A dictionary learning algorithm for multi-subject fMRI analysis based on a hybrid concatenation scheme

Asif Iqbal *, Abd-Krim Seghouane

Department of Electrical and Electronic Engineering, Melbourne School of Engineering, University of Melbourne, VIC, 3010, Australia

ABSTRACT

Analysis of functional magnetic resonance imaging (fMRI) data from multiple subjects is at the heart of many medical imaging studies, and recently, the approaches based on dictionary learning (DL) are noted as promising solutions to the problem. However, the DL-based methods for fMRI analysis proposed to date do not naturally extend to multi-subject analysis. In this paper, we propose a DL algorithm for multi-subject fMRI data analysis which is derived using a hybrid (temporal and spatial) concatenation scheme. It differs from existing DL methods in both its sparse coding and dictionary update stages. It has the advantage of learning a dictionary common to all subjects as well as a set of subject-specific dictionaries, as a result, it is able to generate both group-level spatial activation maps as well as group-level temporal dynamics, which are particularly attractive for task-based fMRI studies. In addition, by simultaneously learning multiple sub-specific dictionaries, it also provides us with unique sub-specific features as well. Performance of the proposed DL method is illustrated using simulated and real fMRI datasets. The results show that it can successfully extract common as well as sub-specific latent components.

© 2018 Elsevier Inc. All rights reserved.

1. Introduction

In recent years, voxel-based functional magnetic resonance imaging (fMRI) activation detection has been extensively used to perform functional brain mappings where task-based or stimulus-driven paradigms have been used to study the brain function. Based on the relative change in the blood oxygen level dependent (BOLD) signal with respect to the baseline during task performance or as a response to a stimulus, inferences are made as to which areas of the brain are activated [15]. The methods used for such analysis can be characterized into two categories; model-based or data-driven. General linear model (GLM) and random field theory [16] are among the widely used model-based methods. However key drawbacks of model-based approaches is that they assume the hemodynamic response function (HRF) [28] to be known *a-priori* and they do not account for its variation across subjects and brain regions [19]. In addition, they can only extract features (spatial maps) corresponding to task-related activity and ignore the intrinsic brain function, which might not be directly linked to the external stimuli [37]. Moreover, model-based methods are not suitable for hard to model experimental paradigms, e.g. resting-state studies or naturalistic paradigms such as listening to music or watching a movie. The data-driven methods, on the other hand, do not make such assumptions and can be used in wide range of experimental paradigms. As a result, data-driven methods are getting more and more traction within the neuroscience community.

In contrast to model-based methods, data-driven approaches minimize the assumptions on the model structure and decompose the observed data based on a factor model and a specific constraint. These methods include principal component analysis (PCA) [3,23], independent component analysis (ICA) [21,33], canonical correlation analysis (CCA) [14,20], and multiset CCA [47,10]. ICA in particular has attracted a lot of interest from the community for analysis of both task-based and resting-state fMRI datasets [6]. More recently, brain networks have been found to be sparse in nature and a healthy debate has emerged regarding the starting point for fMRI data analysis, i.e., to use sparsity or independence [8,7]. The authors in [7] have clarified the issues with [8] and emphasized that the interpretations should be carefully made regarding the role of sparsity vs independence in the final decomposition of the data. Moreover, they went on to highlight the usefulness of sparsity as the starting point in fMRI analysis, which was also noted in [34], the pioneering work that started the activity in data-driven methods for fMRI analysis. As a result, sparse methods have seen increasing interest in solutions to fMRI data analysis.

Currently, ICA is being considered a go-to method for fMRI analysis. However, in recent literature, it is being argued that the ICA

methods are not all without drawbacks. In [58] (and references therein), authors have highlighted one such potential limitation, i.e., the ability to separately learn brain networks which have significant spatial overlap. In [31] (and references therein), it has been argued that certain brain regions appear to work in conjunction to perform various cognitive processes, i.e., a functional network (FN) could utilize multiple heterogeneous neuroanatomic areas. Furthermore, in [58] (and references therein), it has been shown that there exists extensive spatial overlap of large scale FNs in the brain. On the other hand, there exists another data-driven approach which is free of such limitation, this method is known as *Dictionary Learning (DL)*. In [58], authors have performed extended experiments to compare ICA with DL methods. They have concluded that in case of zero to low spatial overlap of FNs, ICA methods performance is better or on par with DL methods. Whereas, in case of moderate to severe spatial overlap of FNs, DL methods perform significantly better than the ICA methods. Thus using DL methods should provide a real benefit in the analysis of fMRI datasets.

Sparse signal representation [59] has seen a growing interest in the recent years and has led to state-of-the-art performance in the fields like signal representation [2,32], image denoising [11], brain tissue segmentation [38], and face recognition [56,54,22]. In these applications the main goal is to represent a given test signal **s** using only a few columns (atoms) from a given basis set (*dictionary*) **D**. Mathematically, the model used is $\mathbf{s} = \mathbf{Da}$ with sparse **a**, hence the term *sparse representation*. In [36], authors have shown that sparse linear codes (over wavelet basis set) for natural scenes develop receptive fields which are similar to the cells found in the primary visual cortex. Following this key motivating result, many DL and sparse coding methods have been proposed for single subject as well as multi-subject fMRI data analysis [55]. In [26], authors propose a sparse general linear model based on K-SVD algorithm [2], in [25] CCA and DL algorithm are used to learn interesting temporal dynamic components, which are then used to generate spatial activation maps via regression analysis. In [1] a fast and incoherent DL algorithm tuned for fMRI analysis is proposed. In [46] authors incorporated the correlation structure of the fMRI dataset into the DL framework and in [43,45] the smooth temporal dynamics are learned using basis expansion and regularization approaches for fMRI analysis. In the DL formulations presented in [57,30,31,60], the authors use [32] to decompose the fMRI dataset **S** into an overcomplete dictionary **D** and a sparse coefficient matrix **A**. Under this framework, atoms of the dictionary represent significant neuronal temporal dynamics and the rows of sparse coefficient matrix represent their respective spatial maps.

In literature, generally speaking, there are two distinct techniques in which the DL methods have been used to analyze multi-subject fMRI datasets. The first category includes the methods that decompose each subject fMRI dataset into a dictionary/sparse code matrix pair separately and perform brain activity analysis by locating components of interest (COI) across a group of subjects. The methods proposed in [26,57,30,31,53] lie in this category. In [30], authors learn a dictionary for each subject dataset and generate group-wise statistical maps by averaging over similar spatial maps. Authors in [53] propose a multi-subject DL algorithm where sub-specific maps are modeled as noisy versions of group-level maps, and learn sub-specific and group-level maps using a joint formulation. One key shortcoming of these methods is that they operate on individual datasets and as a result do not take into consideration the joint information across multiple datasets. In case of the aforementioned papers, performing group-analysis is not straight forward. Once each subject dataset is decomposed into a dictionary/sparse code matrix pair, the components of interest (temporal dynamics or spatial maps) might not be in the same place (column/row index) for all subject dictionary/sparse code matrices,

and may end up in different locations. Thus for joint analysis, matching of components via visual inspection or clustering are needed to be performed, which quickly becomes prohibitive once the number of subjects or size of the dictionary increases.

The second category contains those methods which avoid this shortcoming by using all subject datasets in the DL formulation [35,44,29]. In [35], authors decompose temporally concatenated multi-subject datasets into a single (shared) dictionary/sparse code matrix pair, where the sparse code matrix contains the group-level spatial activation maps and dictionary atoms correspond to sub-specific time courses. This method is suitable for studying the resting state or on-going activity experiments for which an experimental design to model the time courses is not available. On the other hand, authors in [44] have proposed a method tuned for the study of task-based experiments where the authors use spatially concatenated datasets to learn a shared dictionary, containing group-level (paradigm) time courses, and a sparse code matrix, containing subject-level spatial maps. However, we are interested in learning both group-level temporal dynamics and spatial maps which are of particular interest in task-based fMRI studies. To this end, the framework proposed in [29], similar to [44], uses spatially concatenated datasets to extract subject-level spatial maps, which are then decomposed into group-level spatial maps and a sparse loading matrix using [32]. This loading matrix is then used to discriminate different task types using a support vector machine (SVM). Although the framework proposed in [29] uses spatial and temporal concatenation to learn group-level features (time course or spatial map), which inherently introduces bias against unique sub-specific features present in the data. Furthermore, one might ask whether a specific recovered feature is a group-level feature or is only present in few (or certain) subject datasets. Currently available DL frameworks fail to answer this question. Authors in [24] have proposed a framework which tries to tackle this problem, however, they use a pre-averaged dataset to extract group-level features instead of joint analysis using all datasets simultaneously, which inherently is a sub-optimal strategy.

In this paper, we propose a novel DL framework which uses a hybrid (temporal and spatial) concatenation strategy to decompose the multi-subject fMRI datasets into a joint (group-level) dictionary/sparse code pair and unique (sub-specific) ones, resulting in the separation of joint and unique features into different dictionary/sparse code matrices. These pairs can then be used to perform population level analysis using group-level features, or analyze inter-subject variability using the extracted unique features. To perform this decomposition, our goal is to represent each voxels' time series from every subject dataset using only a few atoms from a joint-info dictionary and a sub-specific one. By using a hybrid fMRI data concatenation scheme, we separate similar neural dynamics (joint dictionary) and their respective spatial maps (joint sparse codes) from sub-specific counterparts. This formulation not only leads to group-level time courses (TC) and spatial maps (SM), but it also learns the most dominant sub-specific unique TCs and SMs as well.

In the next section, we briefly review DL and sparse coding problems. In section 2.2, we describe the proposed DL algorithm with the proposed solution derived in section 2.3. Section 3 contains performance analysis of the proposed method using simulated as well as real fMRI datasets. The concluding remarks are given in section 4.

2. Method overview

Notation: In rest of the paper, vectors are denoted by boldface lower case letters, e.g. **a**, matrices are denoted with upper case boldface letters, e.g. **A**, and any constant parameters are denoted by italics, e.g. *K*. A single subject fMRI dataset is denoted as $\mathbf{S}_i$