Internship on "Federated Learning for Heterogeneous Model Architectures"

Supervisor

- Giovanni Neglia, Inria, France, giovanni.neglia@inria.fr, http://www-sop.inria.fr/members/Giovanni.Neglia/
- Chuan Xu, Université Côte d'Azur, France, chuan.xu@inria.fr, https://sites.google.com/view/chuanxu

Location

Inria NEO team (https://team.inria.fr/neo/)

2004 route des Lucioles,

06902 Sophia Antipolis, France

Context

The increasing size of data generated by smartphones and IoT devices motivated the development of Federated Learning (FL) [1,2], a framework for on-device collaborative training of machine learning models. FL algorithms like FedAvg [3] allow clients to train a common global model without sharing their personal data. FL reduces data collection costs and protects clients' data privacy and, in doing so, makes possible to train models on large datasets that would otherwise have been inaccessible. FL is currently used by many big data companies (e.g., Google, Apple, Facebook) for learning on their users' data, but the research community envisions also promising applications to learning across large data-silos, like hospitals that cannot share their patients' data [4].

Most existing algorithms for federated learning train the same model architecture for each user/device (personalized FL algorithms [5-8] allow only the value of model parameters to be different), but, heterogeneous clients may not be able to run the same model (because of computational, memory, or battery constraints) and training the largest model that can be supported by the least capable device can be too restrictive.

Goals

In this internship, we will consider that heterogeneous devices may not be able to run the same model (because of computational, memory, or battery constraints) and will propose new algorithms to train heterogeneous architectures jointly. A key challenge is to design meaningful ways of sharing information across heterogeneous model architectures.

We will consider training approaches based on model subsampling. There is a single neural network architecture. However, while powerful clients store and use the complete model, less powerful clients only store a subset of the model matching their memory and computation capabilities.

This approach has been proposed in [9,10] for convolutional networks, where each client may use a different number of channels. We will explore to what extent this strategy can be extended to different neural network architectures. We will also address open issues about training with subsampling. For example, results in [9] suggest that, at some training iterations, more powerful clients should behave as less powerful clients and update a subset of the architecture. We are missing a solid explanation for these observations and quantitative guidelines about how often clients should train a model different from the largest one they can train. The response likely depends on the distribution of clients' capabilities and their dataset sizes. Moreover, these findings seem to be at odds with other empirical observations suggesting that the subnetworks considered by the different clients should rather be disjoint [11].

The internship will start with an experimental approach. The intern will reproduce results in [9,10,11], try to reconcile their apparent contradictions, and derive practical guidelines for FL training of heterogeneous architectures. He/she will then move to study the problem from an analytical point of view relying on the theory of statistical learning, and ensemble learning in particular.

Pre-requisites

The candidate should have a strong background on machine learning and practical experience with software like PyTorch. The internship can evolve toward a more theoretically-oriented PhD thesis. For this reason, we will prefer candidates with a solid mathematical background (in particular on optimization) and in general keen on using mathematics to model real problems.

How to apply

Send by email to giovanni.neglia@inria.fr your cv and the list of university courses with the corresponding marks.

References

[1] Tian Li, Anit Kumar Sahu, Ameet Talwalkar, and Virginia Smith. Federated learning: Challenges, methods, and future directions. IEEE Signal Processing Magazine, 37 (3), 2020.

[2] Peter Carouse, H Brendan McMahan, Brendan Avent, Aurelien Bellet, Mehdi Bennis, Arjun Nitin Bhagoji, Keith Bonawitz, Zachary Charles, Graham Cormode, Rachel Cummings, et al. Advances and open problems in federated learning. arXiv preprint arXiv:1912.04977, 2019.

[3] Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, and Blaise Aguera y Arcas. Communication efficient learning of deep networks from decentralized data. In Artificial Intelligence and Statistics, PMLR, 2017.

[4] Rieke, N., Hancox, J., Li, W. et al. The future of digital health with federated learning. npj Digit. Med. 3, 119, 2020.

[5] Othmane Marfoq, Giovanni Neglia, Aurelien Bellet, Laetitia Kameni, and Richard Vidal. Federated multi-task learning under a mixture of distributions, NeurIPS 2021.

[6] Othmane Marfoq, Giovanni Neglia, Laetitia Kameni, Richard Vidal, Personalized Federated Learning through Local Memorization, ICML 2022.

[7] Valentina Zantedeschi, Aurelien Bellet, and Marc Tommasi. Fully decentralized joint learning of personalized models and collaboration graphs. volume 108 of Procexedings of Machine Learning Research, pages 864-874, 2020. PMLR.

[8] Tian Li, Shengyuan Hu, Ahmad Beirami, and Virginia Smith. Ditto: Fair and robust federated learning through personalization. In International Conference on Machine Learning, pages 6357-6368. PMLR, 2021.

[9] Enmao Diao, Jie Ding, and Vahid Tarokh. HeteroFL: Computation and Communication Efficient Federated Learning for Heterogeneous Clients. 2021.

[10] Samuel Horvath, Stefanos Laskaridis, Mario Almeida, Ilias Leontiadis, Stylianos Venieris, and Nicholas Lane. FjORD: Fair and Accurate Federated Learning under heterogeneous targets with Ordered Dropout. In Advances in Neural Information Processing Systems, volume 34, pages 12876–12889, 2021

[11] G. Cheng, Z. Charles, Z. Garrett and K. Rush, "Does Federated Dropout actually work?," 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), New Orleans, LA, USA, 2022