Post doc « Advanced statistical methods for analysis of multidimensional databases of human brain imaging»

Description of the Post doc research project

The study of the variability of both normal biological phenomena, and pathological alteration is based on novel experimental design. These typically represent multivariate outcomes (genotype, brain phenotype, complex behavioural observations) and multifactorial (genes, age, education, socio-demographic factors, risk factors, ...). Consequently, the nature of the data/observations is strongly multidimensional nature and can scale up to the order of 10^7 for the genomic data and may be even larger in case of brain imaging. These characteristics result in two major consequences from an experimental point of view:

- these imply the collection of the data on a large number of subjects, ad hoc data, and stratified on a maximum of factors if possible;
- further, these generate a need for novel analytical methods to make the most of such complex data.

The aim of this project is to initiate that stream of research within the IdEx Bordeaux centre by conducting a pilot study involving two high ranked groups:

- the "Probability and Statistics" group (EPS) from the LabEx CPU, focussing on the methodological developments to analyse high dimensional data,
- the "Groupe d'Imagerie Neurofonctionnelle" (Neurofunctional Imaging Group), (GIN) from LabEx TRAIL, which has been leading research on the identification of sources of variability of the human brain with neuroimaging methods, for the past years.

The project will focus on the analysis of variability factors driving hemispheric specialization (HS) of the brain, a human specific character, for which a dedicated database has recently been built by GIN. According to some authors (Crow Laterality 2009), HS is a feature of the brain organization that would define the species. Also referred to as hemispheric dominance, the HS corresponds to the loss of specific brain function in case of lesions in the hemisphere. The search for the neural basis of HS is emerging as a highly competitive research stream since the development of modern imaging techniques, such as MRI, which have revolutionized the (TICS Hervé et al 2013). From these techniques, HS is revealed by asymmetric morphology and/or activity between the two brain hemispheres. It has been shown that, in more than 90 % of right-handed people, language is a function from the left hemisphere (LH), and due to the variability of the hemispheric organization, subjects in whom the right hemisphere (RH) is dominant for language may be observed. Similarly, the RH is usually dominant for visuospatial functions and particularly visual attention, but again the variability between individuals is important and is dependent of the kind of visuospatial activity. There are many factors affecting driving the HS, and handedness has shown to lead to greater variability for language lateralization in left handed subject. Mainly, inherited left-handedness has been associated to a decreased hemispheric lateralization. Several studies have shown an effect of gender on the asymmetries observed during phonological tasks. During the development, asymmetries are up regulated during learning stages and then reverse to lower levels with ageing. One interesting area of research resides in the identification of the relationship between HS for different types of cognitive functions such as language and attention that are mirrored in most right-handed.

Centre d'excellence CPU
351, Cours de la libération
33405 cedex Talence France
T: 33 (0)540002128

Due to the complex nature of the factors and cognitive function relating to hemispheric lateralization, HS appears as a typical example of biomedical theme whose investigation would require innovative mathematical methods accommodating high dimensional multivariate data.
To quantify and understand the role of these factors in the development of different cognitive functions HS, GIN has compiled a database: BIL & GIN (Brain Imaging of Lateralization) combining both anatomic and functional brain MRIs, detailed psychometric data on handedness, verbal and visual-spatial cognitive skills, genetic and demographic data in more than 300 healthy volunteers containing 150 left-handed participants. For each subject, anatomical images from a set of global and regional brain morphometry variables were calculated as well as some measures of asymmetry. In addition, a set of global and regional variables describing the HS for a range of cognitive functions have been derived from functional images. Cognitive functions that have been included GIN & BIL are those with significant hemispheric lateralization. Each subject was screened, during production tasks, during understanding and reading tasks as well as during a semantic task and regarding the language a phonological task. Visual attention was assessed during Saccade tests and bisection line assessment tests. Motion lateralization was described through left and right hand movement recordings as well as a hand mental figuration task. Finally data relating to calculation activities involving both attention networks and language capacity were included to characterize the development of their lateralization with complex computing tasks and evaluation intervals. It is worthwhile to note that each participant enrolled in BIL & Gin also took part in a study aiming at the characterisation of the networks involved in freethinking. Finally, saliva DNA samples for each participant are also available and therefore enable genotyping.
GIN will provide the database and will perform genotyping of fifty loci potentially affecting HS. GIN will also contribute to the project by interacting with researchers from the EPS, by identifying and characterizing the best variables, performing additional analyses, and suggesting appropriate additional variables, especially in the case of the voxel being implemented. GIN will also be involved in the interpretation of the results generated throughout the project.

To analyse high-dimensional data, in a context where the number of individuals is lower than the numbers of variables, number of multivariate methods have been proposed in the literature. These include notably, LASSO regression and its developments (Tibshirani et al, 1996), sPLS (sparse Partial Least Squares , The Cao et al, 2009) and a regression-based Bayesian variable selection approaches (ESS + + , Bottolo et al, 2010). SIR- QZ,will aslo be considered. That new approach (Sliced Inverse Regression based on QZ , Coudret et al algorithm , 2013) was developed within the team " probability and statistics " (R. Coudret , J. Saracco) of EPS in collaboration with Benoît Liquet (U897). Finally and a method combining the classification variables (ClustOfVar , Chavent et al , 2012) and variable selection by random forests (VSURF , Genuer et al, 2010) developed within the EPS (Mr. Chavent J. Saracco) in collaboration with R. Genuer of U897 will be investigated. These different approaches should identify, quantify or select groups of variables that have an impact and influence on the SH.
However, most of these approaches will be extended and developed for the framework where you want to explain several features of the HS variables simultaneously. Particular attention will be paid on dimension reduction model, for example those based on indices (linear combinations of the original variables) as in the SIR approaches or PLS. Indices can be built on both the explanatory part and the dependent part (to explain) of the model. These models ensure the visualization of complex the data reduced-dimension spaces and the interpretation of the impact of each variable. Various approaches to model and analyse high-dimensional data (eg ClustOfVar VSURF or type) could be associated to methods based on indices.
Note that other types of approaches can of course also be considered as part of this project.

Centre d'excellence CPU
351, Cours de la libération
33405 cedex Talence France
T: 33 (0)540002128

The results in terms of application / interpretation are clearly the priority of this project. However, theoretical developments emerging from the project are also expected. Finally, R packages combining new methodologies developed in this project will be implemented and made available to the scientific community.

References:

Bottolo, L, et al. ESS++: a C++ objected-oriented algorithm for bayesian stochastic search model exploration. Bioinformatics, 27(4):587–588, 2011.

Coudret, R, Liquet, B. and Saracco, J. Sliced inverse regression in underdetermined cases. To appear in Journal de la Société Française de Statistique, 2013.

Chavent, M, Kuentz-Simonet, V, Liquet, B, and Saracco, J. ClustOfVar: An R Package for the Clustering of Variables. Journal of Statistical Software, Vol. 50, pp. 1-16, 2012.

Genuer, R, et al. Variable Selection using Random Forests. Pattern Recognition Letters, Vol. 31, pp.2225-2236, 2010.

Le Cao, K.A., et al. Sparse canonical methods for biological data integration: application to a crossplatform study. BMC Bioinformatics, 10(1):34, 2009.

Duration Full time for 20 months

Job statuts Fixed-term contract

Salary after taxes per month : 2.400 euros

Division/School: IdEx of Bordeaux (project inter-LabEx CPU &TRAIL)

Location : University campus in Talence (33, FRANCE)

Date: start : January 2014 (A shift of 1 or 2 months of starting the post-doc is potentially possible.)

Profile of appliquant

Preference will be given to a candidate with a PhD in Statistics or Biostatistics. A strong orientation towards biological applications or in the field of health is essential, and additional experience in imaging and genomics will be considered as an asset. Applications with a strong "bioinformatics" component will be considered if the candidate has strong experience in statistics or biostatistics.

Keywords : statistical modelling, dimension reduction, high dimensional data analysis, variable selection, classification, neuroimaging data, genomic data, computer programming in R.

Supervisors/Contact

- Jérôme Saracco (IMB & CQFD team) : jerome.saracco@math.u-bordeaux1.fr
- Bernard Mazoyer (GIN) : mazoyerb@gmail.com

Application deadline : January 2014