2022-05132 - Post-Doctoral Research Visit F/M Energy-Aware Federated Learning

Contract type: Fixed-term contract
Level of qualifications required: PhD or equivalent
Function: Post-Doctoral Research Visit

About the research centre or Inria department

The Inria Université Côte d’Azur center counts 38 research teams as well as 7 support departments. The center’s staff (about 320 Inria employees) is made up of scientists of different nationalities (250 foreigners) and 50 nationalities), engineers, technicians and administrative staff. 13 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.

Context

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. The development of new low-energy hardware accelerators is an important direction to explore, and neuromorphic hardware for spiking neural networks [3] or new light-based hardware [4] are definitely interesting solutions. In this project, we investigate a more algorithmic/system-level approach to reduce energy consumption for distributed ML training over the Internet.

- References
[4] LightNn, AI for all, everywhere. NLP Extreme scale AI usable through the Muse API (lighton.ai), see their list of publications.

General Information
- Theme/Domain : Optimization, machine learning and statistical methods
- System & Networks (BAP E)
- TownCity : Sophia Antipolis
- Inria Center : CRI Sophia Antipolis - Méditerranée
- Starting date : 2022-10-01
- Duration of contract: 2 years
- Deadline to apply: 2022-07-24

Contacts
- Inria Team: NEO
- Recruiter: Neglia Giovanni / giovanni.neglia@inria.fr

About Inria

Inria is the French national research institute dedicated to digital science and technology. It employs 2,600 people. Its 200 agile project teams, generally run jointly with academic partners, include more than 3,500 scientists and engineers working to meet the challenges of digital technology, often at the interface with other disciplines. The Institute also employs numerous talents in over forty different professions. 900 research support staff contribute to the preparation and development of scientific and entrepreneurial projects that have a worldwide impact.

The keys to success

We are looking for a candidate with a strong background on optimization or energy-aware computing platforms.

Instruction to apply

Defence Security : This position is likely to be situated in a restricted area (ZRR), as defined in Decree No. 2011-1425 relating to the protection of national scientific and technical potential (PPST). Authorisation to enter an area is granted by the director of the unit, following a favourable Ministerial decision, as defined in the decree of 3 July 2012 relating to the PPST. An unfavourable Ministerial decision in respect of a position situated in a ZRR would result in the cancellation of the appointment.

Recruitment Policy :
As part of its diversity policy, all Inria positions are accessible to people with disabilities.

Warning : you must enter your e-mail address in order to save your application to Inria. Applications must be submitted online on the Inria website. Processing of applications sent from other channels is not guaranteed.
Assignment

The question we ask ourselves is:

given a set of available geographically distributed computing units with different energy efficiency and a different mix of energy sources to exploit, how many resources should we allocate and where?

These decisions need to be taken dynamically, as the availability of renewable energies from the sun or the wind changes over short timescales and the amount of resources needed may be a function of the current algorithmic progress of the optimization algorithm (see e.g. [5]).

In particular, we will consider consensus-based distributed optimization approaches [6]. They differ from the usual parameter server framework because each computing unit

1) keeps updating a local version of the parameters and
2) broadcasts its updates only to a subset of nodes (its neighbors).

The remarkable advantages of consensus methods are their flexibility to select the communication topology and to allow some computing nodes to participate only occasionally in the training. These features allow us to reduce the energy footprint of ML training by reducing the amount of communications and activating some computing units only when it is needed.

Giovanni Neglia has started exploring the trade-off between convergence time of consensus methods and communication requirements in [7,8,9] and load balancing among micro-datacenters powered by renewable energy sources in [10]. He has also worked on the control of electrical loads in smart grids [14,15].

Main activities

Working toward publications.

It is possible to be involved in PhD and master students' supervision.

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.

Skills

The working language is English.

Benefits package

- 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

Remuneration

Gross Salary: 2653 € per month