Offre n°2024-07913

Post-Doctoral Research Visit F/M Fast generation of approximate galaxy maps with topology-infused generative models

Le descriptif de l'offre ci-dessous est en Anglais

Type de contrat : CDD

Niveau de diplôme exigé : Thèse ou équivalent

Fonction : Post-Doctorant

A propos du centre ou de la direction fonctionnelle

The Inria center at Université Côte d'Azur includes 42 research teams and 9 support services. The center's staff (about 500 people) is made up of scientists of different nationalities, engineers, technicians and administrative staff. The teams are mainly located on the university campuses of Sophia Antipolis and Nice as well as Montpellier, in close collaboration with research and higher education laboratories and establishments (Université Côte d'Azur, CNRS, INRAE, INSERM …), but also with the regional economic players.

With a presence in the fields of computational neuroscience and biology, data science and modeling, software engineering and certification, as well as collaborative robotics, the Inria Centre at Université Côte d'Azur is a major player in terms of scientific excellence through its results and collaborations at both European and international levels.

Contexte et atouts du poste

The research will be supervised by Mathieu Carrière (DataShape, Centre Inria d'Université Côte d'Azur), with regular meetings organized with collaborators in Area Science Park, Trieste, Italy. The candidate will be located principally at Inria.

Mission confiée

Any candidate should have a PhD degree in applied mathematics and/or computer science, excellent programming skills with an extensive knowledge of the principal data science libraries for (deep) machine learning and optimization (such as Scikit-Learn, TensorFlow, PyTorch) and a lot of research experience on topics including deep generative models, machine learning, dimensionality reduction and non-convex optimization. Some experience in geometric and topological methods in machine learning, as well as approximate cosmological simulations, is a plus. Teamworking, communication and collaboration skills are also essential.

Prerequisite

Academic level: PhD

Place of work: Centre Inria d'Université Côte d'Azur

Language: English (fluent), French (optional)

Libraries: Scikit-Learn, PyTorch/TensorFlow, Gudhi (optional)

Principales activités

The goal of this postdoctoral research position is to develop the theoretical interactions between Topological Data Analysis (TDA) and Deep Generative Models (DGM), and its applications in approximate simulations of galaxy maps from cosmology.

TDA [14] is a field of data science that gained a lot of attention during the last few years due to its ability to generate new descriptors and features for data sets (such as persistence diagrams and Mapper complexes), which encode **topological information**--

-for instance, the number and sizes of topological structures in the data, including, e.g., connected components, loops and cavities---that is usually complementary to traditional features, and can thus greatly improve performances of standard methods in Machine Learning. Moreover, as TDA is firmly rooted in algebraic topology, its descriptors enjoy many properties, such as being invariant to any continuous deformations, as well as being robust to potentially non-linear noise, as proved by the celebrated stability theorem [13].

A recent breakthrough in TDA [1,2,3] is the ability to **differentiate** through topological descriptors, using an appropriate notion of (sub)gradient. This has found many applications, including the definition of new losses based on topology for training neural networks [4], or the automatic optimization of TDA parameters (instead of using fixed ones a priori) [5], but is currently limited to one TDA descriptor only, the so-called *persistence diagram*.

Following this line of work, a promising and exciting yet largely unknown area is the design of deep generative models (variational autoencoders, diffusion models, generative adversarial networks, etc) that are **based on topology**. These would allow to generate data that satisfy complex topological priors, such as constraints on the numbers and sizes of topological structures. Except from very preliminary work [6], this question remains largely open, as most other geometric generative models can only reproduce simple topologies (torus [7], sphere [8], hyperbolic space [9]).

Designing such generative models would matter particularly in the analysis of galaxy maps from cosmological data, which has also gone through tremendous changes recently, due to the fast developments of approximate simulation models, that aim at reproducing dark matter density maps of costly full-fledge physical simulations. Indeed, most approximation models suffer from lack of accuracy: state-of-the-art approximate simulation models, such as Pinocchio [10], only aim at reproducing *statistical summaries* computed on the observations, which usually take the form of low-order correlation functions that capture only part of the geometry of real galaxy maps. On the other hand, TDA descriptors have appeared recently as powerful summaries that can capture non-Gaussianities [11, 12], and that are not as sensitive to deformation and noise induced by experimental biases than standard descriptors.

Hence, the successful candidate will work on developing mathematical frameworks and practical setups for applying TDA methods, descriptors and features in the context of deep generative models. A particular focus will be made first on *multimodal variational autoencoders*, with specific reparameterization tricks adapted to topology (see an example of such model below).

More generally, the candidate will also study the theoretical differentiability properties of TDA descriptors based on *multi-parameter persistence* (a recent breakthrough of TDA), and its actual implementation and comparison against other generative models that use geometry. Then, the models will be applied to the fast generation of approximate yet accurate galaxy maps, and used for constraining the physical parameters corresponding to real observations by incorporating these parameters in the generative models directly.

[1] Mathieu Carrière, Frédéric Chazal, Yuichi Ike, Théo Lacombe, Martin Royer, and Yuhei Umeda.

PersLay: a neural network layer for persistence diagrams and new graph topological signatures.

In 23rd International Conference on Artificial Intelligence and Statistics (AISTATS 2020), pages 2786–2796. PMLR, 2020.

[2] Mathieu Carrière, Frédéric Chazal, Marc Glisse, Yuichi Ike, Hariprasad Kannan, and Yuhei Umeda.

Optimizing persistent homology based functions.

In 38th International Conference on Machine Learning (ICML 2021), volume 139, pages 1294–1303. PMLR, 2021.

[3] Jacob Leygonie, Mathieu Carrière, Théo Lacombe, and Steve Oudot.

A gradient sampling algorithm for stratified maps with applications to topological data analysis.

Mathematical Programming, 2023.

[4] Chao Chen, Xiuyan Ni, Qinxun Bai, and Yusu Wang.

A topological regularizer for classifiers via persistent homology.

In 22nd International Conference on Artificial Intelligence and Statistics (AISTATS 2019), volume 89, pages 2573–2582. PMLR, 2019.

[5] Max Horn, Edward de Brouwer, Michael Moor, Bastian Rieck, and Karsten Borgwardt.

Topological Graph Neural Networks.

In 10th International Conference on Learning Representations (ICLR 2022). OpenReviews.net, 2022.

[6] Thibault de Surrel, Felix Hensel, Mathieu Carrière, Théo Lacombe, Yuichi Ike, Hiroaki Kurihara, Marc Glisse, and Frédéric Chazal.

Ripsnet: A general architecture for fast and robust estimation of the persistent homology of point clouds.

In Topological, Algebraic, and Geometric Learning Workshop, volume 196, pages 96–106. PMLR, 2022.

[7] Danilo Jimenez Rezende, George Papamakarios, Sebastien Racaniere, Michael Albergo, Gurtej Kanwar, Phiala Shanahan, and Kyle Cranmer.

Normalizing flows on tori and spheres.

In 37th International Conference on Machine Learning (ICML 2020), volume 119, pages 8083–8092. PMLR, 2020.

[8] Sung Woo Park and Junseok Kwon.

Sphere generative adversarial network based on geometric moment matching.

In 32nd IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2019), pages 4287–4296. IEEE Computer Society, 2019.

[9] Joey Bose, Ariella Smofsky, Renjie Liao, Prakash Panangaden, and Will Hamilton.

Latent variable modelling with hyperbolic normalizing flows.

In 37th International Conference on Machine Learning (ICML 2020), volume 119, pages 1045–1055. PMLR, 2020.

[10] P. Monaco, E. Sefusatti, S. Borgani, and al.

An accurate tool for the fast generation of dark matter halo catalogues.

Monthly Notices of the Royal Astronomical Society, 433(3):2389–2402, 2013.

[11] Matteo Biagetti, Alex Cole, and Gary Shiu.

The persistence of large scale structures. part i. primordial non-gaussianity.

Journal of Cosmology and Astroparticle Physics, 2021(4), 2021.

[12] Matteo Biagetti, Juan Calles, Lina Castiblanco, Alex Cole, and Jorge Noreña.

Fisher forecasts for primordial non-gaussianity from persistent homology.

Journal of Cosmology and Astroparticle Physics, 2022(10), 2022.

[13] Frédéric Chazal, Vin de Silva, Marc Glisse, and Steve Oudot.

The structure and stability of persistence modules.

SpringerBriefs in Mathematics. Springer-Verlag, 2016.

[14] Frédéric Chazal and Bertrand Michel.

An introduction to topological data analysis: Fundamental and practical aspects for data scientists.

Frontiers in Artificial Intelligence, 4, 2021.

Compétences

Technical skills and level required :

Languages :

Relational skills :

Other valued appreciated :

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
- Contribution to mutual insurance (subject to conditions)

Rémunération

Gross Salary: 2788 € per month

Informations générales

- **Thème/Domaine** : Algorithmique, calcul formel et cryptologie
  Statistiques (Big data) (BAP E)
- **Ville** : Sophia Antipolis
- **Centre Inria** : Centre Inria d'Université Côte d'Azur
- **Date de prise de fonction souhaitée** : 2024-09-01
- **Durée de contrat** : 2 ans
- **Date limite pour postuler** : 2024-07-27

Contacts

- **Équipe Inria** : DATASHAPE
- **Recruteur** :
  Carrière Mathieu / mathieu.carriere@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.

L'essentiel pour réussir

There you can provide a "broad outline" of the collaborator you are looking for what you consider to be necessary and sufficient, and which may combine :

- tastes and appetencies,
- area of excellence,
- personality or character traits,

- cross-disciplinary knowledge and expertise...

This section enables the more formal list of skills to be completed and 'lightened' (reduced) :

- "Essential qualities in order to fulfil this assignment are feeling at ease in an environment of scientific dynamics and wanting to learn and listen."
- " Passionate about innovation, with expertise in Ruby on Rails development and strong influencing skills. A thesis in the field of **** is a real asset."

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.