Medical Image Analysis 18 (2014) 435–448

Contents lists available at ScienceDirect

Medical Image Analysis

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

Discovering brain regions relevant to obsessive–compulsive disorder identification through bagging and transduction

Emilio Parrado-Hernández [a], Vanessa Gómez-Verdejo [a], Manel Martínez-Ramón [b,*], John Shawe-Taylor [c], Pino Alonso [d,e,f], Jesús Pujol [g], José M. Menchón [d,e,f], Narcis Cardoner [d,e,f], Carles Soriano-Mas [d,e]

[a] Dept. Signal Processsing and Communications, Universidad Carlos III de Madrid, Madrid, Spain
[b] Department of Electrical and Computer Engineering, The University of New Mexico, 1 University of New Mexico, Albuquerque, NM 87131-0001, USA
[c] Dept. Computer Science and CSML, University College London, London WC1E 6BT, UK
[d] Bellvitge Biomedical Research Institute-IDIBELL, Psychiatry Service, Bellvitge University Hospital, Barcelona, Spain
[e] Carlos III Health Institute, Centro de Investigación Biomédica en Red de Salud Mental (CIBERSAM), Spain
[f] Dept. of Clinical Sciences, School of Medicine, University of Barcelona, Barcelona, Spain
[g] CRC Hospital del Mar, Barcelona, Spain

ARTICLE INFO

Article history:
Received 27 May 2013
Received in revised form 16 January 2014
Accepted 25 January 2014
Available online 3 February 2014

Keywords:
Machine learning
Obsessive–compulsive disorder
MRI
Voxel selection
Transduction

ABSTRACT

In the present study we applied a multivariate feature selection method based on the analysis of the sign consistency of voxel weights across bagged linear Support Vector Machines (SVMs) with the aim of detecting brain regions relevant for the discrimination of subjects with obsessive–compulsive disorder (OCD, $n = 86$) from healthy controls ($n = 86$). Each participant underwent a structural magnetic resonance imaging (sMRI) examination that was pre-processed in Statistical Parametric Mapping (SPM8) using the standard pipeline of voxel-based morphometry (VBM) studies. Subsequently, we applied our multivariate feature selection algorithm, which also included an L_2 norm regularization to account for the clustering nature of MRI data, and a transduction-based refinement to further control overfitting. Our approach proved to be superior to two state-of-the-art feature selection methods (i.e., mass-univariate *t*-Test selection and recursive feature elimination), since, following the application of transductive refinement, we obtained a lower test error rate of the final classifier. Importantly, the regions identified by our method have been previously reported to be altered in OCD patients in studies using traditional brain morphometry methods. By contrast, the discrimination patterns obtained with the *t*-Test and the recursive feature elimination approaches extended across fewer brain regions and included fewer voxels per cluster. These findings suggest that the feature selection method presented here provides a more comprehensive characterization of the disorder, thus yielding not only a superior identification of OCD patients on the basis of their brain anatomy, but also a discrimination map that incorporates most of the alterations previously described to be associated with the disorder.

1. Introduction

Obsessive–compulsive disorder (OCD) is a chronic psychiatric disorder with an approximate worldwide prevalence of 2% that causes significant interference with familial, work and social functioning (Karno et al., 1988). Prevailing neurobiological models of OCD are based in part on quite solid neuroanatomical findings accumulated over the course of the past years by means of the analysis of structural magnetic resonance imaging (sMRI) data. In particular, Voxel Based Morphometry (VBM) (Ashburner and Friston, 2000) has probably been the most widely used approach to characterize brain structural abnormalities in OCD and other psychiatric disorders (Radaelli et al., 2008). On the one hand, VBM provides a common and largely automated preprocessing framework for morphometry studies, thus facilitating the comparison between results from different centers. On the other hand, statistical inference in VBM studies is based on the principles of the quite intuitive and well-known general linear model (GLM). Such research has enabled the examination of previous neuroanatomical hypotheses relating to the disorders and the proposal of new disease models with solid biological foundations, although the results have been ill translated into clinical practice (Orrù et al., 2012). Consequently, there has been an increasing interest in applying other analysis strategies, such as machine learning (ML), to describe between-group differences in terms of actual discrimination between populations (Lemm et al., 2011).

---

* Corresponding author. Tel.:+1 (505) 277 3008.
E-mail address: manel@unm.edu (M. Martínez-Ramón).

http://dx.doi.org/10.1016/j.media.2014.01.006
1361-8415/

ML is a discipline devoted to the estimation of unknown variables on the basis of empirical observations (Vapnik, 1995). Due to its multivariate nature, the particular application of ML methods to the analysis of sMRI data in patients with mental disorders has made it possible to characterize the brain alterations of such disorders as spatially distributed patterns of interrelated alterations. Likewise, ML methods have also made the characterization of brain alterations possible at an individual level, as opposed to the group level differences provided by other approaches (Orrù et al., 2012). Both assets are extremely valuable for research in biological psychiatry, which characteristically deals with subtle, distributed brain structure alterations in the context of a lack of reliable biomarkers for diagnosis, disease progression and treatment effectiveness monitoring.

Nevertheless, the application of ML algorithms to sMRI data needs to consider the dimensionality of each data point in relation to the number of available data samples (typically, a difference between three and four orders of magnitude). Thus, in the best scenario, one may have hundreds of samples, while each sMRI may contain more than 10^5 voxels. This situation, where the number of instances l is much smaller than the dimensionality of the input space D, is known in ML as a small sample size problem. A small sample size problem implies a high risk of overfitting, i.e., learning models that perform extremely well predicting the training data but very poorly with the test data not available at training time. The classic ML tricks to overcome situations where $l \ll D$ aim at either pretending to effectively increase l or reducing D. A virtual augment of l can be obtained by resampling methods such as bootstrap (Efron, 1979) or bagging (Breiman, 1996). There are two main classes of methods to reduce D, feature extraction (FE) and feature selection (FS).

On the one hand, FE approaches generate a new (reduced) subset of features based on the original ones, they usually apply a transformation of the input variables in a reduced set by minimizing the information loss or maximizing the predictive capability of these new features. A well known disadvantage of FE methods is that the new set of variables losses interpretability, since it is relatively hard to map back the outcome of the algorithm to the original space. Principal Component Analysis (PCA) (Pearson, 1901) is perhaps the most broadly FE method. In fact, Mourão Miranda et al. (2005) apply it to reduce the number of voxels in a brain-mapping problem with fMRI data, although it has to include a post-processing step based on a permutation test to remove irrelevant voxels and, thus, to gain interpretability. The input dimension in a sMRI is so large that indeed most papers presenting FS methods also perform a preprocessing FE step gathering higher level features from groups of neighboring voxels. For instance, Bunea et al. (2011) generate a reduced subset of features based on the neuroimaging and demographic information, whereas other works (see, for instance, Varoquaux et al., 2012 or Wang et al., 2010) use features extracted by representing groups of adjacent voxels with a measure over the cluster as its mean value. The reduced sets of extracted features still suffer from the curse of dimensionality, therefore the FE are combined with other dimensionality reduction method; for instance, Bunea et al. (2011) and Varoquaux et al. (2012) combine the feature extraction stage with a bootstrap approach or Wang et al. (2010) builds an ensemble of bagged SVM learners.

On the other hand, FS techniques assume that the information is just in a subset of the input variables and tries to find it. State-of-the-art FS techniques are usually classified in three groups: (1) Filters, which are based on relevance measures resulting from statistical tests or Mutual Information estimators (Torkkola et al., 2003); (2) wrapper approaches (Kohavi and John, 1997) or (3) embedded methods, such as, the Recursive Feature Elimination (RFE) method proposed by Guyon et al. (2002). A more recent trend in ML proposes to introduce regularizations that lead to a FS in the design of the actual classification or regression algorithm. Examples of these techniques are lasso regression (Tibshirani, 1994), group lasso (Yuan and Lin, 2006), elastic networks (Zou and Hastie, 2005), etc.

Indeed, most of these FS tricks for coping with small sample size situations have already been applied to neuroimaging data and, in particular, to sMRI. Thus, for instance, Mitchell et al. (2004) analyze the performance improvement provided by different relevance measures during the classification of cognitive states, Costafreda et al. (2009) use a FS step based on mass-univariate t-Tests followed by an SVM classification of depressive patients versus healthy controls, or De Martino et al. (2008) and Ecker et al. (2010) apply the RFE method to map and classify fMRI spatial patterns and to discriminate autistic subjects from healthy controls, respectively. In this context, despite recent reports suggesting that classification accuracy may only be increased by increasing sample sizes (Chu et al., 2012), which is not always easy to achieve in studies with clinical populations, further improvements in FS methods may not only lead to an overall improvement in classification accuracy, but also to a more valid and comprehensive characterization of the brain regions relevant for disease characterization.

Recent works propose to embed the FS step within the classification stage. Following this research line, the works of Yamashita et al. (2008) or van Gerven et al. (2010) include regularization terms based on L_1 norms to arrive at models that end up using a very reduced set of the input variables (in the order of the sample size). However, the application of L_1 regularization to raw neuroimaging data, would provide too sparse meaningless solutions where a few isolated voxels making up the whole brain map. This limitation can be overcome combining the L_1 norm with a L_2 one in an elastic net fashion (Carroll et al., 2009; Ryali et al., 2010) or including spatial regularizations (Cuingnet et al., 2011; Michel et al., 2011) to force the selection of sets of spatially close voxels with a significant size. However, the optimization involved in these proposals results in a computational problem. In some cases these optimizations have to be solved in a primal space with hundreds of thousands of variables. In some other cases they result in very specific ad hoc optimization problems that cannot be solved with standard toolboxes.

The fact that OCD is related to brain structural abnormalities results in that the majority of the voxels do not contain relevant information for the discrimination of the disease. For this reason, FS seems the most appropriate complement to tackle the discrimination of the disease with ML. The advantage of using FS methods is twofold. Firstly, by dramatically reducing the number of effective voxels by removing the non-informative ones, it is possible to identify those areas that are relevant for the disease characterization. Such a map would enable the interpretation of the outcome of the ML algorithm by clinicians and neuroimaging experts (Parrado-Hernandez et al., 2012a). Secondly, by removing non-informative signals (i.e., noise), it is expected to increase classification accuracy (Guyon and Elisseeff, 2003). Moreover, an FS method working in the voxel space (without preprocessing FE based on a priori established brain regions or clusters of voxels) would be able to discover structural abnormalities at different levels of detail. In addition, the set of selected voxels must be robust against overfitting; i.e., the selection method should pick up those voxels relevant for the discrimination between out-of-sample patients and healthy subjects (Parrado-Hernandez et al., 2012b).

The present paper further develops a FS scheme built on the starplots strategy proposed by Bi et al. (2003). The starplots method selects, as relevant for the task at hand, those features whose role in the final machine is robust to small, controlled perturbations in the training set. In short, the starplots method begins by training an ensemble of linear machines whose training sets are obtained by undersampling the original dataset. An L_1