Biomedical Signal Processing and Control 33 (2017) 48–65

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

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

A novel layered data reduction mechanism for clustering fMRI data

Xiao-Yan Tang[a,b], Wei-Ming Zeng[a,*], Ni-Zhuan Wang[a], Yu-Hu Shi[a], Le Zhao[a]

[a] *College of Information Engineering, Shanghai Maritime University, Shanghai 201306, China*
[b] *College of Physics & Electronic Engineering, Changshu Institute of Technology, Changshu 215500, China*

ARTICLE INFO

Article history:
Received 28 January 2016
Received in revised form 9 August 2016
Accepted 12 November 2016

Keywords:
Data reduction
FCM
Isolated voxel
Isolated generated cube

ABSTRACT

Original fMRI data often contains a variety of noise caused by the operator, the equipment, the environment, etc. To suppress the noise, many processing methods based on smoothing have been proposed to analyze the fMRI data. In this study, a layered data reduction mechanism is presented to alleviate the influence of noise while retaining the spatial construction of the fMRI data. The layered data reduction method consists of two layers criteria to reduce the noise voxels: (a) the isolated voxel; (b) the isolated generated cube. The 1-layer data reduction procedure aims to remove all those isolated voxels whose corresponding generated cube only contains one single voxel under a preset threshold ξ. The 2-layer data reduction procedure is aimed at removing those isolated generated cubes whose corresponding final cube only contains one single generated cube. A simplified genetic algorithm (SGA) is proposed to determine the optimum threshold ξ adaptively. Meanwhile, to avoid that some useful information would be lost on account of all the isolated voxels and isolated generated cubes being reduced directly, a compensation mechanism taking advantage of multivariate RV measure is used to decrease the probability of incorrect data reduction. The classical FCM (Fuzzy c-means) method is adopted to cluster the data having been implemented by the layered data reduction method. Extensively experimental results show that the proposed layered data reduction method is effective and can efficiently improve the clustering accuracy on the hybrid data and the real fMRI data.

1. Introduction

For a variety of reasons, e.g., the operator, the equipment, and the environment, fMRI data always contain a large amount of spatial random noise. This is problematic for further data processing [1,2]. In order to detect the weak useful signals in very noisy data to analyze local region's spatial activation or connectivity pattern, the fMRI data is often smoothed before performing subsequent process. As a consequence, it also makes the functional connectivity map blurred and obscure or even loses the fine-grained functional connectivity between different nearby regions [3]. In addition, univariate analysis methods are very sensitive to the noise and tend to generate a salt-and-pepper-like connectivity map when using conventional distance metrics methods [4]. Based on the superior properties in a noisy and sensitive case than that of traditional univariate methods, multivariate analysis methods have been got much attention in recent years, and many multivariate analysis methods have been used to analyze fMRI data [3–8]. Harrison et al. used multivariate autoregressive (MAR) models to make inferences about functional integration within the human brain [5]. Baumgartner et al. used Kendall's coefficient of concordance (KCC) to assess the cluster homogeneity in fMRI data [6]. Zang et al. also proposed a regional homogeneity (ReHo) method, which assumed that a given voxel was temporally similar to that of its neighbors, and used the KCC to measure the ReHo of the time series of a given voxel with those of its nearest neighbors [7]. Kriegeskorte et al. used a searchlight and multivariate statistic of Mahalanobis distance to analyze multivariately at each location in the brain [8]. Borrowing the idea of bilateral filtering in imaging processing proposed by Tomasi et al. [9] and the searchlight of Kriegeskorte et al. [8], Zhang et al. obtained a search cube, which contained multiple neighboring voxels with a particular size and shape, centered on a particular voxel and used the RV coefficient to measure the temporal similarities of two local brain regions [3].

The multivariate analysis methods provide efficient ways for us to measure the similarity between two sets of variables. However, the key of a multivariate analysis method is the similarity metrics, and different similarity metrics employed may result in different analysis results. A few important multivariate similarity metrics, e.g., KCC, Mahalanobis distance, RV coefficient and Pearson correlation coefficient, have been successfully applied to the analysis of fMRI data [7,8,10–14]. Due to that it combines the temporal infor-

* Corresponding author.
E-mail address: zengwm86@163.com (W.-M. Zeng).

http://dx.doi.org/10.1016/j.bspc.2016.11.014
1746-8094/

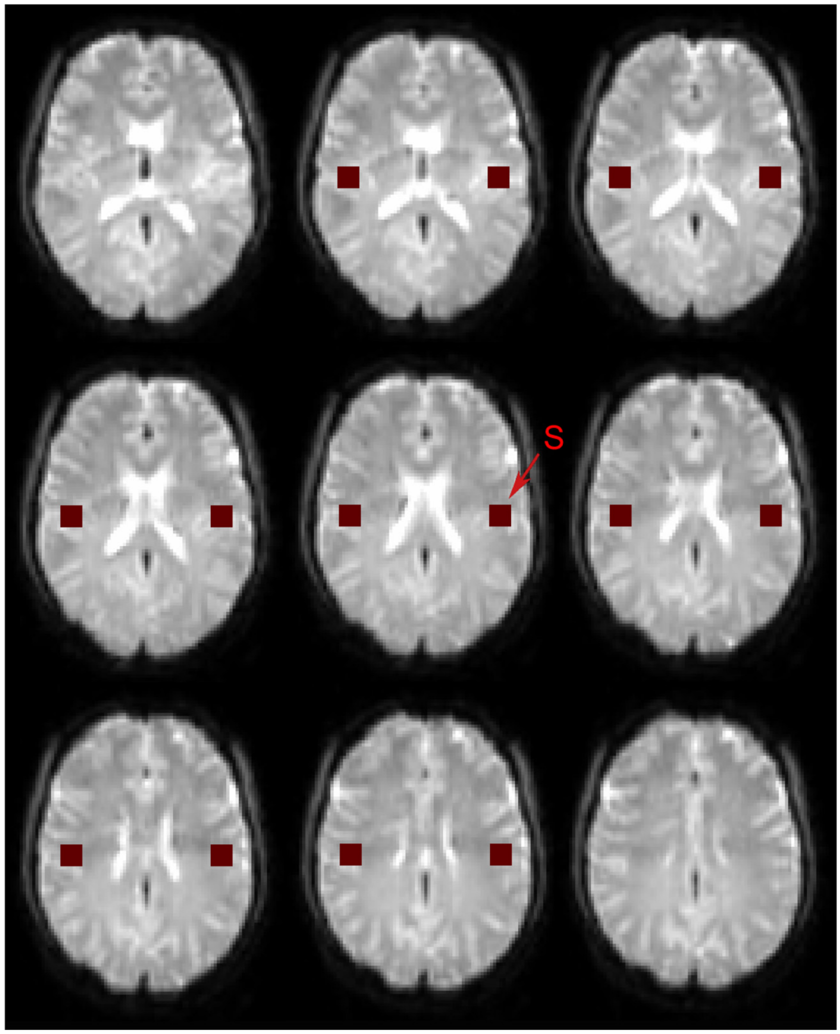

Fig. 1. The spatial maps related to the hybrid data from the 42th to 50th slice.

mation and spatial structure of local multi-voxels, the RV coefficient is demonstrated to have excellent performance in measuring the temporal similarities of two local brain regions [3,10]. In our previous study, we have used the RV coefficient to efficiently measure the similarity between two sets of multi-voxels contained in the adaptive generated cubes, which has demonstrated the quite good performance [10].

Inspired by the idea of ReHo [7] and multivariate analysis, we present a novel layered data reduction mechanism to alleviate the influence of noise while retaining the spatial construction of the original fMRI data. In our layered data reduction method, we think that those voxels, which have nothing to do with some kind of functional activities or certain activated region, could be regarded as noise voxels. So, these noise voxels are useless for the subsequent processing and can be directly reduced rather than being carried out suppressing of noise by smoothing. Pearson correlation coefficient, as one of conventional distance metrics methods, can efficiently measure the correlation degree between two variables and obtain more effective results than that of Euclidean distance [10]. Moreover, adopting the RV coefficient between one voxel and its neighbors as the base of our compensation mechanism can not only take the advantages of multivariate analysis method but also measure the similarity degree of two objects from another perspective different from traditional correlation methods, while KCC or Mahalanobis distance is not suitable in this case. Thus, the layered data reduction mechanism will combine the Pearson correlation coefficient between single variables with the RV measure between multiple variables to adaptively find and remove some obvious noise voxels, and improve the subsequent processing speed as well. In order to reasonably set the threshold of our method, we borrow some parameter optimization methods [15–19], and propose a simplified genetic algorithm (SGA). The main purpose of this study is to take the place of a traditional denoising method, i.e., smoothing, and alleviate the effects of noise while retaining the spatial feature of original data. In addition, we apply this layered data reduction strategy to the classical fuzzy c-means (FCM) algorithm to investigate whether the noise voxels would be removed effectively and the useful voxels would be retained as far as possible. The achieved improvement is validated quantitatively by various comparison metrics both on the hybrid data and on the real fMRI data.

2. Materials and methods

2.1. Data acquisition and processing

Five healthy subjects (two females and three males) were involved in this study. All of them were informed about the purpose of this study and signed a written agreement. Experimental tests consisted of the task-related experiment and the resting-state experiment. All the fMRI data were acquired on a Philips 3.0 T scan-