering techniques. Our experiments also show that in relatively high quality SPMs (contrast to noise ratio (CNR) > 2.5), the performance of SFC and CC is very similar. In the case of the simulated datasets, when the SPMs have poor quality (CNR < 2.5), our method outperforms CC in reducing false positives and improving classification accuracy.

Published by Elsevier Ltd.

1. Introduction

Functional Magnetic Resonance Imaging (fMRI) provides a mechanism for observing hemodynamic changes typically based on blood oxygenation level dependent (BOLD) contrast which is subsequently correlated to neural activity. The analysis of fMRI data is generally carried out in three stages: preprocessing of the data, generation of Statistical Parametric Maps (SPMs) and finally analysis of the generated SPMs. The purpose of preprocessing is to separate function activation from artifacts in such a way as to enable a more accurate analysis of the data. Typical preprocessing steps include: inhomogeneity correction, baseline correction, spatial/ temporal smoothing and motion correction (among others). SPMs are usually formed in the second stage, by statistically comparing images taken during stimulation (ON) periods and those taken during rest (OFF) periods. The results are expressed as an image, where intensity values represent the test statistic under the null hypothesis of no activation at that voxel. Each statistic indicates the significance of the activation of the corresponding voxel by that stimulus. Among the approaches taken to generate SPMs are the General student t test [2], F and χ^2 tests [3], Kolmogorov–Smirnov test [4] and correlation analysis [5]. The merit of these methods is that no assumptions on the signal and noise distribution are needed; while the disadvantage is that the difference between two statistical distributions may not be determined if the difference involves higher order moments of the distributions. The major objective of analyzing SPMs is to detect activation regions that are activated by input stimuli. This can be done by thresholding an SPM at a proper significance p value [6] and/or at a specific cluster size [7]. This can also be viewed as an unsupervised clustering/segmentation [8–13] of voxels into active voxels and inactive voxels during a functional experiment [14]. However, both simple thresholding and conventional clustering algorithms are solely dependent on the intensities of image pixels (voxels) and/or the sizes of the clusters, and thus these techniques ignore spatial correlation, which is believed to be present in SPMs. In other words, there is a basic a priori assumption regarding fMRI that there are intrinsic spatial and temporal correlations in the data and therefore the data tend to have clustered activations. Thus, we expect that a single activation region will comprise spatially connected voxels. In fact, some clustering algorithms have utilized spatial information in the scenario of the segmentation of MR images [15,16].

Spatial regularization by means of random field theory (RFT) [17] has been incorporated into the traditional thresholding scheme. In particular, Gaussian random fields (GRFs) were implemented in [18]. GRFs allow not only the use of a fixed threshold but the more general use of an arbitrary threshold in terms of the spatial extent of the activated region to be used for the detection of significant

* Corresponding author. Tel.: +1 8604863719.
E-mail address: lili@engr.uconn.edu (L. He).

1746-8094/$ – see front matter. Published by Elsevier Ltd.
doi:10.1016/j.bspc.2008.06.003

activations. The GRF assumption necessitates spatial filtering due to its random nature to avoid false activations [19]. However, this process will increase the significance of large activated regions and decrease that of small activated regions such that large weakly activated or small strongly activated regions may be misinterpreted.

On the other hand, Markov Random Fields (MRFs) [19–21] have been utilized to include spatial correlations in the analysis of MRI/fMRI. In [22], Descombes et al. both restored fMRI signals and preserved transitions using a spatio-temporal MRF. Svensen et al. [23] constructed a mixture model based on an MRF for segmentation of the image into regions with different characteristics of the Hemodynamic Response Function (HRF). In addition, instead of modeling raw time series, Holmes [24] proposed to utilize MRFs to analyze SPMs from PET studies. SPMs from fMRI were thereafter analyzed and modeled via MRFs in [1,25,26]. In contrast to the previous hard clustering efforts, we present a spatial fuzzy clustering framework in the paper. One particular technique of interest to us is the method known as Contextual Clustering (CC) [1]. In CC, spatial information is taken into account by considering the number of activated neighbors of a voxel and a subsequent contextual classification rule is derived. On the other hand, here we formalize the interaction between neighboring voxels by a fuzzy neighborhood energy function and the final labeling is determined by a joint fuzzy membership. In the following sections, we discuss and compare CC against the method described here, i.e., spatial fuzzy clustering (SFC). SFC is inspired by [27], where a spatial-spectral fuzzy c-means procedure was implemented to produce thematic maps from remotely sensed multispectral imagery. We extend it to a more generalized and systematic spatial fuzzy clustering framework and apply it to the detection of activation regions in fMRI SPMs. The SFC framework can be adapted to any known general fuzzy clustering method, and it happens that spatial fuzzy c-means is a special case of the more general framework. We validate the method by testing it on simulated data as well as on real fMRI data. The results indicate that SFC is better than voxel-wise thresholding and conventional fuzzy clustering methods in increasing classification accuracy and controlling false positive rates, as well as preserving extremely small (even single voxel) activation regions with significant high statistical values. The proposed SFC performs similar to CC in high CNRs, and is more robust to noise than CC in low CNRs.

2. Materials and methods

2.1. Contextual Clustering

Contextual Clustering (CC) [1] requires the distribution of non-active voxels in a statistical parametric map to be Gaussian. Therefore, in applying CC to SPMs, such SPMs should be computed by a method which leads to the resulting distribution of non-active voxels which are then able to be transformed into a (normal) Gaussian. The intensity value of each voxel in an SPM represents the test statistic at that voxel by which the regions showing significant signal change corresponding to the task might be identified. For the purpose of investigating CC, an unpaired Student t statistic is first computed. The t statistic is calculated on a voxel-by-voxel basis following standard methods in fMRI to compare the task stimulus (ON state) against the "at rest" baseline (OFF state). Corresponding to each statistic is a significance probability (the p value). For a time series X, in which there are (known) n_a images during the task state and n_b images during the rest state, the t statistic is calculated pixel by pixel as follows:

$$t = \frac{\bar{X}_a - \bar{X}_b}{S_{\bar{X}_a - \bar{X}_b}} \tag{1}$$

where,

$$S_{\bar{X}_a - \bar{X}_b} = \sqrt{\frac{s_p^2}{n_a} + \frac{s_p^2}{n_b}} \tag{2}$$

and s_p^2 is the pooled variance

$$s_p^2 = \frac{\sum (X_a - \bar{X}_a)^2 + \sum (X_b - \bar{X}_b)^2}{n_a + n_b - 2} \tag{3}$$

The t statistic given above weights the difference in means by the standard deviations in the ON and OFF states. Large differences with small standard deviations give high t values and small differences with large standard deviations give low t values. This generates a t-distribution with $n_a + n_b - 2$ degrees of freedom (d.f.) under the null hypothesis of no activation. If d.f. is large, the distribution can be considered as being approximately Gaussian. Otherwise, it can be transformed to a Gaussian distribution by computing

$$z = q_z(p_t(t, \text{d.f.})) \tag{4}$$

where q_z is the normal inverse distribution function and $p_t(t, \text{d.f.})$ is the cumulative distribution function for t with (d.f.) degrees of freedom. After the transformation, the z map is produced.

Let f be the probability density function of Gaussian distribution. One defines a critical value z_α as:

$$\int_{z_a}^{\infty} f(z)\,\mathrm{d}z = \alpha \int_{-\infty}^{\infty} f(z)\,\mathrm{d}z \tag{5}$$

For a lower one-sided test, if $z < z_\alpha$, $(1-\alpha)\%$ is the confidence value that the null hypothesis is false.

Given the above distribution, it can be seen that the z map satisfies the prerequisite required by CC. The CC method is then summarized as [1]: (1) label the voxels with $z_i < T$ (where T is a user specified threshold) as *active* and other voxels as *non-active*. (2) compute for all voxels i the number of active neighbor voxels, u_i. (3) relabel the voxels for which

$$z_i + \frac{\beta}{T}\left(u_i - \frac{N}{2}\right) < T \tag{6}$$

as *active* and other voxels as *non-active*. In Eq. (6), N is the number of neighbors. For example, using 8-connectivity, $N = 8$. The parameter β determines the weighting of the contextual information and is usually positive. When β equals 0, Eq. (6) becomes a conventional context-free thresholding rule. (4) the clustering procedure stops when the current labeling is same as the labeling obtained in the previous iteration; otherwise go to step (2).

The parameter β is recommended (in [1]) to be set to $\beta = (T^2/s)$, where s specifies the excess of activated voxels $(u_i - (N/2))$ in the neighborhood required to classify a non-active voxel to be active with probability 0.5. Note that a key part of the success of CC is in the determination of the threshold T. The potential pitfall of CC can arise in a situation where the distribution of the background area is non-normal Gaussian or even a non-Gaussian.

2.2. Markov Random Fields and Gibbs fields

MRF is a probability theory to describe and analyze the spatial or contextual dependencies. We follow standard notation here. Let S be a finite index set (the set of sites/locations). For every site $s \in S$ there is a (finite) space X_s of states x_s. The space of configurations $x = (x_s)_{s \in S}$ is the product $X = \prod_{s \in S} X_s$. We consider probability measures or distributions $\prod$ on X; they can be represented by $\prod = (\prod(x))_{x \in X}$ such that $\prod(x) \geq 0$ and $\sum_{x \in X} \prod(x) = 1$. A collection $\zeta = \{\zeta\{s\} : s \in S\}$ of sites is called a neighborhood system, if $s \notin \zeta\{s\}$