Machine learning in fMRI

Feature Extraction

Alexandre Savio, Maite Termenón, Manuel Graña

[1]Computational Intelligence Group, University of the Basque Country

December, 2010

Outline

Outline

The feature extraction process I

- Feature extraction is a special form of dimensionality reduction.
- There are common algorithms for dimensionality reduction which can be applied to reduce the dimensionality of the data:
  - Principal component analysis (PCA)
  - Independent Component Analysis (ICA)
- Feature extraction methods can be more specific to the type of data we are analysing.
  - The meaning of the data is implicit to the feature extraction method.

The feature extraction process in fMRI

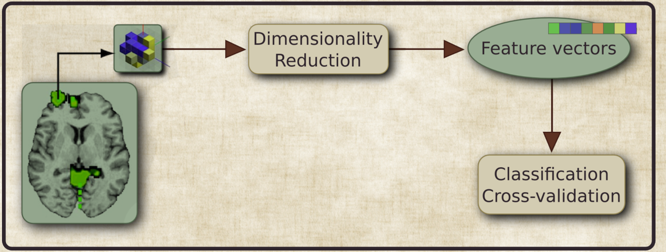

Figure:

The feature extraction process in fMRI I

- The number of possible feature sets we can extract from fMRI acquisitions of a determined group is very large.

The feature extraction process in fMRI II

- It can depend on:
  - Experimental design
  - Number of subjects
  - Type of experiment we want to perform
  - The techniques we can use
  - The techniques used in the literature for similar situations

The feature extraction process in fMRI III

- Group normalization is an issue in fMRI.
  - Finding a feature set with the same meaning and the same size for all the subjects in our data set is not an easy task.
  - In addition, the algorithm should be able to extract the same features from new unseen subjects.

Examples I

- Dimension reduction and feature extraction using ICA[1]

Examples II

- Use average intensity in multiple TRs [2]
    - a drawback of this method is a reduction in the number of samples available for training.

Examples III

- [3] At each stimulus presentation, a trial $t\,(t = 1, \ldots, T)$ is formed considering N_{pre} and N_{post} temporal samples (before and after stimulus onset respectively) of the pre-processed time course of activity.
- A trial estimate of the response at every voxel $v\,(v = 1, \ldots, V)$ is then obtained by fitting a General Linear Model (GLM) with one predictor coding for the trial response and one linear predictor accounting for a within-trial linear trend.
- The trial-response predictor is obtained by convolution of a boxcar with a double-gamma hemodynamic response function (HRF)

Examples IV

- Firstly, let S and R be the sets of selected features and the group of features that might be chosen: we start with $S = \emptyset$ and $R = \{x_i\}, i = 1..N$ and the algorithm will stop when R is empty.
- This algorithm uses an hybrid stepwise selection.
  - The forward strategy adds at each step the most informative feature given the previously selected ones.
  - The backward strategy removes from R all the features which are not informative at this step: we indeed assume that those features will not be informative in the next steps.
- In order to select a feature, we compute at each step, for each dimension x in R, the value $MI1 = MI(S\{x\}, Y)$, which yields the amount of information about Y present in S and x. [4]

Examples V

- To break the complexity of the problem, we first perform a hierarchical clustering of the voxel-based signals, under connectivity constraints, so that only spatially connected clusters are created.
- At that stage, we ignore the target information, but use the variance-minimizing approach of Ward's algorithm [12] in order to ensure that cluster-based averages provide a fair representation of the signal within each cluster. Only adjacent clusters can be merged together.
    - The purpose of this procedure is to use the hierarchical parcellation to guide the search of informative regions within the volume of interest.
- Thus, at a given level in the hierarchy, the data is reduced to NC cluster-based averages, which significantly decreases the computational complexity compared to a voxel-based approach with $N_v \gg NC$ voxels. [5]

Examples VI

- Thus, in order to further reduce the dimensionality of the data, we parcellate this region in 200 parcels with a variant of Ward's algorithm, and we average the signal within each parcels.[6]

Examples VII

- We used PCA to find the bases of reduced dimensionality.
- In the present work, we did not exclude any PC in the analysis, that is, the PCA step is loss–less dimension reduction and represents only a change of the coordinate system to the subspace spanned by the measured brain volumes. [7]

Examples VIII

- After realignment of the functional volumes using SPM5,1 we use the IBASPM toolbox (Tzourio-Mazoyer et al., 2002; Alemán-Gómez et al., 2006) to build an individual brain atlas based on the structural MRI, containing $M = 90$ anatomical regions.
- While this is a relatively coarse atlas, it is an essential step to allow for inter-subject variability and enable inter-subject decoding with good generalisation ability to unseen subjects — using group-level normalisation and atlasing is not an option in this setting.
- Furthermore, the structural atlas serves only as a basis for computing a much lower resolution functional atlas. Using a more fine-grained atlas might result in some regions disappearing completely in the functional atlas.

Examples IX

- Another benefit of using the AAL atlas is that it offers a way of comparing results with several other studies [8]

Summary

- Feature extraction methods is a special form of dimensionality reduction.
- In fMRI there are many different algorithms for feature extraction in the literature.
- The difficulty of a good feature extraction method lies on finding:
  - Common features for all the subjects in the data set (due to spatial normalization problems)
  - The best fit to the experimental design and classification objective of our experiment.

References I

P.K. Douglas, Sam Harris, Alan Yuille, and Mark S. Cohen.
Performance comparison of machine learning algorithms and number of independent components used in fMRI decoding of belief vs. disbelief.
NeuroImage, In Press, Uncorrected Proof.

Elia Formisano, Federico De Martino, and Giancarlo Valente.
Multivariate analysis of fMRI time series: classification and regression of brain responses using machine learning.
Magnetic Resonance Imaging, 26(7):921–934, September 2008.

References II

Federico De Martino, Giancarlo Valente, Noël Staeren, John Ashburner, Rainer Goebel, and Elia Formisano.
Combining multivariate voxel selection and support vector machines for mapping and classification of fMRI spatial patterns.
NeuroImage, 43(1):44–58, October 2008.

V. Michel, C. Damon, and B. Thirion.
Mutual information-based feature selection enhances fMRI brain activity classification.
In *Biomedical Imaging: From Nano to Macro, 2008. ISBI 2008. 5th IEEE International Symposium on*, pages 592–595, 2008.

References III

V. Michel, E. Eger, C. Keribin, J. Poline, and B. Thirion.
A supervised clustering approach for extracting predictive information from brain activation images.
In *Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on*, pages 7–14, 2010.

Vincent Michel, Evelyn Eger, Christine Keribin, and Bertrand Thirion.
Adaptive multi-class bayesian sparse regression - an application to brain activity classification, 2009.

References IV

Janaina Mourão-Miranda, Arun L.W. Bokde, Christine Born, Harald Hampel, and Martin Stetter.
Classifying brain states and determining the discriminating activation patterns: Support vector machine on functional MRI data.
NeuroImage, 28(4):980–995, December 2005.

Jonas Richiardi, Hamdi Eryilmaz, Sophie Schwartz, Patrik Vuilleumier, and Dimitri Van De Ville.
Decoding brain states from fMRI connectivity graphs.
NeuroImage, In Press, Corrected Proof.