What Believe in Multivariate Pattern Analysis? The Skeptical Neuroimager's View

Chair: Bertrand Thirion, INRIA Saclay-Ile-de-France, Parietal team, Neurospin, Gif sur Yvette, France

Multivariate pattern analysis has been used quite intensively in neuroimaging studies during the last few years, because it provides a very sensitive assessment of the link between brain images or brain signals and some stimulus or behavioral variable reflecting the subject's mental state. This approach is sometimes called decoding. Its success is fueled by the active and continuous development of powerful machine learning tools in the last decades, in particular in the field of supervised classification and regression. It is motivated by the neuroscientific idea that mental representations utilize population codes, i.e. the information is combinatorially encoded in patterns of activity. Even within a collection of noisy data, multivariate pattern analysis tools can detect signals linked to the target variable (i.e. carrying information on this variable), and use it to achieve above-chance classification of mental states. In practice, sensitivity gains have been central to the success of these approaches in neuroimaging. It can be noted that machine learning tools used for decoding have come with a more systematic use of cross-validation procedures, which yields a more compelling assessment of the informative content of brain images than analytical criteria. More importantly, pattern analysis has opened the possibility to generalize predictions across experimental conditions, sometimes providing new insights on brain function. The success of pattern analysis is also related to its ability to obtain graded measures of similarity/differences of brain states, stimuli or percepts.

This successful paradigm nevertheless faces several challenges, namely i) the lack of modeling behind most successful brain reading analyses, ii) the non-uniqueness, or degeneracy of patterns that actually convey information on the variable of interest, iii) the lack of consistency of the discriminative patterns used by the classifier, and iv) the difficulty of analyzing the geometry of multivariate representational spaces.

Challenge 1: How to test computational theories of brain information processing? The most frequently used approach in Multivariate pattern analysis (MVPA) consists in training a classifier on a the signals from a set of brain regions, possibly cascading several processing steps to improve the classifier's performance. All these procedures are agnostic to brain mechanisms, and just provide a statistical measurement on the shared information between stimulus or behavioral variables and activity patterns. While the same criticism holds both for univariate activation analysis and MVPA approaches, sophisticated MVPA approaches can in some cases provide more insights when they incorporate computational models of brain information processing (e.g. by introducing explicit priors on the stimulus organization or by introducing latent factors that model the similarity between brain processes).

Challenge 2: How to localize information and understand the spatial organization of neuronal codes? Due to the large size of brain images used in MVPA, a common observation is that equally powerful classifiers trained on one dataset can be based on completely different spatial patterns. More generally, brain signal classifiers rely on distributed rather than focal information, they cannot inform the neuroimager about the precise localization of this information; only the searchlight approach is well suited for such purpose; but even in that case, the spatial encoding of the information remains implicit.

Challenge 3: Do the discriminating patterns provide a consistent estimate of the truly informative regions? Classifiers are instantiated and optimized to yield an optimal response to a prediction problem, and not to recover the ground truth of an activation pattern. For instance, the discriminating pattern inferred from an Support Vector Machines algorithm provide no principled evidence to accurately
delineate task-related regions: elementary simulations show indeed that it will fail to do so in simplistic cases. In general, one should not expect from a good classifier to convey the true model of brain activation pattern. Still, embedded variable selection approaches (i.e., variable selection combined with model learning, such as various sparse approaches), have some potential for recovering relevant regions of brain activity. We wish to provide theoretical and experimental evidence of correct recovery of the underlying spatial structure of the discriminating pattern.

Challenge 4: How to analyze the geometry of multivariate representational spaces? As the space spanned by the activity patterns are intrinsically high-dimensional, their understanding is not easy: beside the traditional issues regarding visualization of multi-dimensional information, the comparison of patterns across individuals, conditions or protocols remains a challenging task.

**Learning Objectives:** Having completed this workshop, participants will be able to:

1. Discuss the success and pitfalls of pattern analysis techniques applied to neuroimaging, carefully considering the ultimate goal of learning procedures used as tools for classification purpose.
2. Review the existing results on the consistency of the recovered spatial patterns, i.e. under which conditions some pattern analysis procedures can give access to the underlying activation pattern.
3. Describe alternative efforts that address the challenges described above: the incorporation of computational models of brain-information processing, the generalization of classification rules across tasks, the construction and characterization of the latent stimulus space from activation data.
4. Understand the use and interpretation of pattern analysis approaches in neuroimaging, based on technical and statistical considerations, as well as our experience on using MVPA on neuroimaging data.

**Feature Selection and Feature Extractions in Multivariate Prediction: Promises and Limitations**

Irina Rish, *IBM T.J. Watson Research Center, Computational Biology Center, NY, USA*

One of the central topics in statistical analysis of neuroimaging data is discovering brain areas relevant to a given stimuli or mental state. In the past years, it has been widely recognized that predictive accuracy of multivariate models can serve as a better relevance measure than the traditional univariate voxel activations (i.e., univariate voxel correlations with a stimulus or mental state), since the latter approach ignores potentially important multivariate voxel interactions. However, since an exhaustive search for the most-informative subset of voxels is clearly intractable, an emphasis was placed on so-called embedded feature selection methods, such as, for example, sparse regression and sparse classification. Sparse multivariate modeling is a promising statistical approach that appears ideally suited for achieving both predictive accuracy and interpretability, as it fits a predictive model simultaneously with selection of predictive variables (e.g., voxels involved in prediction of a mental state), using l1-norm based constraints to eliminate non-essential variables. However, despite its promises, sparse multivariate modeling must be used with care, as interpretability does not equal to sparsity: multiple well-predicting sparse solutions may exist when the variables are highly correlated, as it is the case of neuroimaging data. Thus, while the brain areas included into a highly predictive sparse solution are clearly relevant to the task, it is not clear how unique such solutions are, i.e. how much information about the task is still contained in the rest of the brain. This leads to the following questions: should one expect a sharp boundary between task-relevant and task-irrelevant brain areas, or rather a widespread distribution of task-relevant information across the whole brain? How does the task-related information distribution depend on the properties of the task? Can we find a better
representation (i.e., a better ‘dictionary’, or feature set, than plain voxels) that yield more stable, unique sparse models, highly predictive about the mental states of interest? The last question is known as ‘feature extraction’, ‘feature engineering’ or ‘biomarker discovery’ issue that often contributes more to the success of learning than the choice of any particular predictor. In this talk, we will focus on feature selection and feature extraction in neuroimaging, and discusses recent advances in sparse modeling as well as its limitations.

**Can We Recover Meaningful Spatial Information from Multivariate Pattern Analysis?**
Bertrand Thirion, INRIA Saclay-Île-de-France, Parietal team, Neurospin, Gif sur yvette, France

Brain activity decoding, is a recent paradigm for analyzing functional magnetic resonance imaging (fMRI) data, based on pattern recognition tools. It enables neuroscientists to take into account the multivariate information between voxels and is currently the only way to assess how precisely some cognitive information is encoded by the activity of neural populations within the whole brain. However, it relies on a prediction function that is plagued by the curse of dimensionality, as we have far more features than samples, i.e., more voxels than fMRI volumes. While many machine learning solutions have been proposed to deal with this kind of issues, with more or less sophistication, these techniques are designed to maximize prediction accuracy, thus solving a supervised learning problem. It turns out that neuroimagers are often interested in interpreting the spatial pattern used by the model, a question to which popular machine learning algorithms do not give a satisfactory response: on the theoretical side, many methods do not offer any guarantee to pick a good model, or even to converge to the right model with many observations; in practice, many different models relying on the same spatial pattern can produce strikingly different solutions. We will first review some key concepts of pattern analysis, such as regularization and sparsity, and some known results on the quality of spatial pattern recovery in high-dimensional prediction problems. Then we will discuss recent solutions proposed to deal with this problem, essentially trying to build discriminative patterns based on natural hypotheses of the problem (clustering and spatial smoothness).

**Why We Should Believe in Pattern Analysis and How to Meet the Challenges Ahead**
Nikolaus Kriegekorte, MRC Cognition and Brain Sciences Unit, Cambridge, UK

Pattern-information analysis is motivated by the neuroscientific concept of population coding, the idea that mental content is represented combinatorially in patterns of brain activity. As long as the representational paradigm of population coding remains central to our understanding of brain function, we will need pattern analysis to reveal the information content of brain representations in both cell recording and functional imaging. However, several challenges face the approach: (1) Pattern classifiers have been the dominant variant of pattern analyses, but they require the stimuli (or mental states) to fall into discrete predefined classes. We need to develop parametric variants of pattern information analysis to deal with more complex experimental designs where the stimulus space is sampled by many individual stimuli and the analysis is not biased by a predefined grouping. (2) Statistical models from machine learning help detect information in brain-activity patterns, but they are not optimally suited for revealing the spatial organization or the representational geometry of the neuronal code. (3) Revealing the information in each brain region is an important intermediate goal, but ultimately we would like to be able to test computational models of brain information processing with pattern analysis. I will discuss current developments that address these challenges and illustrate them with a series of insights about brain function that would not have been possible without pattern analysis.
Multi-Voxel Pattern Analysis as a Tool to Look Inside the Modules of the Brain
Hans Op de Beeck, Laboratory of Biological Psychology, Leuven, Belgium

During the past two decades noninvasive functional imaging studies have parceled the brain into smaller chunks and regions. These regions are sometimes referred to as modules and often treated as such: uniform volumes of interest. I will discuss the potential of multi-voxel pattern analyses (MVPA) to understand what happens inside these modules. MVPA clearly offers us a substantial increase in sensitivity and a related increase in the number of conditions that can be differentiated in terms of how they activate these modules. I refer to this as an increase in conceptual resolution. It is less clear, however, whether MVPA also increases the spatial resolution of functional imaging by allowing us to investigate patterns of activity that are organized at a sub-voxel scale within these modules. Direct evidence for this claim is limited, and in this respect I will discuss recent data suggesting an internal functional organization in module-like regions such as the fusiform face area and the visual word form area.

Session Topic(s):
Higher cognitive functions
Informatics
Modeling and Analysis Methods