NeuroImage 78 (2013) 270–283

Contents lists available at SciVerse ScienceDirect

NeuroImage

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

Analytic estimation of statistical significance maps for support vector machine based multi-variate image analysis and classification

Bilwaj Gaonkar *, Christos Davatzikos

Section for Biomedical image analysis, University of Pennsylvania, 3600 Market St., Suite 380, Philadelphia, PA 19104, USA

ARTICLE INFO

Article history:
Accepted 26 March 2013
Available online 10 April 2013

Keywords:
SVM
Statistical inference
Neuroimaging analysis

ABSTRACT

Multivariate pattern analysis (MVPA) methods such as support vector machines (SVMs) have been increasingly applied to fMRI and sMRI analyses, enabling the detection of distinctive imaging patterns. However, identifying brain regions that significantly contribute to the classification/group separation requires computationally expensive permutation testing. In this paper we show that the results of SVM-permutation testing can be analytically approximated. This approximation leads to more than a thousandfold speedup of the permutation testing procedure, thereby rendering it feasible to perform such tests on standard computers. The speedup achieved makes SVM based group difference analysis competitive with standard univariate group difference analysis methods.

Introduction

Statistical parametric mapping (Frackowiak et al., 1997), voxel-based morphometry (Ashburner and Friston, 2000; Davatzikos et al., 2001) and related methods that apply voxel-wise statistical tests have been fundamental tools in modern neuroimaging. These methods have made it possible to quantify group differences and understand spatial patterns of functional activation/brain structure. Methods belonging to this family of mass-univariate methods are amenable to standard statistical inference techniques. Typically these methods associate a statistical significance measure such as a 'p-value' with every voxel. This allows for easy interpretation of the output from these methods. However, during the past decade, the neuroimaging community has recognized that multi-variate relationships among different brain regions cannot be captured by univariate analysis alone. This has lead to the development of multi-variate image analysis methods, which provide a more complete picture of imaging patterns that relate to brain activity, structure and pathology (Craddock et al., 2009; Cuingnet et al., 2011; Davatzikos et al., 2005; De Martino et al., 2008; Fan et al., 2007; Klöppel et al., 2008; Koutsouleris et al., 2009; Langs et al., 2011; Mingoia et al., 2012; Mouro-Miranda et al., 2005; Pereira et al., 1998; Richiardi et al., 2011; Sabuncu and Van Leemput, 2011; Vemuri et al., 2008; Venkataraman et al., 2012; Wang et al., 2007; Xu et al., 2009). Among the most successful of such methods are SVM-based tools (Fan et al., 2007; Klöppel et al., 2008), which have been quite widely used in functional (Craddock et al., 2009; Davatzikos et al., 2005; De Martino et al., 2008; Mouro-Miranda et al., 2005; Wang et al., 2007) and structural (Cuingnet et al., 2011; Fan et al., 2007; Klöppel et al., 2008; Koutsouleris et al., 2009; Vemuri et al., 2008) neuroimaging analysis.

However, interpretation of SVM models is difficult because unlike univariate methods (Ashburner and Friston, 2000), SVMs do not naturally provide statistical tests (and corresponding p-values) associated with every voxel/region of an image. Rather, it is considered normal to evaluate these models as "black boxes" on the basis of cross-validation accuracy, which is a measure of how accurately they detect the presence of disease based on data from an image. While cross-validation provides an overall estimate of the separability between two groups or conditions, it is unclear how each brain region contributes to the construction of the multivariate discriminatory pattern that ultimately drives the detection of disease. Further, while SVM models associate a 'weight coefficient' with every voxel/region of the image space they do not offer an analytic framework for estimating statistical significance of these weights, an issue of fundamental importance. Hence permutation tests have typically been used to understand what regions of the brain drive the SVM decision (Mouro-Miranda et al., 2005; Wang et al., 2007). These permutation tests are extremely expensive computationally. Hence they are largely prohibitive in many practical applications. In this paper, we show that, given the high dimensional nature of neuroimaging data, it is possible to analytically approximate the null distributions that we ordinarily generate using permutation tests. We verify this approximation by comparing it with actual permutation testing results obtained from several neuroimaging datasets. Some of this work is based on concepts first presented by us in Gaonkar and Davatzikos (2012). However, the derivations presented here are more generic. Further, we have added experiments that establish a) the multivariate nature of the inference made using such tests, b) advantages compared to typical univariate testing machinery, and c) advantages compared to inference based on sparse methods.

* Corresponding author.
E-mail addresses: bilwaj@gmail.com, Bilwaj.Gaonkar@uphs.upenn.edu (B. Gaonkar).

1053-8119/$ – see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.neuroimage.2013.03.066

Materials and methods

Background

Support vector machines

The support vector machine attempts to learn a model from data by finding the largest margin hyperplane that separates data from different conditions (e.g. baseline/activation) or groups (e.g. patients/controls). Training is the process of finding this hyperplane using data with known labels (condition, group, etc.). Now, for data with unknown labels (test data), the SVM uses the hyperplane found (during training) to estimate whether it belongs to a patient or to a control. The SVM treats individual data as points located in a high dimensional space. Fig. 1 illustrates the concept of the algorithm in an imaginary 2D space: dots and crosses represent imaging scans taken from two groups or conditions. Even though the two groups cannot be separated on the basis of values along any one dimension the combination of two dimensions gives perfect separation. This corresponds to the situation where a single anatomical region may not provide the necessary discriminative power between groups, whereas the multivariate SVM can still find the relevant hyperplane. Typical imaging data lives in an extremely high dimensional space determined by the number of voxels in each image.

To apply SVMs in neuroimaging data, we convert an image with D voxels into a vector whose dth component is equal to the intensity value at the dth voxel in the image. Thus, we re-organize the ith image into a D-dimensional point that lives in $\mathbb{R}^D$. Let us denote the ith point by $\mathbf{x}_i$ where $i \in 1,\ldots, m$ indexes all subjects in the study. In most imaging studies, we also have a label associated with each image which tells us whether the image belongs to a patient or a control subject. We denote these labels by $y_{(i)} \in \{+1, -1\}$. Then the support vector machine finds 'hyperplane coefficients' denoted by $\mathbf{w}^*$ and b^* such that:

$$\{\mathbf{w}^*, , b^*\} = \operatorname{argmin}_{\mathbf{w},b,\xi_i} \frac{1}{2}||\mathbf{w}||^2 + C\sum_{i=1}^{m} \xi_i \quad (1)$$

$$\text{subj.to} \quad y_i\left(\mathbf{w}^T\mathbf{x}_i + b\right) \geq 1-\xi_i \quad \forall i = 1,\ldots,m$$

$$\xi_i \geq 0 \quad \forall i = 1,\ldots,m.$$

The weight vector $\mathbf{w}^*$ represents the direction in which the SVM deems the two classes (controls and patients) to differ the most. To determine the label associated with a new test subject $\mathbf{x}_{test}$ we use $y_{test} = sign\ (\mathbf{w}^{*T}\mathbf{x}_{test} + b^*)$. Since the data $\mathbf{x}_{(i)}$ are in $\mathbb{R}^D$; the weight vector $\mathbf{w}^*$ is also in $\mathbb{R}^D$. Thus, $\mathbf{w}^*$ can be represented as an image which we call a 'discriminative map'. However, until now there has been a limited use of SVM based discriminative maps in neuroscience. This is because these maps do not provide a measure of statistical significance associated with a voxel/region of an image. What is the probability that a particular image voxel would have a weight vector component at least as large as the one observed in an experiment due to pure chance alone? To answer such a question, one needs to establish a null distribution on the weight vector components at each image voxel. An empirical approach for obtaining such a null distribution is through the use of permutation tests. We describe permutation testing in the next section.

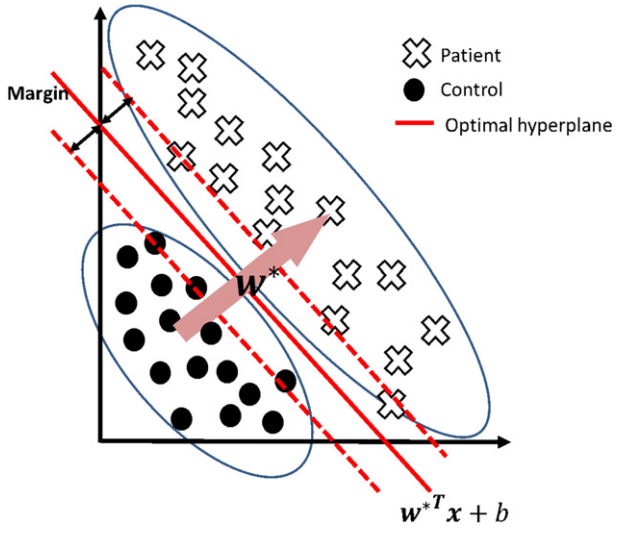

Fig. 1. Illustration of the SVM concept in two dimensions.

Permutation tests

Permutation testing can be used to establish a null distribution on the weight vector components at each image voxel. The permutation testing procedure is illustrated in Fig. 2. This procedure for permutation testing was applied in the context of neuroimaging by Mouro-Miranda et al. (2005) and Wang et al. (2007). In Fig. 2, the dots denote controls and the crosses denote patients. The first step involves the generation of a large number of shuffled instances of data labels by random permutations. Each shuffled instance is used to train one SVM. For each instance of shuffled labels, this generates one hyperplane parameterized by the corresponding vector $\mathbf{w}$. Then for any component of $\mathbf{w}$, we have one value corresponding to a specific shuffling of the labels. Collecting the values corresponding to any one component of $\mathbf{w}$ allows us to construct a null distribution for that component of $\mathbf{w}$. Recall that each component of '$\mathbf{w}$' corresponds to a voxel location in the original image space. Thus, we now have a null distribution associated with every voxel in the image space. Comparing each component of $\mathbf{w}^*$ with the corresponding null distribution allows us to estimate statistical significance.

While we run tests on each coefficient separately, it is crucial to note that permutation testing based inference is distinct from univariate inference. These tests are capable of identifying multivariate phenomenon that univariate tests cannot find. We further clarify this point using experiments on simulated data presented in the 'Experiments and results' section.

Further, it is also vital to note that the permutation test based inference method described here is distinct from thresholding SVM weights themselves which has been popular in literature. However, the thresholding approach is problematic and has also been repeatedly criticized in machine learning literature because a larger weight value does not necessarily indicate higher feature relevance. Limitations of the weight vector component thresholding do not simply carry over to the permutation testing methodology described here. We have included a simulated experiment to establish this fact. In the Experiments and results: comparison with prior art section, we show using simulated data that the proposed approach continues to work when SVM weight thresholding fails.

It is obvious that running 1000 permutation tests requires training 1000 support vector machine classifiers. This requires a significant amount of time (a few hours in our case) (In many applications, one might need 10,000 permutations or more). In contrast traditional SPM based methods (Frackowiak et al., 1997) can run in a few minutes. Further, some SVM applications involve running separate SVMs on local 3D windows in MR images in order to identify group differences (Rao et al., 2011; Xiao et al., 2008). In such cases, it is computationally infeasible to run the required number of permutation tests experimentally.

Analytical approximation to permutation tests: the case of balanced data

The primary aim of this work is to show that the permutation testing procedure described above can be replaced by an analytic alternative that can be computed in a small fraction of the time (a few