

Post-doctoral researcher in Statistics/Machine Learning at CentraleSupélec and NeuroSpin

Title: Multimodal analysis of the Imagen NeuroimagING dataset (MINING)

Start date: From November 2015

Contract duration: 12 months

Gross salary: to be defined

Primary Contact: Arthur Tenenhaus (arthur.tenenhaus@centralesupelec.fr), Laurent Le Brusquet (Laurent.lebrusquet@centralesupelec.fr) and Vincent Frouin (Vincent.frouin@cea.fr)

Description of the proposal. IMAGEN is a European Research Project which aim is to identify and learn more about biological and environmental factors that might have an influence on mental health in teenagers. This knowledge will allow developing better prevention strategies and therapies in the future. The IMAGEN database includes, for about two thousand 14 years old adolescents: (i) demographics data, (ii) neuropsychological assessments, psychometry, medical questionnaires, (iii) multimodal neuroimaging (including MR functional, structural and diffusion weighted neuroimaging) and (iv) omics (SNP and methylation) data. All these datasets are already centralized at NeuroSpin. The IMAGEN dataset gathers all the challenges that have to be faced in modern multivariate data analysis. The first bottleneck is the high complexity of the data that stems from (i) various sources: genetics, neuroimaging, etc. (ii) the number of neuroimaging modalities and (iii) the multi-centric nature of the data. The second bottleneck is the high number of measurements (~1M) in both genetic and neuroimaging data which involves the computation of billion(s) of associations. A successful investigation of such a dataset requires developing a computational and statistical framework that fits both the peculiar structure of the data as well as its heterogeneous nature. In that context, the MINING project aims specifically to focus on the tensor structure of the IMAGEN dataset which arises naturally when considering multi-modal neuroimaging.

The goal of the project is to develop a proper multiway/tensor-based framework that allows considering conjointly several neuroimaging modalities and studying their associations with various categorical or continuous outcomes (e.g. sub-depression, binge-drinking status, genetic factors). This multiway framework may be seen as a dedicated way to introduce prior information and control the over-fitting issue leading to more interpretable results. NEUROSPIN and L2S gather experiences in the IMAGEN database, in the statistical analyses already carried out and in the field of structured data analysis [1, 2, 3].

Requirements (training/expertise) and profile. PhD in statistics/machine learning. Past experience in multivariate data analysis/Machine Learning applied to neuroimaging and/or biological data will be strongly appreciated.

To apply, submit a cover letter indicating past research experience, motivation for the position, expected availability date and curriculum vitae directly to arthur.tenenhaus@centralesupelec.fr, Laurent.lebrusquet@centralesupelec.fr and Vincent.frouin@cea.fr.

References

- [1] Bro, R., 2000, Multi-way Analysis in the Food Industry - Models, Algorithms, and Applications ICSLP Proceedings
- [2] Le Brusquet L., Lechuga G., Tenenhaus A., 2014, Régression Logistique Multivoie, 46eme Journée de Statistique
- [3] Le Brusquet L., Tenenhaus A., Lechuga G., Perlberg V., Puybasset L., Galanaud D., 2015, Une pénalité de groupe pour des données multivoie de grande dimension, 47eme Journée de Statistique