Whole Brain Voxel-based Analysis Using Registration and Multivariate Statistics

C.E. Thomaz[1], J.P. Boardman[2], D.L.G. Hill[3], J.V. Hajnal[2],
A.D. Edwards[4], M.A. Rutherford[2], D.F. Gillies[1], and D. Rueckert[1]

[1]Department of Computing, Imperial College, London, UK
[2]Imaging Sciences Department, Imperial College, London, UK
[3]Division of Imaging Sciences, King's College, London, UK
[4]Division of Paediatrics, Obstetrics and Gynaecology, Imperial College, London, UK

1 Introduction

Whole brain voxel-based morphometry and statistical pattern recognition methods have been used to classify and describe anatomical structures of MR images. Most of these methods are based on statistical learning techniques applied to either segmented images or a number of features pre-selected from specific image decomposition approaches. Although such pre-processing strategies have overcome the difficulty of dealing with the inherent high dimensionality of 3D brain image data, most of these approaches rely on optimisation techniques that are time consuming and do not provide a simple way of mapping the classification results back into the original image domain for further interpretation.

In this paper, we use the general multivariate statistical methodology (PCA+LDA) to identify the most discriminating hyper-plane separating two populations. We introduce some novel techniques to overcome the well-known instability of the LDA within-class scatter matrix and increase the computational efficiency of the approach. Our goal is to analyse all the data simultaneously rather than feature by feature. The result is an efficient and practical method for separating two populations and visually analysing their differences.

2 Methodology

Before we can analyse the MR images we need to map all images into a common atlas coordinate system. This pre-processing step is essential because the construction of the multivariate statistical model relies on anatomical correspondences when comparing patterns across subjects. We have randomly chosen the image of one subject as reference or atlas. In order to map the anatomy of each subject into the anatomy of the atlas we have first applied an affine registration [1] followed by non-rigid registration based on free-form deformations [2]. Both algorithms are based on the maximisation of normalised mutual information as a voxel-based similarity measure.

2.1 PCA

After registration, the Principal Components Analysis (PCA) technique is performed. PCA is a feature extraction procedure concerned with explaining the covariance structure of a set of variables through a small number of linear combinations of these variables. It is a common statistical technique that has been used in several image recognition problems, especially for dimensionality reduction.

Although there is always the question of how many principal components to retain in order to reduce the dimensionality of the original training sample, Yang and Yang [3] have proved recently that the number of principal components to retain for a best LDA classification performance should be equal to the rank m of the total covariance matrix S composed of all the training patterns and given by

$$S = \frac{1}{(N-1)} \sum_{j=1}^{N} (x_{i,j} - \bar{x})(x_{i,j} - \bar{x})^T , \qquad \textbf{(1)}$$

where $x_{i,j}$ is the n-dimensional pattern j from class π_i, N is the total number of samples, and $\bar{x}$ is the grand mean vector given by

$$\bar{x} = \frac{1}{N}\sum_{j=1}^{N} x_{i,j} \ . \quad \textbf{(2)}$$

The m principal components can then replace the initial n features and the original data set, consisting of N measurements on n variables, is reduced to a data set consisting of N measurements on m principal components. For this representation to make sense in statistical classification problems we are making the assumption that the distributions of each class or group are separated by their corresponding mean differences.

2.2 LDA

The primary purpose of Linear Discriminant Analysis (LDA) is to separate samples of distinct groups by maximising their between-class separability while minimising their within-class variability. Let the between-class scatter matrix S_b be defined as

$$S_b = \sum_{i=1}^{g} N_i (\bar{x}_i - \bar{x})(\bar{x}_i - \bar{x})^T \ , \quad \textbf{(3)}$$

and the within-class scatter matrix S_w be defined as

$$S_w = \sum_{i=1}^{g} (N_i - 1) S_i = \sum_{i=1}^{g}\sum_{j=1}^{N_i} (x_{i,j} - \bar{x}_i)(x_{i,j} - \bar{x}_i)^T \ , \quad \textbf{(4)}$$

where $x_{i,j}$ is the m-dimensional pattern j from class π_i, N_i is the number of training patterns from class π_i, g is the total number of classes or groups, and $\bar{x}$ is the grand mean vector defined in equation (2). The vector $\bar{x}_i$ and matrix S_i are respectively the unbiased sample mean and sample covariance matrix of class π_i.

The main objective of LDA is to find a projection matrix P_{lda} that maximises the ratio of the determinant of the between-class scatter matrix to the determinant of the within-class scatter matrix (Fisher's criterion), that is

$$P_{lda} = \arg\max_{P} \frac{\left| P^T S_b P \right|}{\left| P^T S_w P \right|} \ . \quad \textbf{(5)}$$

It is a proven result [4] that if S_w is a non-singular matrix then the Fisher's criterion is maximised when the projection matrix P_{lda} is composed of the eigenvectors of $S_w^{-1} S_b$ with at most $(g-1)$ nonzero corresponding eigenvalues. This is the standard LDA procedure.

However, the performance of the standard LDA can be seriously degraded if there are only a limited number of total training observations N compared to the dimension of the feature space m. Since the within-class scatter matrix S_w is a function of $(N-g)$ or less linearly independent vectors, its rank is $(N-g)$ or less. Therefore in the problem under investigation where the number of training patterns is comparable to the number of features, S_w might be singular or mathematically unstable and the standard LDA cannot be used to perform the task of the classification stage.

2.3 MLDA

In order to avoid both the singularity and instability critical issues of the within-class scatter matrix S_w when LDA is used in such limited sample and high dimensional problem, we have proposed a maximum uncertainty LDA-based approach (MLDA) to overcome the instability of the S_w matrix [5]. It is based on the maximum entropy covariance selection method developed to improve quadratic classification performance on limited sample size problems [6].

The proposed method considers the issue of stabilising the S_w estimate with a multiple of the identity matrix by selecting the largest dispersions regarding the S_w average eigenvalue. The following selection algorithm expands only the smaller and consequently less reliable eigenvalues of within-class scatter matrix S_w:

i. Find the Φ eigenvectors and Λ eigenvalues of S_p, where $S_p = S_w / [N-g]$;

ii. Calculate the S_p average eigenvalue $\overline{\lambda}$ using

$$\overline{\lambda} = \frac{1}{m}\sum_{j=1}^{m}\lambda_j = \frac{tr(S_p)}{m};$$

iii. Form a new matrix of eigenvalues based on the following largest dispersion values

$$\Lambda^* = diag[\max(\lambda_1, \overline{\lambda}), \max(\lambda_2, \overline{\lambda}), ..., \max(\lambda_m, \overline{\lambda})];$$

iv. Form the modified within-class scatter matrix

$$S_w^* = S_p^*(N-g) = (\Phi\Lambda^*\Phi^T)(N-g).$$

The maximum uncertainty LDA is constructed by replacing S_w with S_w^* in the standard Fisher's criterion formula described in equation (5). It is a straightforward method that overcomes both the singularity and instability of the within-class scatter matrix S_w when LDA is used in limited sample and high dimensional problems.

3 Experiments

To demonstrate the effectiveness of the approach, we have used a neonatal MR brain data set that contains 67 preterm infants at term equivalent age (mean 29.7, range 24-34 weeks post-menstrual age), and 12 term born controls (mean 39.3, range 36-42 weeks post-menstrual age). Ethical permission for this study was granted by the Hammersmith Hospital Research Ethics Committee and informed parental consent was obtained for each infant. Infants were sedated for the examination but did not require mechanical ventilation at the time of MR imaging. Pulse oximetry, electrocardiographic and televisual monitoring were used throughout the examination which was attended by a paediatrician. A 1.5 T Eclipse MR System (Philips Medical Systems, Cleveland, Ohio) was used to acquire high resolution T1 weighted images (TR=30ms, TE=4.5ms, flip angle = 30o). In addition to conventional T1 and T2 weighted image acquisition, volume datasets were acquired in contiguous sagittal slices (in-plane matrix size 256 x 256, FOV = 25cm) with a voxel size of 1.0 x 1.0 x 1.6 mm3.

We have performed two main tasks: classification and visual analysis. First a training matrix composed of N zero mean *n*-dimensional image vectors is used as input to compute the PCA transformation matrix. The columns of this *n* x *m* transformation matrix are eigenvectors, in eigenvalues descending order. The N zero mean image vectors are projected on the principal components and reduced to *m*-dimensional vectors representing the most expressive features of each one of the pre-processed *n*-dimensional image vector. Afterwards, this *N* x *m* data matrix is used as input to calculate the MLDA discriminant eigenvector. The most discriminant feature of each one of the *m*-dimensional vectors is obtained by multiplying the *N* x *m* most expressive features matrix by the MLDA linear discriminant eigenvector. An analogous procedure, but in reverse order, has been used to convert any point on the most discriminant space back to its corresponding *n*-dimensional image vector. More specifically, first we multiply that particular point by the transpose of the linear discriminant vector previously computed, then we multiply its *m* most expressive features by the transpose of the principal components matrix, and finally we add the average image calculated in the training stage to the *n*-dimensional image vector.

4 Results

Figure 1 presents the leave-one-out recognition rate (rr) of the two-stage linear classifier using the affine and non-rigid registration algorithms as pre-processing techniques. As expected, the classification results obtained by the non-rigid registration algorithms are higher than the one obtained by an affine transformation, achieving a maximum recognition rate of 97.47% with a control point spacing of 5mm.

Figure 2 highlights the statistical differences between the preterm infants (shown on the top) at term equivalent age and the control group (bottom) mapped back (without the mean) into the image domain. We can see clearly differences in the ventricular system, the posterior limb of the internal capsule, the corpus callosum area, and the inter-hemispheric fissure.

Registration	rr (%)
Affine	91.14
Non-rigid (10mm)	96.20
Non-rigid (5mm)	97.47
Non-rigid (2.5mm)	93.67

Figure 1. Classification results.

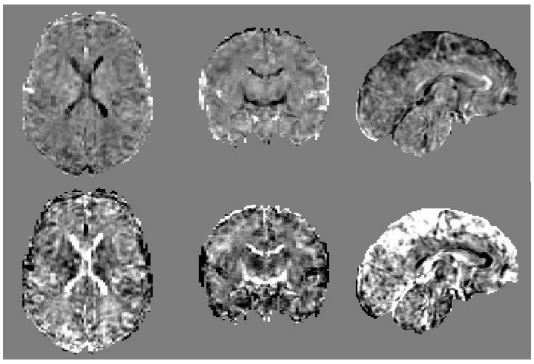

Figure 2. Visual statistical differences.

5 Conclusion

This paper describes the idea of using PCA plus the maximum uncertainty LDA-based approach to classify and analyse MR brain images. The methodology proposed has been performed directly on the MR intensity images rather than on segmented versions of the images. Our results indicate that the use of non-rigid registration in the pre-processing step and the two-stage linear classifier make clear the statistical differences between the control and preterm neonatal samples, showing a classification accuracy of 97.47% using the leave-one-out method.

Although the experiments carried out were based on a specific preterm infants database, we believe that such multivariate statistical strategy for targeting limited sample and high dimensional problems provides a suitable framework for characterising and analysing the high complexity of MR images in general.

Acknowledgements

This work is part of the UK EPSRC e-science project “Information eXtraction from Images” (IXI). Also, the first author was partially supported by the Brazilian Government Agency CAPES under Grant 1168/99-1.

References

1. C. Studholme, D. L. G. Hill and D. J. Hawkes. “An overlap invariant entropy measure of 3D medical image alignment”, *Pattern Recognition*, vol. 32, no. 1, pp 71-86, 1999.
2. D. Rueckert, L. I. Sonoda, C. Hayes, D. L. G. Hill, M. O. Leach, and D. J. Hawkes. “Non-rigid registration using free-form deformations:Application to breast MR images”, *IEEE Transactions on Medical Imaging*, vol. 18, no. 8, pp. 712-721, 1999.
3. J. Yang and J. Yang, “Why can LDA be performed in PCA transformed space?”, *Pattern Recognition*, vol. 36, pp. 563-566, 2003.
4. K. Fukunaga, *Introduction to Statistical Pattern Recognition*, second edition. Boston: Academic Press, 1990.
5. C. E. Thomaz. “Maximum Entropy Covariance Estimate for Statistical Pattern Recognition”, *PhD Thesis*, Department of Computing, Imperial College London, 2004.
6. C. E. Thomaz, D. F. Gillies and R. Q. Feitosa. "A New Covariance Estimate for Bayesian Classifiers in Biometric Recognition", *IEEE Transactions on Circuits and Systems for Video Technology, Special Issue on Image- and Video-Based Biometrics*, vol. 14, no. 2, pp. 214-223, February 2004.