Machine Learning for Neurological Disorders

Matthew Blaschko

Machine Learning for Neurological Disorders

Matthew B. Blaschko

The last two decades have seen tremendous advances in our understanding of human brain structure and function, particularly at the level of systems neuroscience, where neuroimaging methods have led to better delineation of brain networks and brain modules. Brain understanding is one of the greatest challenges of our century with enormous potential impact in a number of fields, including medicine. Recent progress in the hardware side has made possible the in-vivo acquisition on top of structural/anatomical data, functional information (through emerging image modalities like functional magnetic resnonance imaging (fMRI), diffusion tensor imaging (DTI), magneto encephalogram (MEG), electro encephalogram (EEG), etc.) in a non-invasive manner depicting task-specific states of the brain. Such information can be of great interest towards understanding of neurogenerative diseases, and providing means of assessing the impact of different therapeutic strategies.

These data are by necessity high dimensional and complex, driving the widespread application of machine learning techniques for their analysis. Machine learning in this context works by learning a regressor from the space of brain recordings (e.g. voxels of a fMRI scan) to an output variable describing a condition to be explained. The form of the regressor is important as it encodes the class of interrelationships within the brain that can be learned, e.g. linear relationships, or non-linear interactions encoded in a graph structure. Equally important is the choice of output variable, e.g. the presence or absence of a neural disorder, or the identity of a stimulus or task undertaken by the subject during recording. Through the form of the regression task, a large number of neuroscientific questions may be elucidated, e.g. the neural correlates of visual processing, emotion, addiction, Alzheimer's, or autism.

In the Center for Visual Computing, our research in this area has been guided by the translation of neuroimaging problems into rich mathematical representations such as graphs for use in machine learning algorithms. By employing novel machine learning algorithms and representations, we expand our ability to understand brain function in a way that is reproducable, interpretable, and statistically sound.

A key area of neuroscience that we have been working on is the understanding of the neural mechanisms of cocaine addiction. In collaboration with neuroscientists at the Icahn School of Medicine in New York and computer scientists at Stony Brook University, we have employed recordings of cocaine addicted and control subjects to find regions of the brain that discriminate the two groups based on functional activations in a task carefully designed to differentially invoke associations with word meanings. The work has been successful not only in showing the applicability of novel machine learning techniques, but in its confirmation of the deactivation of the rostral anterior cingulate cortex (rostral ACC) in cocaine addicted vs. control groups. This is an important result as it provides evidence that certain categories of theraputic interventions previously applied in other addictions may be effective in cocaine addiction as well. As a result of this research, Katerina Gkirtzou successfully defended her doctoral dissertation entitled "Sparsity regularization and graph-based representation in medical imaging" in December, 2013.

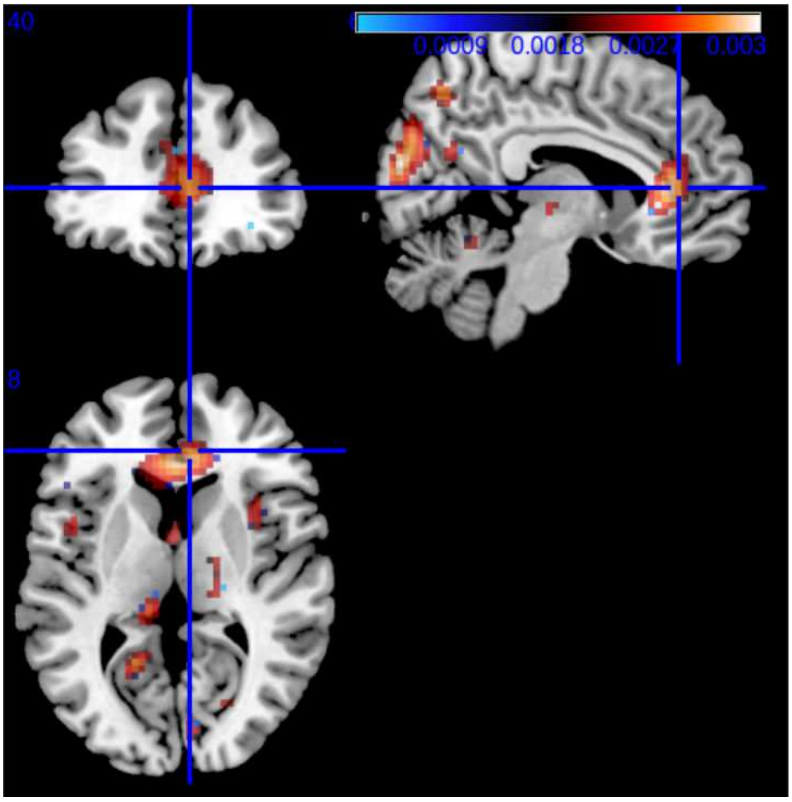

FIGURE 1. Caption: A visualization of the areas of the brain identified as being discriminative between cocaine addicted and control subjects. The rostral anterior cingulate cortex is prominent in the selected areas implicating a role in cocaine addiction (Gkirtzou et al., ISBI 2013).

A focus of research in our group has been on the development of methodologies for neuroscientific studies, rather than simply providing an answer to a specific question. We are therefore addressing applications to several neurological disorders, including Alzheimer's disease and autism. Our work is greatly assisted through collaboration and integration with multiple partners on the Saclay Plateau, and assisted by resources made available to support such integration. Wacha Bounliphone is a doctoral student jointly funded by Centrale and Supelec, and cosupervised by Matthew Blaschko (Centrale & Inria) and Arthur Tenenhaus (Supelec). Her doctoral research is focused on statistical tools for the integration of neuroimaging and genetics information. Such tools are essential to understanding both the neural and genetic mechanisms of Alzheimer's, and the statistical tools developed will be immediately applicable to multi-modal data analysis across a wide range of domains.

Our work on autism is primarily supported through Digiteo, an organization created to support scientific research and collaboration across dozens of laboratories on the Saclay Plateau, including Ecole Centrale. We are fortunate to have both a visiting chair (Prof. Dimitris Samaras, Stony Brook University), and a doctoral student funded by Digiteo. Due to the largely fractured nature of neuroscience data collection, and the statistical benefits of larger sample sizes, a recent development in neuroscience has been the creation of amalgamated databases incorporating recordings taken at multiple sites, such as the Autism Brain Imaging Data Exchange, or the Functional Imaging Biomedical Informatics Research Network. These databases are sufficiently anonymized to ensure patient privacy, and provide sample sizes infeasible to collect at a single site. They have begun to develop a culture of open data in neuroscience, and have enabled scientific discoveries in a short time-span. Nevertheless, such collections inevitably contain heterogeneities, both in recording conditions and annotation. It is within this context that we are developing new machine learning methodologies to enable large scale inference of neuroscientific principles from amalgamated fMRI data sets. Due to the large scale nature of the data, learning must be efficient. Furthermore, the

learning algorithm is to extract relevant annotations from noisy or incomplete structured data fields associated with the fMRI recordings.

The range of mathematical methods developed at Centrale for neuroimaging studies is diverse, spanning discriminative latent variable models, structured sparsity regularization, the representation of brain recordings as graphs, and structured output prediction. By applying a range of statistical innovations to a diverse set of interesting neuroscientific questions and problems, our lab is endeavoring to maximize its impact from the perspective of (i) innovating methodologies for neuroscience studies, (ii) addressing a wide range of neuroscientific questions and applications, and (iii) ultimately driving research that leads to increased understanding of the mind and its pathologies, as well as supporting improved theraputic strategies driven by direct observation of functional activations of the brain. Our research is strongly supported by partners across the Saclay Plateau including Inria, Supelec, Digiteo, and Neurospin (a neuroimaging laboratory of the CEA).

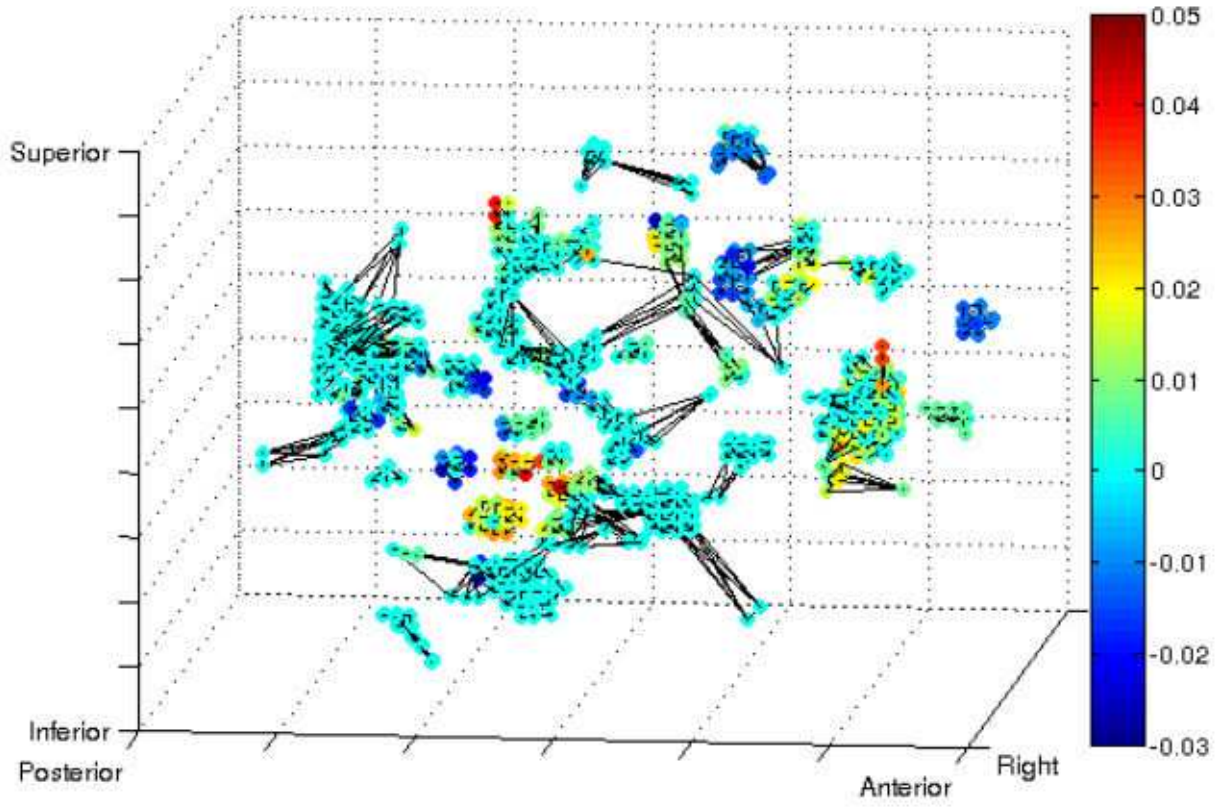

FIGURE 2. Caption: We are exploring graph theoretic methods for fMRI analysis, including novel visualizations of brain activity (Gkirtzou et al., MLMI 2013).

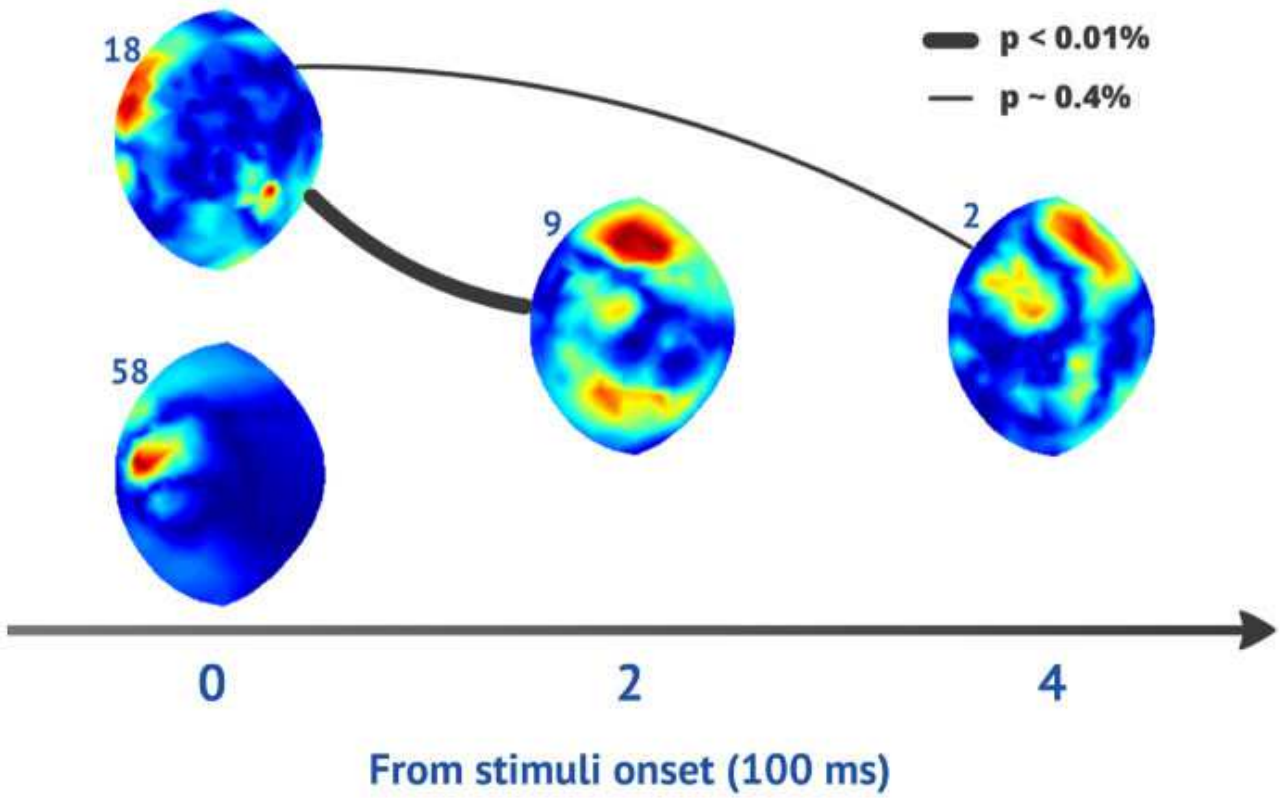

FIGURE 3. Caption: We have made use of discriminative latent variable models to infer functional connectivity of the brain from M/EEG recordings (Zaremba et al, IPMI 2013).

References

Katerina Gkirtzou, Jean Honorio, Dimitris Samaras, Rita Goldstein, and Matthew B. Blaschko. fMRI analysis of cocaine addiction using k-support sparsity. In Biomedical Imaging (ISBI), IEEE 10th International Symposium on, pages 1078-1081, 2013.

Katerina Gkirtzou, Jean Honorio, Dimitris Samaras, Rita Goldstein, and Matthew B. Blaschko. fMRI analysis with sparse Weisfeiler-Lehman graph statistics. In Guorong Wu, Daoqiang Zhang, Dinggang Shen, Pingkun Yan, Kenji Suzuki, and Fei Wang, editors, Machine Learning in Medical Imaging, volume 8184 of Lecture Notes in Computer Science, pages 90-97. Springer, 2013.

Wojciech Zaremba, M. Pawan Kumar, Alexandre Gramfort, and Matthew B. Blaschko. Learning from M/EEG data with variable brain activation delays. In James C. Gee, Sarang Joshi, Kilian M. Pohl, William M. Wells, and Lilla Zöllei, editors, Information Processing in Medical Imaging, volume 7917 of Lecture Notes in Computer Science, pages 414-425. Springer, 2013.