

Neuronal Clustering of Brain fMRI Images

Nicolas Lachiche, Jean Hommet, Jerzy Korczak, and Agnès Braud

LSIIT, Pôle API, Bd Brant, F-67400 Illkirch, France
{lachiche, hommet, jjk, braud}@lsiit.u-strasbg.fr

Abstract. Functional Magnetic Resonance Imaging (fMRI) allows the neuroscientists to observe the human brain in vivo. The current approach consists in statistically validating their hypotheses. Data mining techniques provide an opportunity to help them in making up their hypotheses. This paper shows how a neuronal clustering technique can highlight active areas thanks to an appropriate distance between fMRI image sequences. This approach has been integrated into an interactive environment for knowledge discovery in brain fMRI. Its results on a typical dataset validate the approach and open further developments in this direction.

1 Introduction

Functional Magnetic Resonance Imaging (fMRI) provides a unique opportunity to observe the internal behaviour of living organisms. In particular it can be used to observe how the human brain works. This is useful either to gain a better understanding of the human brain in general, or to prepare a surgical operation for a particular patient.

fMRI aims at localizing areas of activity. This is done by considering variations in the blood flow and in the oxygen concentration. Indeed internal activities increase the consumption of oxygen and thus induce the so-called BOLD effect (Blood Oxygenation Level Dependent) which increases the fMRI signal. The studied area is splitted up into cubes with sides usually around 3mm. The intensity of the signal is measured for each of those cubes. A 3D image is formed by associating each cube to a 3D pixel called a voxel. This image shows the intensity of the signal and thus the areas of activity. Compared to other current methods fMRI offers a very good trade-off between spatial and temporal resolutions.

We are interested in brain images. Sequences of fMRI images of the brain are used to check the evolution of the neuronal activity of a patient. The problem is then that the signal on noise ratio is very low, which makes it difficult to identify the areas with a real activity. Moreover, it generates a large amount of data (around 300000 voxels, for which between 100 and 1000 observations are acquired).

Aiming at identifying active areas, the use of mining algorithms seems to be interesting, but they have not reached their full potential yet [9]. Due to the amount of data, most of the studies do not try to explore the data, but use them to test a model with univariate statistics, e.g. Statistical Parametric Mapping

(SPM) [5], AFNI [2], Brain Voyager [7]. Methods to identify the variations which are meaningful can be divided into two families. The most common approach is based on multivariate statistics such as MANCOVA, PCA, PLS, canonical analysis or more recently ICA [1]. The second family gathers all the data mining methods, such as clustering, genetic algorithms and neuromimetic networks [9].

Visual Data Mining allows the integration of the user, an expert, in the knowledge discovery process. In this paper, we present a new interactive approach to fMRI images mining, guided by the data. The originality of our approach relies in the fact that we do not only use data mining techniques which have not been used for that purpose yet, but above all those techniques will be extended by injecting a priori knowledge in the mining process and interactivity. The expert will thus be directly involved in the learning process to identify functional areas of the brain and their organization.

In the next section we explain how a neuronal clustering technique can be applied to identify brain areas that have the same activity, and how it has been integrated into an interactive environment for knowledge discovery in brain fMRI. The results of experiments on a typical dataset are reported in section 4 and validate the use of data mining techniques to complement other approaches in this domain.

2 Neuronal Clustering

Clustering algorithms usually depend on a distance. A distance between voxels has to be defined. A 3D distance between voxels is irrelevant to identify voxels having the same activity. Taking the 3D distance into account would make close voxels -close from a 3D perspective- look more similar than far-away voxels having the same activity. The distance between voxels should only be defined according to their activities. Let us also emphasize that a clustering based on the activity of the voxels without any influence of their localisation is clearly different from segmentation techniques also used to “identify areas” in fMRI images that relies on a comparison of neighbouring voxels.

The activity of a voxel is a continuous signal, in practice a sampling of that signal. One approach consists in generating different attributes describing the signal, e.g. its average, minimum and maximum values, and then using those attribute-values in a traditional attribute-value clustering system. In such an approach, the built-in distance of the clustering system is calculated on the intermediate attributes, e.g. the euclidian distance between the respective average, minimum and maximum values of each voxel. Its success depends on how well the built-in distance and the generated attributes fit together. Lots of attributes can be tested. For instance, [3] made use of wavelets to transform the signal, though they made use of hidden markov models rather than a clustering technique.

We considered an alternative approach, where the distance is directly calculated on the signal. Its obvious disadvantage is that it is then impossible to use an existing clustering system right out of the box. This is balanced by having a distance that exactly measures the difference between two signals. First we

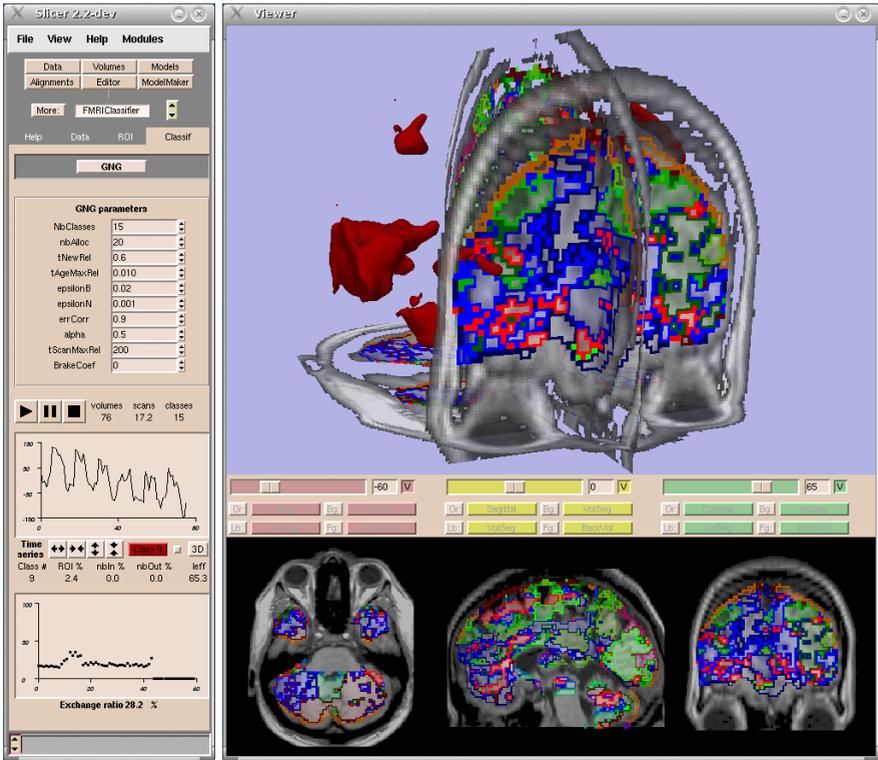


Fig. 1. An environment for interactive knowledge discovery in brain fMRI

remove the average value, then the distance we use is the area between the two variations of the signals around their means.

We have chosen a neuronal clustering algorithm. Kohonen’s Self-Organising Maps [8] is the most known neuronal clustering algorithm. However its fixed topological structure would not help in our application, since there is no a priori topological relationship between the classes¹. Thus we preferred the Growing Neural Gas (GNG) algorithm [6]. Its main advantage is that the number of classes is not fixed in advance as in most clustering algorithms. The class centers can increase as well as decrease during the learning process. Moreover this algorithm easily fits in an interactive knowledge discovery application.

Our aim is indeed to provide a software environment that enables the neuroscientist to improve his understanding of the human brain. While the current approach supposes that the human expert builds up the hypothesis and the software, e.g. SPM [5], is only used to validate that hypothesis, data mining

¹ In order to avoid any confusion between the biological neurons of the brain and the artificial neurons of the clustering algorithms, we will denote the former by the voxels -which actually represent/contain thousands of biological neurons- and the latter by the class centers.

techniques can complement that approach by guiding the expert in his generation of new hypotheses, in particular by automatically showing up activated areas, and highlighting dependencies between those areas.

We have extended the SLICER system for visualisation (<http://www.slicer.org/>) with clustering and interactive exploration of fMRI images features. In its current version, our software environment allows the user to interact graphically with the clustering process, e.g. by modifying parameters of the GNG algorithms, by focusing on specific parts of the brain, etc. A screenshot of our system is presented on figure 1.

3 Experiments

In this section we demonstrate our approach on a typical dataset: an auditory dataset from the Wellcome Department of Imaging Neuroscience of the University College London (<http://www.fil.ion.ucl.ac.uk/spm/>). It consists of 96 brain MRI acquisitions with a 7s time interval. Each acquisition consists of 64 contiguous slices (64x64x64 3mm x 3mm x 3mm voxels). The auditory stimulation consists of

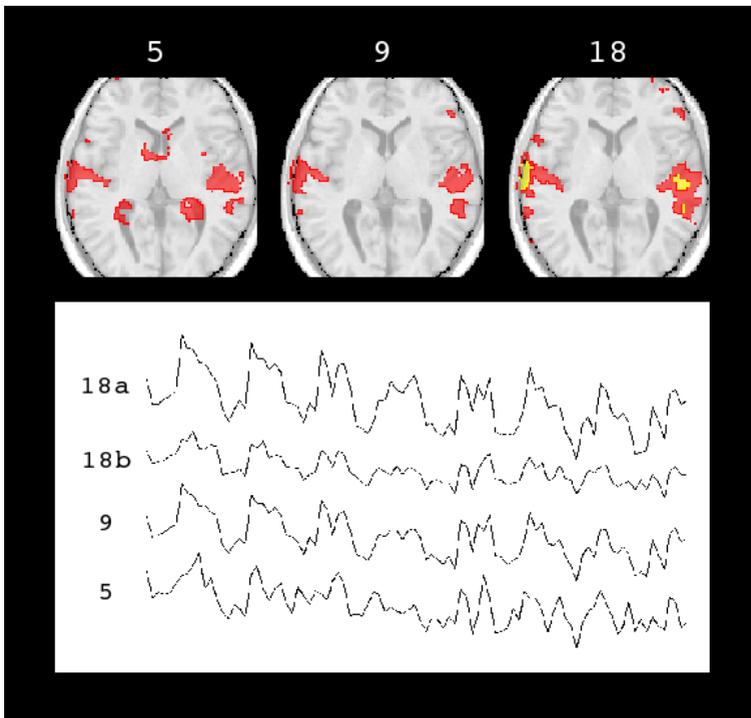


Fig. 2. Active areas and their activities wrt to the number of classes considered in the clustering

bi-syllabic words presented binaurally at a rate of 60 per minute during 42 seconds (6 images), followed by a 42 second silence. It is repeated 8 times.

We illustrate that the clustering algorithm successfully identifies the active areas and their signal. The figure 2 shows the results for three runs with a different number of classes considered in the clustering: 5, 9 and 18. With 5 classes only, the active class overlaps on inactive areas, thus the activity signal do not correspond to the stimulation signal as tightly as with 9 classes. With 18 classes, the active area is decomposed into two parts corresponding to two intensity levels of activation.

4 Conclusion

While the current approach in brain fMRI consists in statistically validating hypotheses made up by the human experts, data mining techniques can help the neuroscientist in the generation of hypotheses. In this paper we showed that a clustering algorithm such as the Growing Neural Gas can successfully highlight active areas in brain fMRI images. It required to express the neuroscientist's interest in terms of a problem that can be solved by a machine learning approach. In particular we had to define a distance between voxels of fMRI images. We argued that such a distance should be based on the signal only, without any influence of the 3D localisation, and that the difference between the variation of the signals is a straightforward alternative to all propositionalisation approaches that could be used to generate attributes then apply usual attribute-value distances. We have integrated our approach on a well-known visualisation platform in order to get a complete interactive environment for knowledge discovery in brain fMRI.

However the identification of active areas is very difficult. fMRI data are inherently very complex: there is a very low signal to noise ratio, lots of different noises are involved, etc. Considering evenemential signals reduces even further the signal to noise ratio. Taking the stimulation signal into account as in [3] can help by synchronising the signals. However it is no longer possible with spontaneous events [4].

Acknowledgements

The authors kindly thank K. Friston et G. Rees for the auditory dataset, C. Scheiber, J. Foucher and D. Gounod for their help in understanding the fMRI domain, and the following students for their contribution to the software development: H. Hager, P. Hahn, V. Meyer, J. Schaeffer and O. Zitvogel.

References

- [1] C. Beckmann and S. M. Smith. Probabilistic independent component analysis for functional magnetic resonance imaging. *IEEE Trans. on Medical Imaging*, 2003.
- [2] R. W. Cox. AFNI: Software for analysis and visualization of functional magnetic resonance neuroimages. *Computers and Biomedical Research*, 29:162–173, 1996.

- [3] S. Faisan, L. Thoraval, J.-P. Armspach, and F. Heitz. Unsupervised learning and mapping of brain fmri signals based on hidden semi-markov event sequence models. In *Medical Image Computing and Computer Assisted Intervention (MICCAI)*, pages 75–82, 2003.
- [4] J. R. Foucher, H. Otzenberger, and D. Gounot. Where arousal meets attention: a simultaneous fmri and eeg recording study. *Neuroimage*, 22(2):688–697, 2004.
- [5] K. J. Friston, A. P. Holmes, K. J. Worsley, J. P. Poline, C. D. Frith, and R. S. J. Frackowiak. Statistical parametric maps in functional imaging: A general linear approach. *Human Brain Mapping*, 2:189–210, 1995.
- [6] B. Fritzsche. A growing neural gas network learns topologies. In G. Tesauro, D. S. Touretzky, and T. K. Leen, editors, *Advances in Neural Information Processing Systems 7*, pages 625–632. MIT Press, Cambridge MA, 1995.
- [7] R. Goebel. *BrainVoyager: Ein Programm zur Analyse und Visualisierung von Magnetresonanztomographiedaten*. T. Plessner and P. Wittenburg, Forschung und wissenschaftliches Rechnen, 1997.
- [8] T. Kohonen. Self-organized formation of topologically correct feature maps. *Biological Cybernetics*, 43:59–69, 1982.
- [9] F. T. Sommer and A. Wichert. *Exploratory Analysis and Data Modeling in Functional Neuroimaging*. The MIT Press, Massachusetts Institute of Technology, Cambridge, Massachusetts, 2003.