NeuroImage 56 (2011) 531–543

Contents lists available at ScienceDirect

NeuroImage

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

Dimensionality estimation for optimal detection of functional networks in BOLD fMRI data

Grigori Yourganov [a,b,*], Xu Chen [c], Ana S. Lukic [d], Cheryl L. Grady [b,e], Steven L. Small [f], Miles N. Wernick [g,d], Stephen C. Strother [a,b,d]

[a] Institute of Medical Science, University of Toronto, Toronto, ON, Canada
[b] Rotman Research Institute, Baycrest Centre for Geriatric Care, Toronto, ON, Canada
[c] Case Center for Imaging Research, Case Western Reserve University, Cleveland, OH, USA
[d] Predictek, Inc., Chicago, IL, USA
[e] Department of Psychology, University of Toronto, Toronto, ON, Canada
[f] Department of Neurology, University of Chicago, Chicago, IL, USA
[g] Medical Imaging Research Center, Illinois Institute of Technology, Chicago, IL, USA

ARTICLE INFO

Article history:
Received 27 November 2009
Revised 10 September 2010
Accepted 14 September 2010
Available online 19 September 2010

Keywords:
Dimensionality estimation
Model order selection
Linear discriminant Analysis (LDA)
Principal component analysis (PCA)
Signal detection
fMRI

ABSTRACT

Estimation of the intrinsic dimensionality of fMRI data is an important part of data analysis that helps to separate the signal of interest from noise. We have studied multiple methods of dimensionality estimation proposed in the literature and used these estimates to select a subset of principal components that was subsequently processed by linear discriminant analysis (LDA). Using simulated multivariate Gaussian data, we show that the dimensionality that optimizes signal detection (in terms of the receiver operating characteristic (ROC) metric) goes through a transition from many dimensions to a single dimension as a function of the signal-to-noise ratio. This transition happens when the loci of activation are organized into a spatial network and the variance of the networked, task-related signals is high enough for the signal to be easily detected in the data. We show that reproducibility of activation maps is a metric that captures this switch in intrinsic dimensionality. Except for reproducibility, all of the methods of dimensionality estimation we considered failed to capture this transition: optimization of Bayesian evidence, minimum description length, supervised and unsupervised LDA prediction, and Stein's unbiased risk estimator. This failure results in sub-optimal ROC performance of LDA in the presence of a spatially distributed network, and may have caused LDA to underperform in many of the reported comparisons in the literature. Using real fMRI data sets, including multi-subject group and within-subject longitudinal analysis we demonstrate the existence of these dimensionality transitions in real data.

Introduction

Typical fMRI data sets consist of a relatively small number of temporally correlated observations recorded for a large number of spatial locations (voxels). When trying to create activation maps with multivariate analysis and treating each fMRI volume as a vector, one is faced with an ill-posed problem, because the number of variables (voxels) greatly exceeds the number of observations (fMRI volumes). In addition, not all dimensions of the vector space are useful, as many are dominated by noise. A well-known strategy to reduce the dimensionality and noise prior to analysis is to discard a subset of the components in a principal component analysis (PCA) (Friston et al., 1995a; Strother et al., 1997; Tegeler et al., 1999; Hansen et al., 1999; Laconte et al., 2003). This requires estimation of the dimensionality of the signal-carrying subspace, a task for which many solutions have been proposed in the literature.

Intrinsic PC dimensionality of the data is defined as the number of principal components (PCs) that contain signal and should be retained for further analysis. Some early methods (or, more accurately, rules-of-thumb) of intrinsic dimensionality estimation are described in Mardia et al. (1979). For example, one can retain the PCs that, taken together, explain 90% of the variance in the data; or one can look for the "knee" in a "scree plot" (the point at which the eigenvalue spectrum of the covariance matrix flattens out, which in a white-noise model indicates noise-dominated components). Both of these methods are subjective, because the threshold of 90% is an arbitrary choice, and the scree-plot method involves visual inspection.

Beckmann and Smith (2004) discuss (alongside other methods of dimensionality estimation) a more-sophisticated technique based on scree plots, which works as follows. When the data are Gaussian-distributed, the estimated covariance matrix has a Wishart distribution, and one can calculate the expected distribution matrix's eigenvalues

* Corresponding author. Rotman Research Institute, Baycrest Centre for Geriatric Care, 3560 Bathurst St., Toronto, ON, Canada M6A 2E1. Fax: +1 416 785 2862.
E-mail address: gyourganov@rotman-baycrest.on.ca (G. Yourganov).

1053-8119/$ – see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.neuroimage.2010.09.034

analytically. PCs corresponding to eigenvalues that are not significantly different from the expected eigenvalues of a Wishart matrix are discarded. However, this method assumes a well-posed problem (the number of samples should be at least equal to the number of voxels) making it impractical for fMRI data analysis without voxel-based feature selection. Peres-Neto et al. (2005) gives a review of several other methods of dimensionality estimation that work for well-posed problems, and compared their performance when the data are not Gaussian-distributed. The best-performing methods were based on permutation tests, where independent observations were randomly shuffled 999 times.

Several techniques for estimating the number of signal-carrying PCs have been developed specifically for ill-posed problems. These methods can be classified into two categories: analytic and empirical. Analytic methods are based on information-theoretic criteria to estimate the optimal number of principal components. We have studied several analytic methods, such as minimal description length (Calhoun et al., 2001; Li et al., 2007), Stein's unbiased risk estimator (Ulfarsson and Solo, 2008), and Laplace approximation to Bayesian evidence (Minka, 2000), and examined their utility in situations usual to fMRI research, where sample size is small and signal-to-noise ratio is low.

Empirical methods of dimensionality estimation select the number of principal components to optimize some metric of performance calculated with resampling techniques; therefore, these methods are typically more computationally expensive than analytic methods. Early examples of empirical dimensionality estimation are Wold (1978) and Eastment and Krzanowski (1982); see also Krzanowski and Kline (1995). They proposed a method when the matrix is approximated with a subset of principal components, and dimensionality is estimated by the number of principal components that results in approximation with an optimal predicted residual sum of squares (PRESS) statistic. Each element of the matrix is compared with its corresponding element in a PC approximation. However, we need to make sure that the information about the element was not used in the approximation of this element (i.e. to ensure the independence of training and test data). For this purpose, the PC approximation of an element is computed on a matrix from which the row and the column containing this element have been removed. This leads to a rather cumbersome and very computationally expensive procedure of numerous PC decompositions on the matrices from which a particular row and column of the original data matrix were removed.

More recently, Hansen et al. (1999) proposed another method based on cross-validation, in which the data were separated into independent training and test sets and the number of PC components that minimize generalization error of the test set was used as an intrinsic dimensionality estimate. Strother et al. (1997, 2002) introduced the split-half resampling framework, where the data are split into two independent sets of roughly the same size. The number of PCs to be retained is selected to optimize the reproducibility of activation maps (calculated separately on the two half-sets), or classification accuracy (when one half serves as a training set, and the other as a test set). There is usually a tradeoff between reproducibility and classification accuracy, and one might wish to optimize a combination of these two metrics (LaConte et al., 2003; Strother et al., 2004; Jacobsen et al., 2008).

Several authors have compared the efficacy of various dimensionality estimation methods. Minka (2000) has shown that his method of optimizing Bayesian evidence is significantly more accurate than 5-fold cross-validation if the sample size and number of voxels in simulated data are both small (<15). His method was also better than cross-validation when the data were non-Gaussian, which was the reason to use his method in probabilistic independent component analysis (Beckmann and Smith, 2004), in which the signal sources are assumed to be non-Gaussian. When the number of observations was larger (>60, in both well-posed and ill-posed situations), optimization of Bayesian evidence was found to be slightly, but not significantly, better than cross-validation in estimating the dimensionality of simulated Gaussian data. Cordes and Nandy (2006) have shown that estimates of dimensionality calculated with analytic methods are strongly influenced by sample size (when the number of observations grows, so do the dimensionality estimates, although the underlying intrinsic dimensionality stays the same). Li et al. (2007) have addressed this problem by subsampling the data, when a large portion of observations is discarded prior to dimensionality estimation so the remaining observations are independent and identically distributed. Ulfarsson and Solo (2008) have shown their method to be more accurate than both Minka's method and minimum description length in simulated Gaussian data. In their simulations, the ratio of the number of variables to the number of observations did not exceed 5/2.

Using the simulation framework introduced in Lukic et al. (2002), we tested a wide selection of dimensionality estimation methods, and compared their performance in situations when the task-related signal was organized into a spatial network of functionally connected loci. In the simulations, we sampled the signal from a multivariate Gaussian distribution, and embedded it in additive Gaussian noise. We used linear discriminant analysis (LDA) on a PC subspace as our basic method of analysis. When the loci are not correlated, signal detection is optimal when the number of PCs in our analysis is roughly equal to the number of loci. However, as the correlations increase and the spatial network starts to become apparent in the data, intrinsic dimensionality goes through a transition from many dimensions to a single dimension, and signal detection is optimized when we use just one PC. This transition is captured when we estimate the intrinsic dimensionality by optimizing the reproducibility of activation maps; all other methods of estimation fail to capture the transition.

Using a metric of global signal-to-noise ratio (gSNR) based on the reproducibility of independent spatial activation maps (Strother et al., 2002, 2010), we demonstrate an asymptotic relationship between optimal dimensionality and gSNR: when gSNR is high enough, the network can be captured with a single dimension, but as gSNR drops, the optimal dimensionality starts to rise sharply. We have shown this asymptotic relationship in two sets of real data: analysis of multi-subject groups from a study of cognitive impairment associated with aging (Grady et al., 2006), and within-subject analysis of a longitudinal stroke study (Small et al., 2002).

This result has a parallel in statistical physics: in order to capture the structure of the data, the ratio of number of observations to the number of dimensions has to reach a certain critical level. A phase transition happens at this point, and, if this level has not been reached, it is impossible to identify the signal-carrying components in the noisy data although good signal detection is still possible under some circumstances (Watkin and Nadal, 1994; see also Results and Discussion below).

Materials and methods

Simulated data

In this study we used computer-generated data to simulate a block-design experiment with two conditions: *activation* and *baseline* (refer to Lukic et al. (2002) for details of the simulation). All images contained the same simplified single-slice "brain-like" background structure with additive Gaussian noise. An elliptical background structure contained in a 60×60 pixel image consisted of "grey matter" in the center and on the rim of the phantom, and "white matter" in between; the amplitude of the background signal in the "grey matter" was 4 times higher than in the "white matter". Gaussian noise was spatially smoothed using a Gaussian filter with full-width-at-half-maximum (FWHM) of 2 pixels. After smoothing, the standard deviation of the noise was 5% of the background signal. Images in the "activation" condition contained 16 Gaussian-shaped signal "blobs" distributed over the image (12 in the "grey matter" and 4 in the "white matter") and added to the smoothed noisy background