Unsupervised classification of whole-brain fMRI data with artificial neural networks

Arnaud Fournel, Emanuelle Reynaud

UNSUPERVISED CLASSIFICATION OF WHOLE BRAIN fMRI DATA WITH ARTIFICIAL NEURAL NETWORKS

Arnaud Fournel[1], Emanuelle Reynaud[1]
1. Laboratoire EMC, EA 3082, Université Lyon 2, Université de Lyon
5 avenue Pierre Mendès-France 69676 Bron cedex - France
arnaud.fournel@univ-lyon2.fr, emanuelle.reynaud@univ-lyon2.fr

ABSTRACT

In the present study, we apply the Self Organizing Map (SOM) algorithm for classifying cognitive states from fMRI data without prior selection of spatial or temporal features. In addition, we compare our method with two other models. We applied the method to single-subject as well as multi-subject classification.
BOLD signals from subjects viewing emotional pictures of positive, neutral and negative valences were acquired during a block design experiment, and classified with an unsupervised non-linear method, the SOM. We demonstrate here that, in terms of classification performance, the SOM algorithm outperforms an SVM algorithm when processing whole brain data, and performs as well as methods (SVM and KCCA) working with temporal compression or spatial feature selection. Our method presents three phases: data dimensionality reduction : where non functional data are deleted, SOM algorithm training : where statistic regularities relevant for classification are extracted, SOM algorithm test : where the subject's brain state is predicted from his brain activity.

KEY WORDS

Machine Learning methods; Kohonen Self Organizing Maps; Multi-Voxel Pattern Analysis; fMRI data analysis; Brain-reading

1. Introduction

The key question in cognitive neuroscience is to bind cognitive states with brain structures.
Functional magnetic resonance imaging (fMRI) is a ubiquitous tool of investigation. The technique allows to measure brain activity in a non-invasive way in response to experimental conditions.
During an fMRI scan, subjects perform active and control tasks. Statistical analysis methods are then applied to identify voxels significantly more activated by a particular task, compared to a control task. Basically, univariate methods are applied on voxels time series. This method often leads to ignore the spatial correlations that may exist in the data (ie the spatial relationship between voxels). For example, Haxby and colleagues [1] have shown that this spatial pattern of brain activity can carry important information on cognitive processing.

In recent years, pattern recognition methods have begun to be used as a multivariate technique for fMRI data analysis[2] [3]. Breaking with the traditional approach that tries to relate tasks to significant voxel responses, multivariate method tries to map an observed fMRI volume onto a pattern of brain activity (ie the cognitive state of a subject at a time *t*). For multivariate methods, fMRI data is treated as spatial patterns, on which machine learning algorithms will try to find statistical relationships underlying the subject's cognitive state (ie the specific pattern of brain activity for a *task*). For a review of machine learning for analyzing neuroimaging data see [4].

Most of the research in the field has used region of interest (ROI) approach or data compression for trying to improve the Signal to Noise Ratio, naturally high in fMRI data. Nevertheless, even if these techniques decrease the total amount of noise in data, they also erase potentially relevant information. Moreover, adopting an ROI approach demands to have strong *a priori* hypotheses about the subsequent brain regions to focus on. Data compression by averaging brain volumes over an experimental block might also mask the potential differences that could exist in brain activity over time.

We propose in this paper to use Self Organizing Map [5] for classifying fMRI data. This artificial neural network is well know and currently used in Machine Learning and Data Mining for exploring high dimensional data. Because our method is based on whole brain functional data analysis, and due to the unsupervised nature of our algorithm, neuroimaging data can be analyzed without any prior assumptions. This can lead to new hypotheses on data organization and therefore on cognitive processing.

2. Material and Methods

2.1- Subjects

FMRI data were acquired on 16 male right-handed healthy US College students (mean : 20,25 years old).
All subjects had a normal vision and didn't have any history of neurological or psychiatric disorders. This experiment was performed in accordance with local Ethics Committee of the University of North Carolina (USA). All participants gave written informed consent to participate in the study after the study was explained to them.

2.2- Acquisition

We measured blood-oxygenation-level-dependent (BOLD) responses using 3 T Allegra Head Only MRI System at the Magnetic Resonance Imaging Research Center at the University of North Carolina. The parameters for scanning were as follow: voxel size = 3x3x3 mm3, TR = 3 s, TE = 30 ms, FA = 80°; FOV = 192x192 mm; matrix = 64x64x49. For each participant, 254 functional volumes were acquired.

2.3- Experimental Design

Data were acquired during a block design, composed of 3 different conditions : passively viewing unpleasant stimuli (photos of dermatological diseases), viewing neutral stimuli (photos of neutral scenes with humans) and viewing pleasant stimuli (nude girls or girls in swimsuits).
There were 6 blocks, each block consisting of 7 images, each presented for 3 seconds. Each block was followed by a "rest" block, with only a fixation cross.

FIGURE 1: show examples of the three images conditions that the subjects show during experiment.

2.4- Pre-Processing

The data were pre-processed with the FSL Software Library (FSL is written by the Analysis Group, FMRIB, Oxford, UK, http://www.fmrib.ox.ac.uk/fsl/).
In single subject condition, all volumes were simply motion corrected (removing subject movements) and detrended (removing scanner drift), to preserve the fine-grained local patterns.
In multi-subject condition, data were motion corrected, detrended, smoothed (FWHM 5mm kernel) and spatially transformed into standard space [6]. In this condition we apply smoothing to minimize small spatial deformations between subjects, due for example to imperfect spatial transformations.

2.5- Classifier

The Self-organizing Map (SOM) is an unsupervised artificial neural network that provides a non-linear mapping from a high-dimensional input space to a low dimensional output space (most often a 2D or 3D output grid). SOM is a popular and widely used method for data mining, pattern recognition and exploratory data analysis [7]. The main property of this algorithm is that it consistently conserves the original topological and metrics relationships of the input space.
This algorithm consists of a set of *i* input units, corresponding to the input data set and a set of *j* output units arranged in a 2D or 3D grid. Each output unit has a weight vector w_i associated.

The SOM algorithm process can be described as follow:

1- An input vector X_i is presented to the neural network .

2- Artificial neurons (output units) compete with each other, the Best Matching Unit (BMU)– ie the closest unit – is the winner. (distance used here is Euclidean distance)

3- The BMU and its neighbours update their values, and are moved towards the input vector.
The update rule for the BMU and its neighbours can be mathematically described as:

$$(1):\ w_k(t+1) = w_k(t) + \alpha(t) h_k(t)\left[x(t) - w_k(t)\right]$$

After updating their weights, the BMU and neighbours are supposed to represent more accurately the input vector.
In (1), $\alpha(t)$ is a decreasing function of time that controls the learning rate in the network and $h_k(t)$ is a function computing the size of the BMU neighbourhood. $\alpha(t)$ and $h_k(t)$ can be controlled by two other parameters : the "learning rate" and "neighbourhood radius". The size of the map (ie the numbers of neurons) is an another parameter. All the parameters described above can influence SOM algorithm results. In this study we use the classic SOM algorithm denoted on-line SOM, given that the weights vectors are updated at each step.

4- This process is repeated until the stopping criteria are met (in this study stopping criteria are: a fixed number of iterations are completed and/or accuracy rate is above a predefined threshold.

3. Classification protocol

Pre-processed BOLD signals emanating from the brain activity at a given time point are 3D volumes of 64*64*49 components. These volumes are transformed into vectors. Our data set consists of 126 vectors per subjects, divided in 3*42 vectors per experimental condition (pleasant/unpleasant/neutral).

The SOM map is then used as a non-linear classification tool through two processing steps.

3.1-The learning phase

Here, statistic regularities relevant for classification are extracted. During this phase we use 80% of the data set (for single subject condition) or 15 out of the 16 subjects (for multi-subjects condition) as the training set. The SOM algorithm adjusts its output map to group the fMRI 3D volumes according to their specific spatial pattern of activity. As this activity is elicited by distinct experimental conditions, we would expect, in a perfect world, volumes to be grouped in three clusters, reflecting our three experimental conditions.

3.2- The testing phase

This second phase challenges the cluster organization made during the learning phase. An input example is presented. For this new vector, the BMU can belong to the same experimental condition as the input vector (correct prediction) or to another (misclassification). Here the input vector is the subject's brain activity. The BMU class represents the predicted cognitive state, inferred from the new fMRI volume.

The test set consists in novel input vectors taken from 20% of the data set (single-subject condition), or from the data of a single subject not used in the training set (leave-one-out procedure for the multi-subject condition).
The SOM performance was measured by the correct prediction rate on test examples.

4. Results

4.1- Single-subject condition.

In this condition, the pre-processing steps described in section 2 are applied to all the fMRI volumes scanned in a single subject performing our passive viewing experiment.
The SOM parameters used are a map size of 6*6 neurons, a learning rate set to 0.1, a neighbourhood radius of 3.
Figure 1 shows the average prediction rate of the 16 single-subjects runs for the three conditions (viewing pleasant, unpleasant, and neutral photos).
The average performance on all conditions is 90.99 % (+- 4.5%) of correct recognition (ie : a brain activity volume is predicted to belong to the correct experimental condition).

It is worth mentioning that if the scans corresponding to the neutral pictures are taken out from the data set, the classification problem is reduced to a two-class problem, and our method reaches 100% of correct predictions in the testing phase.

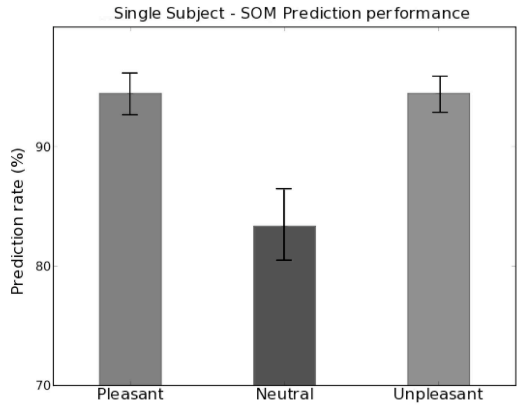

FIGURE 1: Average performance prediction of SOM algorithm for the 16 single subjects.

We can observe that pleasant and unpleasant condition are better recognized than the neutral condition. We will address this point later in the discussion section.

4.2- Multi-subject condition

For multi-subject condition after pre-processing the data (see section 2), we apply the AAL template [8] to remove all non grey matter.
The self organizing map parameters were the same as in the single-subject condition.
As in the single-subject condition, results show that pleasant and unpleasant conditions lead to better performances than the neutral condition (see fig. 2).
After 10 runs in the multi-subject condition, the average prediction rate for the three experimental conditions is 83.33 %.

As in the single subject condition, if the “neutral” scans are pulled out from training and test datasets, performances reach 95% correctness.

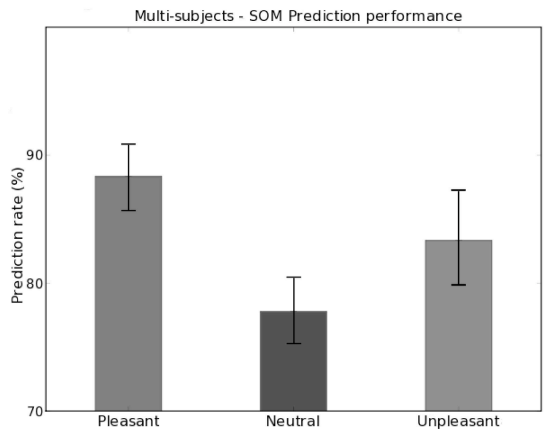

FIGURE 2: Average performance prediction of SOM algorithm for 10 runs in multi subject condition.

4.3- SOM versus two other models

Here, we compare our technique with two others existing models. An supervised model based on Support Vector Machine algorithm (SVM)[9] and an unsupervised algorithm called Kernel Canonical Correlation Analysis (KCCA)[10], that use the temporal compression to increase accuracy rate. The three models use the same data set, and are therefore available for comparison. Figure 3 shows that in single-subject condition, our model has similar performances as KCCA and performs better than the SVM technique. In multi-subjects condition, SVM and SOM reach nearly the same accuracy rate when SVM uses temporal compression, but SOM outperforms SVM when whole-brain data is used for input (see fig 4).
This result shows that achieving high-quality classification is possible even with high-dimension and low signal-to-noise ration data as is whole-brain fMRI data.

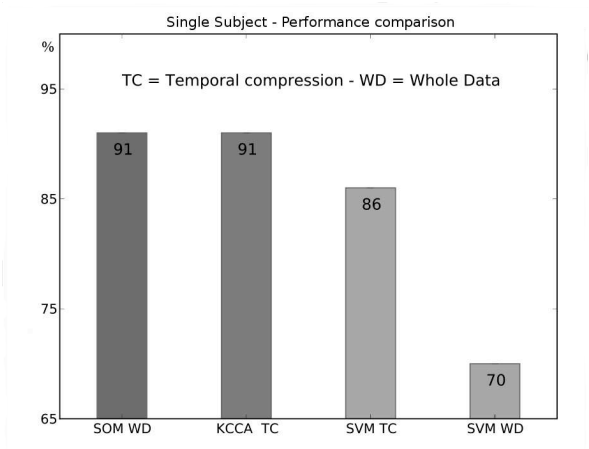

FIGURE 3: Single-subject - Comparison of the 3 models (SVM, KCCA, SOM) prediction performance (TC = Temporal compression, WD = Whole data)

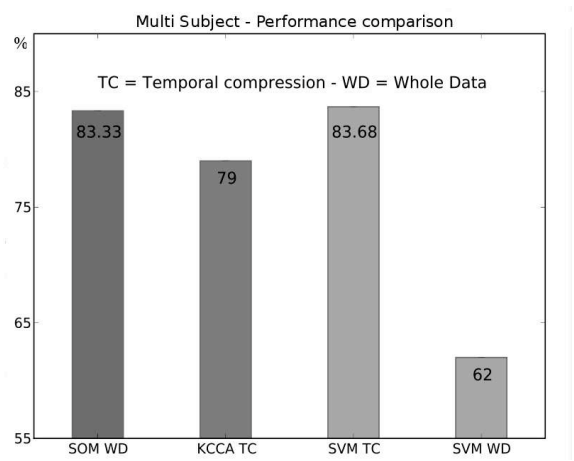

FIGURE 4: Multi-Subject - Comparison of the 3 models (SVM, KCCA, SOM) prediction performance (TC = Temporal compression, WD = Whole data)

5 .Conclusion

Our method reveals the possibility to analyze whole brain data without any prior assumptions. Nevertheless, the next step after this work is to confirm the performance on an another dataset. Moreover, we hope to test the method sensibility (ie what are the finest differences that can be detected with this method).

As can be seen with the results from single subject and multi subject conditions, the neutral condition is less efficiently recognized than the pleasant and unpleasant condition. We think that this can be explained by the nature itself of the neutral stimuli used (Figure 5 shows an example of three neutral condition images presented during the experiment). For example the study lead by Kanwisher [11] has shown that looking at body parts versus looking at faces do not imply the same brain activity patterns. SOM algorithm might pick up these small differences and, as a consequence, it can not view neutral stimuli as an homogeneous stimuli class. This would explain that the algorithm fails to classify perfectly brain activity from neutral condition when mixed with other from pleasant and unpleasant conditions.

FIGURE 5: shows three images of the neutral condition

Our method allows us to observe another element. Data organization do not only reflect the experimental conditions (through clustering behaviour of SOM algorithm) but also the moment of presentation in the experiment seems to impact the classification. The cognitive processes that we observe here seem to evolve over time. This possible temporal nature of cognitive treatments of emotional stimuli is currently under investigation.

Finally, our method allows to study cognitive functions distributed over the whole brain. This could make it a very good candidate for “brain-reading” [12] experiments.

Acknowledgements

The authors warmly thank S.Mazza, B. Lété and H. Paugam for their comments and ideas on this work.

References

[1] J.V. Haxby, M.I. Gobbini, M.L. Furey, a. Ishai, J.L. Schouten, and P. Pietrini, Distributed and overlapping representations of faces and objects in ventral temporal cortex, *Science (New York, N.Y.), 293*, 2001, pp. 2425-30.

[2] D.D. Cox and R.L. Savoy, Functional magnetic resonance imaging (fMRI) brain reading: detecting and classifying distributed patterns of fMRI activity in human visual cortex, *NeuroImage,19*, 2003, pp. 261-270.

[3] J. Mourão-Miranda, A.L. Bokde, C. Born, H. Hampel, and M. Stetter, Classifying brain states and determining the discriminating activation patterns: Support Vector Machine on functional MRI data, *NeuroImage,28*, 2005, pp. 980-95.

[4] F. Pereira, T. Mitchell, and M. Botvinick, Machine learning classifiers and fMRI: a tutorial overview, *NeuroImage,45*, 2009, pp. S199-209.

[5] T. Kohonen, *Self-Organizing Maps*. Berlin : Springer-Verlag, 1995.

[6] Talairach, P., Tournoux, J.,*A Stereotactic Coplanar Atlas of the Human Brain*. Thieme, 1988, Stuttgart.

[7] Ong J.; Raza, S., Data mining using Self-Organizing Kohonen maps: A Technique for effective Data Clustering and Visualisation, *International Conference on Artificial Intelligence*, 1999, pp. 1-4

[8] Tzourio-Mazoyer N, Landeau B, Papathanassiou D, Crivello F, Etard O, Delcroix N, et al. Automated anatomical labeling of activations in spm using a macroscopic anatomical parcellation of the MNI MRI single subject brain. Neuroimage, *2002, 15*, pp. 273-289.

[9] J. Mourão-Miranda, E. Reynaud, F. McGlone, G. Calvert, and M. Brammer, The impact of temporal compression and space selection on SVM analysis of single-subject and multi-subject fMRI data, *NeuroImage, 33,* 2006, pp. 1055-65.

[10] D.R. Hardoon, J. Mourão-Miranda, M. Brammer, and J. Shawe-Taylor, Unsupervised analysis of fMRI data using kernel canonical correlation, *NeuroImage, 37*, 2007, pp. 1250-9.

[11] Kanwisher, N., McDermott, J., & Chun, M. (1997, Jan). The fusiform face area : a module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience, 17*, pp. 4302-11.

[12] Norman, K.A., Polyn, S.M., Detre, G.J. and Haxby, J.V. (2006). Beyond mind-reading: multi-voxel pattern analysis of fMRI data. *Trends in Cognitive Sciences, 10*(3).