INRIA Saclay, Équipe Parietal http://team.inria.fr/parietal NeuroSpin, bat 145, CEA Saclay

Template estimation for arbitrary alignments

Research theme: machine learning, functional brain imaging
Keywords: regression, fMRI, optimal transport.
Duration & salary: 4 to 6 months, between 500 €and 800 € monthly
Research team: Parietal (INRIA Saclay and CEA)
Adviser: Bertrand Thirion
Contact: bertrand.thirion@inria.fr
Application: Interested candidate should send CV and motivation letter

Context: Inferring a template from a sample of complex signals (medical images, high-frequency signals) is a challenging task. Standard statistical approaches revolve around the idea of averaging, yet this most often deletes important information: the resulting average is qualitatively not similar to the input signal. This means that the Euclidean or L_2 distance implicit to standard averaging is not relevant for that task. Other, possibly domain-dependent, metrics need to be considered to obtain more meaningful templates and make inference on their properties.
We propose to revisit this question in the context of brain images, where a standard representation is sought to summarize the information carried by a population of images. Euclidean averaging unavoidably loses information [3].
The ongoing accumulation of brain data at the individual level opens a new perspective: that of a feature-based alignment of brain voxels[7, 6, 4]. However, none of the existing works has solved the problem of template inference; heuristics were used instead. In this internship, we propose to address this question, taking advantage of the accumulation of data in individual brains, obtained through intensive scanning in the Individual Brain Charting project http://project.inria.fr/IBC.

Proposed work:
We will consider two main approaches to solve the problem.

- The use of non-Euclidean metrics to compute the distance between samples: the celebrated Wasserstein metric suggests itself as a good candidate [2], yet it must deal with three issues: *i)* the samples are not probability distributions *ii)* the samples are potentially high-dimensional *iii)* these distances are expensive to compute and minimize. We will consider and extend recent work [5] to deal with these issues.

- The use of generative models: A more natural approach may be to revisit the definition of a template, as a parameter of a generative model of the data. In the case of brain maps, sparse low-rank linear models probably capture the most of the information, hence we will consider inferring this type of models from the data. For this we will revisit and extend the so-called shared response model [1], using local dictionaries that are sparsely represented in the data.

We will then carry out validation experiments, whereby we will measure the accuracy of the template in predicting new data.

Required skills: The successful candidate will be interested in applications of machine learning and in the understanding of human cognition. Knowledge of scientific computing in Python (Numpy, Scipy, scikit image, Pandas) is encouraged. All the work will be done in Python based on the Nilearn library http://nilearn.github.io.

[1] Po-Hsuan (Cameron) Chen, Janice Chen, Yaara Yeshurun, Uri Hasson, James Haxby, and Peter J Ramadge. A reduced-dimension fmri shared response model. In C. Cortes, N. D. Lawrence, D. D. Lee, M. Sugiyama, and R. Garnett, editors, *Advances in Neural Information Processing Systems 28*, pages 460–468. Curran Associates, Inc., 2015.

[2] Marco Cuturi and Arnaud Doucet. Fast computation of wasserstein barycenters. In Eric P. Xing and Tony Jebara, editors, *Proceedings of the 31st International Conference on Machine Learning*, volume 32 of *Proceedings of Machine Learning Research*, pages 685–693, Bejing, China, 22–24 Jun 2014. PMLR.

[3] Evelina Fedorenko, Michael K. Behr, and Nancy Kanwisher. Functional specificity for high-level linguistic processing in the human brain. *Proceedings of the National Academy of Sciences*, 108(39):16428–16433, 2011.

[4] Matthew F. Glasser, Timothy S. Coalson, Emma C. Robinson, Carl D. Hacker, John Harwell, Essa Yacoub, Kamil Ugurbil, Jesper Andersson, Christian F. Beckmann, Mark Jenkinson, Stephen M. Smith, and David C. Van Essen. A multi-modal parcellation of human cerebral cortex. *Nature*, advance online publication, July 2016.

[5] Peyré G. Gramfort, A. and M. Cuturi. Fast optimal transport averaging of neuroimaging data. In *Proc. IPMI 2015*, July 2015.

[6] James V. Haxby, J. Swaroop Guntupalli, Andrew C. Connolly, Yaroslav O. Halchenko, Bryan R. Conroy, M. Ida Gobbini, Michael Hanke, and Peter J. Ramadge. A common, high-dimensional model of the representational space in human ventral temporal cortex. *Neuron*, 72:404–416, 2011.

[7] M.R. Sabuncu, B.D. Singer, B. Conroy, R.E. Bryan, P.J. Ramadge, and J.V. Haxby. Function-based intersubject alignment of human cortical anatomy. *Cereb Cortex*, 20:130, 2010.