Offre n°2024-08249

Internship: Generative methods for online adaptive deep learning training

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

Type de contrat : Convention de stage

Niveau de diplôme exigé : Bac + 4 ou équivalent

Fonction : Stagiaire de la recherche

A propos du centre ou de la direction fonctionnelle

The Centre Inria de l'Université de Grenoble groups together almost 600 people in 23 research teams and 9 research support departments.

Staff is present on three campuses in Grenoble, in close collaboration with other research and higher education institutions (Université Grenoble Alpes, CNRS, CEA, INRAE, …), but also with key economic players in the area.

The Centre Inria de l'Université Grenoble Alpe is active in the fields of high-performance computing, verification and embedded systems, modeling of the environment at multiple levels, and data science and artificial intelligence. The center is a top-level scientific institute with an extensive network of international collaborations in Europe and the rest of the world.

Contexte et atouts du poste

Advisers: Bruno Raffin (Bruno.Raffin@inria.fr), Sofya Dymchenko (sofya.dymchenko@inria.fr)

The internship will take place at the DataMove team located in the IMAG building on the campus of Saint Martin d'Heres (Univ. Grenoble Alpes) near Grenoble. The length of the internship is *4 months minimum* and the start date is flexible, but need a 2 month delay before starting the interhsip due to administrative constraints. The DataMove team is a friendly and stimulating environment that gathers Professors, Researchers, PhD and Master students all leading research on High-Performance Computing. The city of Grenoble is a student-friendly city surrounded by the Alps mountains, offering a high quality of life and where you can experience all kinds of mountain-related outdoor activities.

Mission confiée

Subject context

n supervised learning, successfully training advanced neural networks requires annotated data of sufficient quantity and quality. In natural sciences (physics, chemistry, weather modeling), observational data remains to be a limiting factor. One alternative is to numerically create synthetic training data. This offers several advantages: synthetic data can be generated at will, in potentially unlimited amounts, the quality can be degraded in a controlled manner for more robust trainings, and the coverage of the parameter space can be adapted to focus training where relevant. Today, a large variety of simulation codes to create such data are available, from computer graphics, computer engineering, computational physics, biology and chemistry, and so on. When training data is produced from simulation codes, it can be generated along with the training.

This approach has multiple benefits. First, there is no need to store and move a huge pre-created data set: float matrices of data can take terrabytes of memory, and reading them from the disk every training iteration might take more time than the iteration itself. Instead, data is stored in working memory and created "on-the-fly": when new data point is created it substitutes an old one. This allows the model to see terrabytes of data throughout its lifetime while storing only a smaller part of it at a time. Second, the training is not done with the same repeated data as in epoch-based approach. Continiously updated training set potentially improves the generalization quality of the model. More importantly, *the update of the training set and creation of new data can be adaptive*, driven by the observed behavior of the neural network during training. However, this adaptive data generation is a challenging question.

Active learning adresses this challenge by adaptively sampling the input parameters of simulators based on training progress, aiming to generate more relevant data. Thus, faster and higher-quality training is expected. In current approaches, active learning for simulations-based training often follows a phased algorithm: 1) generate an initial training set by uniformly sampling input points 2) (re)train the model on the trainng set 3) use feedback from the model's performance to generate/augment new training set and return to (2). Fundamentally, the methods differentiate by choice of "feedback" metric (aquisition

function) and the way the next training set is created (aquisition algorithm).

Our research

Our team's research is focused on exploring and developing new online active learning methods for efficient training of surrogates -- neural networks that meant to substitute simulation codes. We have developped *Breed* for online adaptive surrogate training, such as Physics Informed Neural Networks (PINNs), Neural Operators, and basic Dense Neural Networks, within our *MelissaDL* framework that allows the training to be highly distributed and the training data to be created on-the-fly.

Our related publications

- MelissaDL x Breed: Towards Data-Efficient On-line Supervised Training of Multi-parametric Surrogates with Active Learning, SC AI4S 2024: https://hal.science/hal-04712480v1
- Training Deep Surrogate Models with Large Scale Online Learning, ICML 2023: https://hal.science/hal-04102400v1
- Loss-driven sampling within hard-to-learn areas for simulation-based neural network training, NeurIPS ML4Phys 2023: https://hal.science/hal-04305233v1
- Melissa: Simulation-Based Parallel Training, NeurIPS AI4S 2022:https://hal.science/hal-03842106v1

Principales activités

This intership is focused on investigating use of generative methods for active learning, e.g., diffusion posterior sampling to generate input points based on models uncertainty. Currently, Breed method uses importance sampling technique and loss statistics.
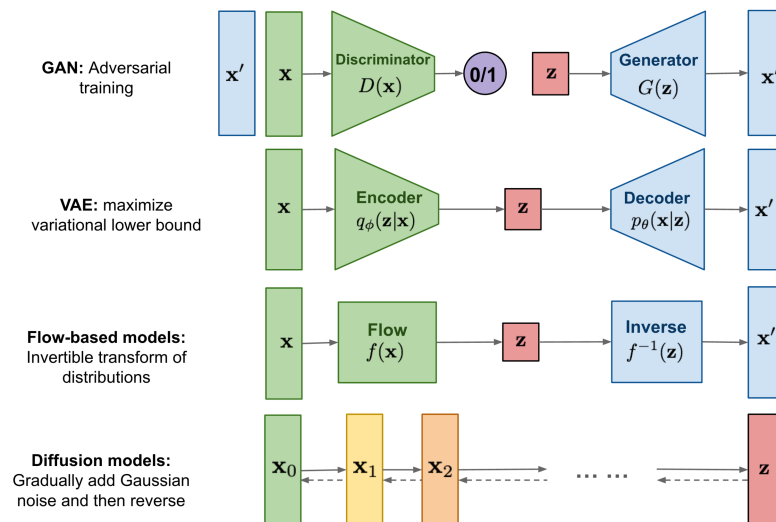
In the beginning, the objective is to get familiar with the domain and read about existing work: surrogates, neural operators, active learning, online training, Bayesian methods. Then -- start to work on possible generative methods for active learing (**normalizing flows, diffusion models, generative-adversarial networks, energy-based models, etc.**), develop and evaluate their performance through experiments with use cases such as heat equation and fluid dynamics equations. Currently, we work in a team consisting of a PhD student, a research engineer and a research director (Bruno Raffin), we have regular meetings and a daily communication - you will not be alone!

The perfect candidate has basic knowledge of generative deep learning, confident programming skills to develop ML/DL algorithms in Python, motivation to quickly learn new things, and, most importantly, an interest to application of AI to physical sciences!

Related papers

- Population Monte Carlo with Normalizing Flow.https://arxiv.org/abs/2312.03857
- All-in-one simulation-based inference. https://arxiv.org/abs/2404.09636
- Adaptive Generation of Training Data for ML Reduced Model Creation. https://www.osti.gov/biblio/1923172
- A comprehensive study of non-adaptive and residual-based adaptive sampling for physics-informed neural networks. https://arxiv.org/abs/2207.10289
- Mitigating Propagation Failures in Physics-informed Neural Networks using Retain-Resample-Release (R3) Sampling. https://arxiv.org/abs/2207.02338
- Deep Active Learning by Leveraging Training Dynamics.https://arxiv.org/abs/2110.08611

Compétences

Technical skills: Python (numpy, pytorch), Git, Jupyter notebooks.

The main communication language is English.

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 (90 days / year) and flexible organization of working hours
- Social, cultural and sports events and activities
- Access to vocational training
- Social security coverage under conditions

Rémunération

€4.35 per hour of actual presence at 1 January 2024.

About 590€ gross per month (internship allowance)

Informations générales

- **Thème/Domaine** : Optimisation, apprentissage et méthodes statistiques
  Calcul Scientifique (BAP E)
- **Ville** : Saint Martin d'Hères
- **Centre Inria** : Centre Inria de l'Université Grenoble Alpes
- **Date de prise de fonction souhaitée** : 2025-02-01
- **Durée de contrat** : 6 mois
- **Date limite pour postuler** : 2024-11-21

Contacts

- **Équipe Inria** : DATAMOVE
- **Recruteur** :
  Dymchenko Sofya / sofya.dymchenko@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

The perfect candidate has basic knowledge of probabilities, deep learning, physics, confident programming skills to develop ML/DL algorithms in Python, motivation to quickly learn new things, and, most importantly, an interest to application of AI to physical sciences!

> **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

CV + cover letter

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

Applications must be submitted online via the Inria website. Processing of applications submitted via other channels is not guaranteed.

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.