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. But in this project, we investigate a more algorithmic/system-level approach to reduce energy consumption for distributed ML training over the internet.

- References

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

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].

Principales activités

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.

Compétences

The 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
- Access to vocational training
- Social security coverage

Rémunération

Gross Salary: 2653 € per month