



Contents lists available at ScienceDirect

NeuroImage

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

Q1 Sample-poor estimation of order and common signal subspace with 2 application to fusion of medical imaging data

Q2 Yuri Levin-Schwartz ^{a,*}, Yang Song ^b, Peter J. Schreier ^b, Vince D. Calhoun ^{c,d}, Tülay Adal ^a

^a Department of Computer Science and Electrical Engineering, University of Maryland Baltimore County, Baltimore, MD 21250, USA

^b Signal and System Theory Group, Universität Paderborn, Germany

^c The Mind Research Network, Albuquerque, NM 87106, USA

^d Department of Electrical and Computer Engineering, University of New Mexico, Albuquerque, NM 87131, USA

9 ARTICLE INFO

10 Article history:

11 Received 21 July 2015

12 Revised 8 March 2016

13 Accepted 23 March 2016

14 Available online xxxx

15 Keywords:

16 Multimodal fusion

17 PCA-CCA

18 fMRI

19 sMRI

20 EEG

21 Schizophrenia

ABSTRACT

Due to their data-driven nature, multivariate methods such as canonical correlation analysis (CCA) have proven very useful for fusion of multimodal neurological data. However, being able to determine the degree of similarity between datasets and appropriate order selection are crucial to the success of such techniques. The standard methods for calculating the order of multimodal data focus only on sources with the greatest individual energy and ignore relations across datasets. Additionally, these techniques as well as the most widely-used methods for determining the degree of similarity between datasets assume sufficient sample support and are not effective in the sample-poor regime. In this paper, we propose to jointly estimate the degree of similarity between datasets and their order when few samples are present using principal component analysis and canonical correlation analysis (PCA-CCA). By considering these two problems simultaneously, we are able to minimize the assumptions placed on the data and achieve superior performance in the sample-poor regime compared to traditional techniques. We apply PCA-CCA to the pairwise combinations of functional magnetic resonance imaging (fMRI), structural magnetic resonance imaging (sMRI), and electroencephalogram (EEG) data drawn from patients with schizophrenia and healthy controls while performing an auditory oddball task. The PCA-CCA results indicate that the fMRI and sMRI datasets are the most similar, whereas the sMRI and EEG datasets share the least similarity. We also demonstrate that the degree of similarity obtained by PCA-CCA is highly predictive of the degree of significance found for components generated using CCA.

© 2016 Published by Elsevier Inc.

1. Introduction

The collection of data from multiple modalities has become common in neurological studies, since different modalities are expected to provide complementary views of complicated systems, such as the study of brain activity (James & Dasarathy, 2014). Thus, full utilization of all common information forms the fundamental goal of performing a joint analysis on multimodal data. Since little is known about the intermodality relationships, it is important to minimize the underlying assumptions placed on the data and let the data “speak for itself.” Because of this fact and their ability to treat separate modalities in a symmetric manner, multivariate data-driven methods have proven to be quite popular for the fusion of multimodal neurological data, see e.g., (James & Dasarathy, 2014; Calhoun & Adal, 2009; Adal et al., 2015). To this end, canonical correlation analysis (CCA), which maximizes the correlation of sources across datasets (Hotelling, 1936), has proven to be an effective multivariate and data-driven fusion method, see e.g., (Adal

et al., 2015; Correa et al., 2008; Sui et al., 2010; Chen et al., 2014). However, if the covariances are unknown and must be estimated from the samples, then CCA requires sufficient sample support (Pezeshki et al., 2004). This is particularly an issue when performing multimodal data fusion, since the number of samples, *i.e.* subjects, is typically much less than the dimension of the neurological data that is used. Thus, special attention must be paid both before performing an analysis, *i.e.*, when determining the similarity between datasets and their order—the dimension of the signal subspace—and while performing the analysis itself.

In this paper, we define the similarity between two datasets as the number of common components that both datasets share, *i.e.*, those components that are correlated across datasets, raising the issue of how to determine this number when the sample size is limited. One of the most popular exploratory techniques to estimate the number of common components between two datasets is based on the canonical correlation coefficients (CCCs) calculated using CCA (Hotelling, 1936) and defining a threshold for the level of the correlation, see e.g., (Hoefs, 1983; Bush et al., 1986; Kennedy et al., 1990; Lin et al., 2006). Other methods for estimating the number of common and distinctive sources include: orthogonal *n*-way partial least squares

* Corresponding author.

E-mail address: ylevins1@umbc.edu (Y. Levin-Schwartz).

(OnPLS) (Löfstedt & Trygg, 2011), generalized singular value decomposition (Alter et al., 2003), and distinctive and common components with simultaneous-component analysis (DISCO-SCA) (van Deun et al., 2013). These methods all assume sufficient sample support and thus perform poorly when the number of samples is not significantly greater than the number of observations. CCA, in particular, suffers greatly in the sample-poor regime, where all CCCs are significantly misestimated (Song et al., 2015) and the highest CCCs, usually of greatest interest, may saturate at 1 (Pezeshki et al., 2004), meaning that they provide no information about the true relationship between the datasets.

Since multimodal data is often quite noisy and of high dimensionality, dimension reduction using principal component analysis (PCA) is a crucial preprocessing step for avoiding the problem of overfitting in subsequent analyses. However, the effectiveness of PCA is intimately tied to the problem of order selection. For a single dataset, the most popular order selection methods define the order based on information theoretic criteria (ITC) (Wax & Kailath, 1985), i.e., by using a function of the estimated eigenvalues of the data and the number of model parameters. These methods include: Akaike's information criterion (AIC) (Akaike, 1973), minimum description length (MDL) (Rissanen, 1978) or Bayesian information criterion (BIC) (Schwarz, 1978), and extensions of those methods, see e.g., (Li et al., 2011). Though these methods have found widespread application in multimodal fusion, they are not directly applicable for two major reasons. The first is that almost all of these eigenvalue-based methods, with the notable exception of (Nadakuditi & Edelman, 2008), assume sufficient sample support. If this is not true, such as for multimodal fusion using CCA, where the number of subjects is much less than the dimension of the data, the performance of these methods deteriorates rapidly because the eigenvalues cannot be estimated accurately (Nadakuditi & Edelman, 2008). Additionally, these methods only report on the sources that have greatest energy in each dataset individually. Since we are interested in common components that are linked across datasets, the use of methods that focus solely on a single dataset is not a desirable solution to the question of order selection for multimodal fusion. This provides the incentive to consider the problems of determining the degree of similarity and order jointly. Though not used in the context of medical imaging, there are methods that consider these two problems jointly, see e.g., (Zwack & Velicer, 1986; Hwang et al., 2013), however these techniques are heuristic and will fail in the sample-poor regime (Roseveare & Schreier, 2015).

In this paper, we discuss an effective method, PCA and CCA (PCA-CCA) along with the order selection rule from (Song et al., 2015), for jointly determining the number of common sources for datasets and their order, in the sample-poor regime and demonstrate its importance as a preliminary step for multimodal fusion. To the best of our knowledge, this method is the only one that addresses the issues of common source detection and order selection for the sample-poor case encountered when using CCA for multimodal fusion. The versatility and high performance of this technique are first demonstrated through simulations. We then apply this new method to the pairwise combinations of functional magnetic resonance imaging (fMRI), structural magnetic resonance imaging (sMRI), and electroencephalogram (EEG) data drawn from 14 patients with schizophrenia and 22 healthy controls performing an auditory oddball (AOD) task and relate these results to the pairwise fusion results obtained using CCA. Through this application, we demonstrate a strong correlation between the number of common components estimated using PCA-CCA, i.e., the similarity between datasets, and number of statistically significant components estimated during the fusion analysis. This technique of investigating the pairwise combinations of datasets drawn from the same subjects provides unique insight into the degree of complementarity between related data of different modalities.

2. Materials and methods

2.1. Theory

2.1.1. Traditional and sample-poor hypothesis test

Let us assume that we have M independent and identically distributed (i.i.d.) paired samples of $x^{[1]} \in \mathbb{R}^n$ and $x^{[2]} \in \mathbb{R}^m$ from the two-channel measurement model (Song et al., 2015),

$$\begin{aligned} x^{[1]} &= A^{[1]}s^{[1]} + n^{[1]} \\ x^{[2]} &= A^{[2]}s^{[2]} + n^{[2]}, \end{aligned} \tag{1}$$

where $s^{[k]} \in \mathbb{R}^{d+f}$, $k = 1, 2$, are zero-mean jointly Gaussian random vectors with cross-covariance matrix, $R_{s_1s_2} = E\{s^{[1]}(s^{[2]})^T\}$, given by

$$R_{s_1s_2} = \begin{bmatrix} \text{diag}(\rho_1\sigma_1^{[1]}\sigma_1^{[2]}, \dots, \rho_d\sigma_d^{[1]}\sigma_d^{[2]}) & 0_{d \times f} \\ 0_{f \times d} & 0_{f \times f} \end{bmatrix},$$

where $\sigma_i^{[k]}$ is the unknown standard deviation of signal component $s_i^{[k]}$ and ρ_i is the correlation coefficient between $s_i^{[1]}$ and $s_i^{[2]}$. Thus, both $s^{[1]}$ and $s^{[2]}$ have d correlated signals and f uncorrelated signals. Without loss of generality, we assume the auto-covariance matrices of $s^{[1]}$ and $s^{[2]}$ to be diagonal. The noise terms $n^{[1]}$ and $n^{[2]}$ are independent of each other, independent of the signals, and zero-mean Gaussian with unknown covariance matrices. Additionally, without loss of generality we assume that $A^{[1]}$ and $A^{[2]}$ are of full column rank and, like the dimensions d and f , are fixed but unknown.

We collect the M sample pairs into data matrices $X^{[1]} = [x_1^{[1]}, \dots, x_M^{[1]}]$ and $X^{[2]} = [x_1^{[2]}, \dots, x_M^{[2]}]$. When performing CCA when $M < m + n$, at least $m + n - M$ of the sample canonical correlation coefficients, k_i , $i = 1, \dots, q$, $q = \min(m, n)$, will be identically 1 regardless of the values of ρ_i and thus do not provide any information about the relationship between $s^{[1]}$ and $s^{[2]}$ (Pezeshki et al., 2004). Moreover, even in the case where M is greater, but not significantly greater, than $m + n$, the sample canonical correlations may significantly overestimate the population canonical correlations (Song et al., 2015). This result provides the incentive to estimate a suitable rank, r , in order to reduce the dimensions of $X^{[1]}$ and $X^{[2]}$, thus allowing accurate estimation of the number of correlated signals.

A classical way of estimating d is by assuming that the sources are drawn from a multivariate Gaussian distribution and applying a sequence of binary hypothesis tests (Bartlett, 1941; Lawley, 1959). The test begins with $s = 0$ and compares the two hypotheses $H_0: d = s$ and $H_1: d > s$. If the null hypothesis is rejected, then s is increased by one and the test is repeated, until either the null hypothesis is not rejected or $s = q$. This test is based on the Bartlett–Lawley statistic (Bartlett, 1941; Lawley, 1959), which is given by

$$C(s) = \left(M - s - \frac{m+n+1}{2} + \sum_{i=1}^s k_i^{-2} \right) \ln \prod_{i=s+1}^q (1 - k_i^2), \tag{2}$$

and is asymptotically distributed under H_0 as χ^2 with $(m-s)(n-s)$ degrees of freedom. The fact that the test statistic is distributed according to the χ^2 distribution enables the determination of a threshold, $T(s)$, to meet a given probability of false alarm, P_{FA} , for the test. A major constraint of the traditional framework is the assumption of sufficient samples, i.e., that the k_i 's are accurate estimates of the true k_i 's, making it not applicable for the sample-poor regime.

As proposed in (Song et al., 2015), the sample-poor version of the classical hypothesis test selects

$$d = \max_{r \in \mathbb{R}} \min_{s \in \mathbb{S}} \{s : C(s, r) < T(s, r)\}, \tag{3}$$

Download English Version:

<https://daneshyari.com/en/article/6023349>

Download Persian Version:

<https://daneshyari.com/article/6023349>

[Daneshyari.com](https://daneshyari.com)