

2022-05133 - Energy-aware Machine Learning Training

Type de contrat : CDD
Niveau de diplôme exigé : Bac + 5 ou équivalent
Fonction : Ingénieur scientifique contractuel

A propos du centre ou de la direction fonctionnelle

The Inria Université Côte d'Azur center counts 36 research teams as well as 7 support departments. The center's staff (about 500 people including 320 Inria employees) is made up of scientists of different nationalities (250 foreigners of 50 nationalities), engineers, technicians and administrative staff. 1/3 of the staff are civil servants, the others are contractual agents. The majority of the center's research teams are located in Sophia Antipolis and Nice in the Alpes-Maritimes. Four teams are based in Montpellier and two teams are hosted in Bologna in Italy and Athens. The Center is a founding member of Université Côte d'Azur and partner of the I-site MUSE supported by the University of Montpellier.

Contexte et atouts du poste

Deep neural networks have enabled impressive accuracy improvements across many machine learning tasks. Often the highest scores are obtained by the most computationally-hungry models [1]. As a result, training a state-of-the-art model now requires substantial computational resources which demand considerable energy, along with the associated economic and environmental costs. Research and development of new models multiply these costs by thousands of times due to the need to try different model architectures and different hyper-parameters.

A recent paper [2] has estimated the amount of energy and the corresponding CO2 emissions required to train different models.

For example, the full neural architecture search described in [1] to train a big transformer model for machine translation is estimated to have consumed 650 kWh and generated the equivalent of 284 tons of CO2.

As a comparison, the average American citizen produces 16 tons of CO2 per year and a New York City-San Francisco round-trip flight of a Boeing 777 with 300 passengers produces 260 tons. As the role of AI becomes more pervasive in our society, its sustainability needs to be addressed.

- References:

[1] D. R. So, C. Liang, and Q. V. Le, The evolved transformer, 36th Intl. Conference on Machine Learning (ICML), 2019.

[2] E. Strubell, A. Ganesh, and A. McCallum, Energy and Policy Considerations for Deep Learning in NLP, Annual Meeting of the Association for Computational Linguistics (ACL), 2019.

Mission confiée

As machine learning training operations are now often distributed across multiple computation nodes, the development of tools to assign computation tasks to the most sustainable node within a pool is an important direction to explore. In this project we would like to implement several codes and script in order to automatically:

- 1) Evaluate a node's energy consumption and GHG emissions
- 2) Configure a distributed deep learning training framework to wisely assign a training task to the right nodes
- 3) Stop and start computing node while not in use

As part of the project, it may be considered integration with external components or algorithms to provide inputs about preferred nodes. This may include a forecast of the node's GHG or availability evolution

Pre-existing libraries or API may be used during the project such as:

- Apache Spark™ - Unified Engine for large-scale data analytics
- electricityMap API Documentation
- Scaphandre
- Cloud Carbon Footprint - An open source tool to measure and analyze cloud carbon emissions
- Carbon Footprint Evaluation from Cloud providers: AWS / GCP / Azure

Principales activités

- install and deploy software for distributed machine learning training
- design and carry on experiments to evaluate energy consumption of machine learning training
- write reports
- participate to the preparation of scientific papers

This offer is part of a collaboration between the NEO research team and the company Accenture Labs based in Sophia Antipolis. The candidate will be co-supervised, and hosted mainly at Accenture Labs for the duration of the project

Compétences

Good programming skills and knowledge of Unix systems.

Working 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 (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

Informations générales

- **Thème/Domaine** : Systèmes distribués et intergiciels
Système & réseaux (BAP E)
- **Ville** : Sophia Antipolis
- **Centre Inria** : CRI Sophia Antipolis - Méditerranée
- **Date de prise de fonction souhaitée** : 2022-10-01
- **Durée de contrat** : 2 ans
- **Date limite pour postuler** : 2022-08-31

Contacts

- **Equipe Inria** : NEO
- **Recruteur** :
Neglia Giovanni / Giovanni.Neglia@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.

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.

- Access to vocational training
- Social security coverage