



Offre n°2024-08415

Stage de Master 2 : Développement d'un modèle à blocks latents profond pour le co-clustering

Type de contrat : Stage

Niveau de diplôme exigé : Bac + 5 ou équivalent

Fonction : Stagiaire de la recherche

A propos du centre ou de la direction fonctionnelle

Le centre Inria d'Université Côte d'Azur regroupe 42 équipes de recherche et 9 services d'appui. Le personnel du centre (500 personnes environ) est composé de scientifiques de différentes nationalités, d'ingénieurs, de techniciens et d'administratifs. Les équipes sont principalement implantées sur les campus universitaires de Sophia Antipolis et Nice ainsi que Montpellier, en lien étroit avec les laboratoires et les établissements de recherche et d'enseignement supérieur (Université Côte d'Azur, CNRS, INRAE, INSERM ...), mais aussi avec les acteurs économiques du territoire.

Présent dans les domaines des neurosciences et biologie computationnelles, la science des données et la modélisation, le génie logiciel et la certification, ainsi que la robotique collaborative, le Centre Inria d'Université Côte d'Azur est un acteur majeur en termes d'excellence scientifique par les résultats obtenus et les collaborations tant au niveau européen qu'international.

Contexte et atouts du poste

Le stage proposé s'inscrit dans le contexte du co-clustering qui consiste à regrouper simultanément les lignes et les colonnes d'un tableau de données [1], ce qui est particulièrement utile pour résumer de grands ensembles de données (voir la figure 1). Un modèle de co-clustering probabiliste populaire est le latent block model [3] (LBM), il suppose que les clusters dans chaque ligne et chaque colonne sont tirés indépendamment de deux distributions multinomiales et que, compte tenu de ces clusters, toutes les entrées du tableau de données sont indépendantes, et que chaque entrée suit une distribution qui dépend uniquement de ses clusters dans la ligne et dans la colonne. Dans le stage, nous proposons de développer une extension du LBM dans le cas de données binaires en supposant que chaque ligne et chaque colonne peut être codée par une position latente dans un espace euclidien et que le paramètre de la distribution de chaque entrée ne dépend que de ces positions latentes, comme dans le cas de [5]. Ce modèle permettra d'effectuer à la fois le co-clustering et la visualisation des données à travers les positions latentes comme dans [2]. Pour l'inférence des paramètres, nous envisagerons une approche variationnelle comme dans [2] en utilisant une architecture de réseau neuronal pour la distribution postérieure approximative des variables latentes.

Références

[1] Christophe Biernacki, Julien Jacques, and Christine Keribin. A survey on model-based co-clustering: High dimension and estimation challenges. 2022.

[2] Rémi Boutin, Pierre Latouche, and Charles Bouveyron. The deep latent position topic model for clustering and representation of networks with textual edges, 2024.

[3] Vincent Brault and Mahendra Mariadassou. Co-clustering through latent bloc model: A review. *Journal de la Société Française de Statistique*, 156(3):120–139, 2015.

[4] Geert Govaert and Mohamed Nadif. Block clustering with bernoulli mixture models: Comparison of different approaches. *Computational Statistics and Data Analysis*, 52(6):3233–3245, 2008.

[5] Mark S Handcock, Adrian E Raftery, and Jeremy M Tantrum. Model-based clustering for social networks. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 170(2):301–354, 2007.

Mission confiée

La mission principale du stage sera d'écrire le modèle mathématique et ses paramètres d'inférence, et d'effectuer ses implémentations sur Python. De plus, la précision de la méthodologie proposée sera également étudiée sur des données réelles. Un sujet de thèse pourra être proposé à la suite de ce stage.

Principales activités

- Recherche bibliographique
- Calculs mathématiques
- Rédaction de l'état de l'art
- Programmation
- Interprétation des résultats

Compétences

Compétences techniques et niveau requis :

- Programmation Python : Niveau avancé, avec une expérience des bibliothèques telles que NumPy, Pandas, Scikit-learn, PyTorch ou TensorFlow.
- Expérience avec les frameworks et outils de machine learning.
- Connaissances en modélisation statistique, techniques d'optimisation et prétraitement des données.

Langues :

- Anglais : Niveau professionnel (pour la documentation et la collaboration dans une équipe internationale).
- Français : Facultatif mais apprécié.

Compétences relationnelles :

- Capacité à travailler en collaboration au sein d'une équipe multidisciplinaire.
- Esprit critique et aptitude à résoudre des problèmes de manière proactive.
- Bonnes compétences en communication pour présenter des résultats et rédiger des rapports.

Autres compétences appréciées :

- Expérience préalable dans des projets de recherche ou des stages en IA/apprentissage automatique.
- Intérêt pour la contribution à des projets open source.

Avantages

- Restauration subventionnée
- Transports publics remboursés partiellement
- Congés: 7 semaines de congés annuels + 10 jours de RTT (base temps plein) + possibilité d'autorisations d'absence exceptionnelle (ex : enfants malades, déménagement)
- Possibilité de télétravail (après 6 mois d'ancienneté) et aménagement du temps de travail
- Équipements professionnels à disposition (visioconférence, prêts de matériels informatiques, etc.)
- Prestations sociales, culturelles et sportives (Association de gestion des œuvres sociales d'Inria)
- Accès à la formation professionnelle
- Participation mutuelle (sous conditions)

Rémunération

Gratification selon temps de présence.

Informations générales

- **Thème/Domaine** : Optimisation, apprentissage et méthodes statistiques Statistiques (Big data) (BAP E)
- **Ville** : Sophia Antipolis
- **Centre Inria** : [Centre Inria d'Université Côte d'Azur](#)
- **Date de prise de fonction souhaitée** : 2025-03-01
- **Durée de contrat** : 6 mois
- **Date limite pour postuler** : 2025-02-28

Contacts

- **Équipe Inria** : [MAASAI](#)
- **Recruteur** :
Vandewalle Vincent / Vincent.Vandewalle@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 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.

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

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

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.