



Offre n°2024-07904

Doctorant F/H Foundation Models of human brain function

Type de contrat : CDD

Niveau de diplôme exigé : Bac + 5 ou équivalent

Fonction : Doctorant

Niveau d'expérience souhaité : Jeune diplômé

A propos du centre ou de la direction fonctionnelle

The Inria Saclay-Île-de-France Research Centre was established in 2008. It has developed as part of the Saclay site in partnership with Paris-Saclay University and with the Institut Polytechnique de Paris since 2021.

The centre has 39 project teams, 27 of which operate jointly with Paris-Saclay University and the Institut Polytechnique de Paris. Its activities occupy over 600 scientists and research and innovation support staff, including 54 different nationalities.

Contexte et atouts du poste

One of the major directions for future neuroscience is to build on the expertise accumulated in AI-powered cognitive systems, such as architectures that process language or visual content, but in the future will also include motor actions, planning and navigation. While both AI and neuroscience will benefit from comparing brain activity data with AI systems, one difficulty is that the links between AI models and brain activity have been made in very specific contexts [1, 2, 3] and may not generalise beyond a few standard situations (static images, language, sounds). A recent and beneficial shift in recent years has been the development and public sharing of large-scale brain imaging datasets, whether performed on large populations [4, 5, 6] or on small groups of individuals but with very large amounts of data available [7, 8, 9] and <http://www.cneuromod.ca> – a context known as deep phenotyping. Given the availability of such data, which are only partially or inconsistently annotated [10], the question is: can one identify core structures of these networks that would provide relevant primitives for fitting AI models?

Mission confiée

In this PhD, we aim to build basic models of brain function using large-scale imaging resources, from which a network of components would be identified.

- Components refer to dictionary-like decompositions of brain data, factorising the data into sparse but structured regions as in [11]. These topographies should be multiscale [12] and, unlike previous work, could be provided with a precise semantic specification [13, 14, 10] inherited from the annotations associated with the brain data.
- Network refers to a graphical model underlying the properties of these components, which would describe the interactions that exist between them. The network structure can be learned together with the above components.
- The model should be adaptive to new individuals (personalization): given some new brain imaging data, it should be relaxed to accommodate idiosyncrasies of the new data, based on the known properties of the components and the features visible in the data.

While dictionary learning has already proven to be a powerful tool, the semantic, network and adaptive features of this decomposition are novel. For the adaptive part, we will rely on an optimal transport approach recently developed in our group, which has been shown to be the most powerful in mapping inter-individual variability [15].

Principales activités

The work will start by collecting as many publicly available datasets as possible to enable large-scale learning. We will then develop a multi-resolution dictionary learning strategy similar to [11] to 2k or even 4k components (the most accurate models are currently only 1k). We will augment the learned model with contrastive strategies that produce merging of the data, revealing statistical relationships between models. For this we will draw inspiration from [19, 16]. After this step, we will extend the model with the following features (these tasks can be performed in parallel and are weakly interdependent).

- We will associate semantics with components by investigating stable predictive associations between cognitive annotations of the data and the brain components whose signal predicts the occurrence of these labels across datasets.
- We will learn the dependence structure of the model; this can be summarised by a classical covariance structure, but embeddings produced by contrastive models actually contain an implicit dependence structure that could be captured by conditional distributions of the signals in the model components.
- We will propose a scheme for adapting the model to new individual data based on the model in [15].

Unlike existing tools, this correspondence model should be able to account for all available information: not only image domain signals, but also semantics and connectivity.

Technical developments To this end, we will revisit and extend the large-scale dictionary learning model of [11, 12], and investigate non-linear contrastive variants of these models. We will use large-scale inference tools to learn clean semantic associations from the data.

Validation on brain imaging analysis tasks Validation is a core step of the procedure, necessary to validate any successive improvements to the initial model. It will consist of a selection of several core imaging tasks that represent the expected virtues of the sought brain model: i) good representation of the existing data, whether they come from public image databases not used to build the model, or from the existing literature; ii) perform decoding at scale with good accuracy [10]; iii) fit of the AI model, as measured by the amount of explained variance.

Compétences

Compétences techniques et niveau requis : Maître de Python scientifique, neuroimagerie, apprentissage automatique.

Langues : anglais

Compétences relationnelles : -

Compétences additionnelles appréciées :-

Avantages

- Subsidized meals
- Partial reimbursement of public transport costs
- Leave: 7 weeks of annual leave + 10 extra days off due to RTT (statutory reduction in working hours) + possibility of exceptional leave (sick children, moving home, etc.)
- Possibility of teleworking and flexible organization of working hours
- Professional equipment available (videoconferencing, loan of computer equipment, etc.)
- Social, cultural and sports events and activities
- Access to vocational training
- Social security coverage

Rémunération

Minimum remuneration: 2,100 € gross/month

Order of August 29, 2016 setting the remuneration of contract doctoral students

Informations générales

- **Thème/Domaine** : Neurosciences et médecine numériques
Biologie et santé, Sciences de la vie et de la terre (BAP A)
- **Ville** : Palaiseau
- **Centre Inria** : [Centre Inria de Saclay](#)
- **Date de prise de fonction souhaitée** : 2024-08-01
- **Durée de contrat** : 3 ans
- **Date limite pour postuler** : 2024-07-31

Contacts

- **Équipe Inria** : [MIND](#)
- **Directeur de thèse** :

A propos d'Inria

Inria est l'institut national de recherche dédié aux sciences et technologies du numérique. Il emploie 2600 personnes. Ses 215 équipes-projets agiles, en général communes avec des partenaires académiques, impliquent plus de 3900 scientifiques pour relever les défis du numérique, souvent à l'interface d'autres disciplines. L'institut fait appel à de nombreux talents dans plus d'une quarantaine de métiers différents. 900 personnels d'appui à la recherche et à l'innovation contribuent à faire émerger et grandir des projets scientifiques ou entrepreneuriaux qui impactent le monde. Inria travaille avec de nombreuses entreprises et a accompagné la création de plus de 200 start-up. L'institut s'efforce ainsi de répondre aux enjeux de la transformation numérique de la science, de la société et de l'économie.

L'essentiel pour réussir

The successful candidate will be interested in applications of machine learning and in understanding human cognition. Note that the work will take place in a multidisciplinary environment (physics, neuroscience, computer science, modelling, psychology), neuroscience, computer science, modelling, psychology). Prior experience on deep model is a major asset, as it makes it easier for the candidate to understand the concepts and tools involved. Knowledge of scientific computing in Python (Numpy, Scipy, Pytorch) is required. All the work will be done in Python based on standard machine learning libraries and the Nilearn library for neuroimaging aspects. The candidate will benefit from the numerous development of the Mind and Unicog teams for computational facilities and expertise in the various domains involved (machine learning, optimization, statistics, neuroscience, psychology).

Attention: Les candidatures doivent être déposées en ligne sur le site Inria. Le traitement des candidatures adressées par d'autres canaux n'est pas garanti.

Consignes pour postuler

Sécurité défense :

Ce poste est susceptible d'être affecté dans une zone à régime restrictif (ZRR), telle que définie dans le décret n°2011-1425 relatif à la protection du potentiel scientifique et technique de la nation (PPST). L'autorisation d'accès à une zone est délivrée par le chef d'établissement, après avis ministériel favorable, tel que défini dans l'arrêté du 03 juillet 2012, relatif à la PPST. Un avis ministériel défavorable pour un poste affecté dans une ZRR aurait pour conséquence l'annulation du recrutement.

Politique de recrutement :

Dans le cadre de sa politique diversité, tous les postes Inria sont accessibles aux personnes en situation de handicap.