Application of Discrminant Principal Component Analysis to Distinguish Schizophrenic Subjects from Normal Controls Based on Fractional Anisotropy Measurements.

A. Caprihan[1], G. D. Pearlson[2,3], and V. D. Calhoun[1,4]

[1]The MIND Institute, Albuquerque, New Mexico, United States, [2]Olin Neuropsychiatry Research Center, Hartford, Connecticut, United States, [3]Deparment of Psychiatry, Yale University, United States, [4]Electrical Engineering, University of New Mexico, Albuquerque, New Mexico, United States

Introduction

Principal component analysis (PCA) is widely used for exploratory data analysis to reduce data dimensionality before applying other methods such as independent component analysis or non-linear pattern recognition. If the goal of a study is to discriminate between two or more groups, then applying standard PCA can undesirably eliminate features that discriminate and primarily keep features that best represent both groups. An alternative method for selecting features has been proposed by Chang (1) which maximizes the Mahalanobis distance between two groups. We apply this method, the discriminatory PCA (DPCA), to a fractional anisotropy (FA) data set from schizophrenia (SZ) subjects and age-matched healthy controls (HC).

Theory Let $X_1 = (x_{11}, x_{12}, ..., x_{1n1})$, and $X_2 = (x_{21}, x_{22}, ..., x_{2n2})$ be samples from two groups with sample means $m_i = (\sum_{j=1}^{n_i} x_{ij}) / n_i$, $i = 1, 2$, the pooled mean $m = (n_1 m_1 + n_2 m_2) / n$, with $n = n_1 + n_2$, and the pooled covariance matrix $R = (\sum_{i=1}^{2} \sum_{j=1}^{n_i} (x_{ij} - m)(x_{ij} - m)^T) / n$. The Mahalanobis distance is $\Delta^2 = (m_1 - m_2)^T W^{-1} (m_1 - m_2)$, where the mean group covariance matrix $W = (\sum_{i=1}^{2} \sum_{j=1}^{n_i} (x_{ij} - m_i)(x_{ij} - m_i)^T) / n$. A large value of Δ is an indicator of greater separation between the two groups. In the standard PCA for dimensionality reduction we choose the q eigenvectors of *R* correspond to its *q* largest eigenvalues. This selection does not necessarily maximize Δ. Chang has shown that if we order the eigenvectors of *R* according to $(e_k^T (m_1 - m_2))^2 / \lambda_k$, where e_k is the eigenvector of *R* corresponding to the eigenvalue λ_k, and choose the corresponding first *q* eigenvectors then Δ is maximized for that q-dimensional subspace.

Method and Subjects

Data from 45 SZ subjects and 45 HC was obtained on a 3.0T Siemens Allegra scanner at the Olin Center, Institute of Living, with a 12 directions standard Siemens sequence with b=1000 s/mm^2. Other parameters were: FOV=200mm, slice thickness =3mm, 45 slices, TE=83 ms, and TR=5900 ms. Data processing consisted of eddy current correction (FSL), calculating FA (dtifit, FSL), registering the images to a FA template (tbss, FSL). The spatially normalized output was masked such that FA > 0.05 across all subjects for each voxel to include only voxels with non-zero anisotropy, and written as a 149206x90 matrix, where each column is the image. The eigenvectors were calculated for the sample covariance matrix *R,* then ordered according to the standard PCA, and the proposed DPCA. The error for approximating to a lower dimension subspace and the Δ was calculated.

We tested the predictive capability of FA to distinguish the two groups for the reduced data sets by the leave-one out and Fisher's linear discriminant (FLD) classification method (2). The discriminant function being $w = W^{-1}(m_1 - m_2)$. We sequentially left 1 of the subjects out of processing, selected a certain number of components and classified the subject as SZ or NC and counted the errors. In order to reduce the effect of subject variability and registration errors on classification we smoothed the data by Gaussian filters of widths 4mm, 6mm, and 8mm and the performance evaluated by the leave-one out method. Finally, to identify anatomic regions contributing more towards discrimination, we used the Johns Hopkins University (JHU) white matter tract atlas (3), now part of FSL. The FLD function used for classification was correlated with the map of the 20 tracts in the JHU atlas.

Results and Discussion

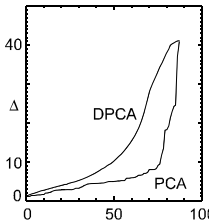

Fig. 1: No. of Components

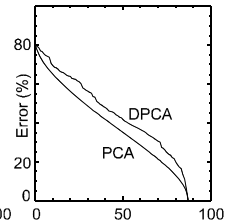

Fig. 2: No. of Components

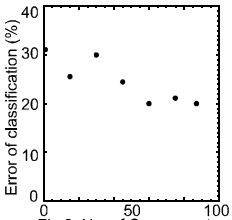

Fig 3: No. of Components

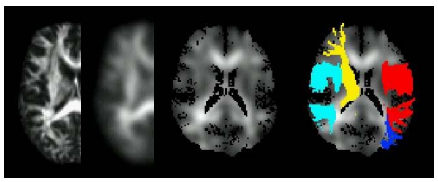

Fig. 4: Processing Pipeline

Initial analysis showed that smoothing with 6mm width Gaussian filter resulted in the least classification error and the following results are for this case. Figure 1 compares the Mahalanobis distance and Fig. 2 the approximation error between standard PCA & DPCA. As expected, when all components are chosen both methods are equivalent and by chance the first component is the same for both methods. The distance Δ is like the t-statistic for the difference between the two groups. With 50 components the distance between the two groups is twice as large with DPCA then with PCA. As expected the approximation error is higher with DPCA (Fig. 2). If the FLD function is calculated from all the samples then the two groups can be perfectly separated with 30 components. The predictive performance of FLD classification is seen by the leave-one out method in Fig. 3. Even with all the components selected we have 20% error, rising to 30% when only one component is selected. In this example after 60 components are selected there is no further reduction in error. We reduced a 149206x90 matrix to a 60x90 matrix while maintaining similar classification error as the total data set. The back projection of the 60 component image is shown in Fig. 4C. We calculated the contribution of 20 different white matter tracts to discrimination . The four most important components were superior longitudinal fasciculus (SLF)-left (red) and right (light blue), inferior LF-left (dark blue), and anterior thalamic radiation-right (yellow).

Conclusions

We have shown that the proposed method and the results are important when large data sets from different modalities are collected and it is important to retain group differences in the exploratory dimensionality reduction step. In our example data set, we were able to reduce the image dimension to 60 without any significant loss in classification performance. The SLF in both the left and the right side of the brain had the maximum contributions towards discrimination between schizophrenia subjects and healthy controls.

Reference 1)W-C. Chang. Appl. Statist. 1983. 32:267-275. 2) Duda, Hart, and Stark. Pattern Class. 2001. 3) www.fmrib.ox.ac.uk/fsl/fslview/atlas

Acknowledgment This research was supported by the NIH (1R01EB006841, 1R01EB005846) and NSF (0612076).