Internship: Adaptive Optimization for Decentralized Deep Neural Networks training

Keywords: Decentralized deep learning, meta-learning, hyperparameter optimization
Duration, salary, starting date: max. 6 months, min. 600€ per month. The internship could start as soon as possible and should be extended to a fully funded PhD.
Research teams: MLIA, ISIR (Sorbonne University and CNRS)
Adviser: Edouard Oyallon (CNRS, Sorbonne University)
Location: Jussieu, Sorbonne University, Paris
Contact: `edouard.oyallon@cnrs.fr`
Application: Interested candidate should send CV, grade transcript and potentially one or two referees who could provide a recommendation letter.

Context: Decentralized deep learning systems [4] have gained significant attention due to their ability to leverage distributed computational resources. To fully realize the potential of such systems, it is crucial to develop effective techniques for adaptive optimization and resource management, which will refine the hyper-parameters of a model. This internship aims to explore advanced meta-learning techniques and develop novel algorithms for hyper-parameters, optimizing resource allocation and communication bandwidth of complex communication topology in time-varying settings.

Proposed work: The intern will focus on the following tasks during the internship period: **(a)** Conduct a literature review on existing meta-learning techniques, such as gradient-based meta-learning [1], Bayesian optimization [5], and reinforcement learning [3], to identify hyperparameters for optimal and local learning rate schedulers, update rules, and communication strategies. A specific focus on local learning method [2] will be of high interest for this internship. **(b)** Develop adaptive load balancing, straggler mitigation, and task-aware scheduling algorithms to optimize resource allocation and communication bandwidth among nodes in decentralized deep learning systems. **(c)** Implement and test these algorithms on various decentralized deep learning scenarios and evaluate their effectiveness in improving overall system efficiency and performance. **(d)** Document the research findings and contribute to writing research papers or reports.

An ideal candidate should:

- Be at the Master 2 level (or equivalent),
- Have a strong background in applied mathematics.
- Have strong programming skills in Python and experience with Pytorch
- Be familiar with distributed computing and optimization techniques.

Fundings This project is supported by Project ADONIS.

[1] Marcin Andrychowicz, Misha Denil, Sergio Gomez, Matthew W Hoffman, David Pfau, Tom Schaul, Brendan Shillingford, and Nando De Freitas. Learning to learn by gradient descent by gradient descent. *Advances in neural information processing systems*, 29, 2016.

[2] Eugene Belilovsky, Michael Eickenberg, and Edouard Oyallon. Decoupled greedy learning of cnns. In *International Conference on Machine Learning*, pages 736–745. PMLR, 2020.

[3] Hadi S Jomaa, Josif Grabocka, and Lars Schmidt-Thieme. Hyp-rl: Hyperparameter optimization by reinforcement learning. *arXiv preprint arXiv:1906.11527*, 2019.

[4] Anastasia Koloskova, Nicolas Loizou, Sadra Boreiri, Martin Jaggi, and Sebastian Stich. A unified theory of decentralized sgd with changing topology and local updates. In *International Conference on Machine Learning*, pages 5381–5393. PMLR, 2020.

[5] Jia Wu, Xiu-Yun Chen, Hao Zhang, Li-Dong Xiong, Hang Lei, and Si-Hao Deng. Hyperparameter optimization for machine learning models based on bayesian optimization. *Journal of Electronic Science and Technology*, 17(1):26–40, 2019.