2022-05177 - Post-Doctoral Research Visit F/M Neural methods for relational data

Contrat renouvelable : Oui
Niveau de diplôme exigé : Thèse ou équivalent
Fonction : Post-Doctorant

A propos du centre ou de la direction fonctionnelle

Located at the heart of the main national research and higher education cluster, member of the Université Paris Saclay, a major actor in the French Investments for the Future Programme (Idex, LabEx, IRT, Equipex) and partner of the main establishments present on the plateau, the centre is particularly active in three major areas: data and knowledge; safety, security and reliability; modelling, simulation and optimisation (with priority given to energy).

The 500 researchers and engineers from Inria and its partners who work in the research centre’s 30 teams, the 60 research support staff members, the high-level equipment at their disposal (image walls, high-performance computing clusters, sensor networks), and the privileged relationships with prestigious industrial partners, all make Inria Saclay Île-de-France a key research centre in the local landscape and one that is oriented towards Europe and the world.

Contexte et atouts du poste

Many data science problems, for instance in health or business, start from relational data, whether it is in explicit relational databases or in a set of tables. The data are not a numerical table, and an important part of the statistical modeling consists in creating a variety of transformation to turn it into numerical vectors: discrete elements are one-hot encoded –though high cardinality needs more sophisticated encoding (Cerda and Varoquaux, 2020); information may be assembled across multiple tables, joining and aggregating on common entities. For instance, a good prediction of housing prices requires assembling various information about the neighborhood –the access to education, transportation, parks, job, shops– more global trends of geographical growth... This information is available spread across multiple source, for instance on multiple internet pages. Crafting all the transformation required to turn these information in numerical vectors requires many manual data preparation steps and is arguably the number one time sink in data science.

In the soda team, we are adapting to relational data modern representation learning tools –those behind the deep learning revolution–, with the specific goal of learning vectorial embeddings of all the information in a database and thus greatly facilitate data preparation for data science.

The soda team is a newly created team doing research at the intersection between machine-learning, databases, and quantitative social sciences (eg empirical economy, epidemiology...). It hosts the team developing scikit-learn at Inria. The team has access to multiple large compute nodes with GPUs, an internal compute cluster, as well as the Jean Zay large supercomputer with GPUs.

This is a joint project with the Dataiku startup, a leading provider of data-science tools. The post-doctoral researcher will be co-hosted between the R&D team of Dataiku and the Soda team at Inria.

Mission confiée

Assignments : The recruited person will work under the direct supervision of Gaël Varoquaux (Soda) and Léo Dreyfus-Schmidt (head of research, Dataiku).

Collaboration : The work will be done within the subgroup of soda working on automating preprocessing and analysis of relational data: 2 engineers, 2 students (soon 3), and wider collaborations with experts on NLP, knowledge bases, and deep learning such as Alexandre Allauzen, Fabian Suchanek, and Edouard Oyallon.

For a better knowledge of the proposed research subject : a detailed scientific description of the research program is available on https://team.inria.fr/soda/job-offers/ . The output of our previous research project on dirty-data is available on https://project.inria.fr/dirtydata/publications/ .

Principales activités

Main activities :

- Design and validate deep learning architectures to capture the information in relational data
- Experiment on data-science tasks to understand the benefits brought by these architecture
- Write publications explaining these progresses

Additional activities :

- Collaborate with data scientists to understand the challenges in a variety of applications (such as health or socio-economic questions)
- Release software demonstrating the methods developed
Compétences

Technical skills and level required: understanding of the workings of deep learning will be valued

Languages: English is the only required language, however a good ability to writing clear, didactic scientific publications is important.

Relational skills: kindness and enthusiasm make happy teams.

Other valued appreciated: Curiosity, and a desire to learn.

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 (after 6 months of employment) 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

2653 €/month (gross salary)

Informations générales

- Thème/Domaine: Optimisation, apprentissage et méthodes statistiques
  Statistiques (Big data) (BAP E)
- Ville: Palaiseau
- Centre Inria: CRI Saclay - Île-de-France
- Date de prise de fonction souhaitée: 2022-10-01
- Durée de contrat: 2 ans
- Date limite pour postuler: 2022-07-29

Contacts

- Equipe Inria: SODA
- Recruteur: Varoquaux Gael / Gael.Varoquaux@inria.fr

A propos d'Inria

Inria est l'institut national de recherche dédié aux sciences et technologies du numérique. Il emploie 2600 personnes. Ses 200 équipes-projets agiles, en général communes avec des partenaires académiques, impliquent plus de 3500 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 180 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

- A strong ability for numerical experimentation: large-scale empirical validation of data-processing pipeline, exploration of results
- Programming skills in implementing algorithms (eg numerical computing, or matching)
- Knowledge of machine learning or applied maths (mathematical optimization and statistics)
- Familiarity with fitting deep neural networks (typically pytorch)
- Good paper-writing skills, in English
- Desire and skills in helping to supervised more junior researchers

Candidates without all these, but a strong desire and ability to learn, will also be considered.

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.

**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.