Offer #2023-06130

PhD Position F/M Topology Design for Decentralized Federated Learning

Contract type : Fixed-term contract

Level of qualifications required : Graduate degree or equivalent

Fonction : PhD Position

About the research centre or Inria department

The Inria centre at Université Côte d'Azur includes 37 research teams and 8 support services. The centre's staff (about 500 people) is made up of scientists of di?erent nationalities, engineers, technicians and administrative staff. The teams are mainly located on the university campuses of Sophia Antipolis and Nice as well as Montpellier, in close collaboration with research and higher education laboratories and establishments (Université Côte d'Azur, CNRS, INRAE, INSERM ...), but also with the regiona economic players.

With a presence in the fields of computational neuroscience and biology, data science and modeling, software engineering and certification, as well as collaborative robotics, the Inria Centre at Université Côte d'Azur is a major player in terms of scientific excellence through its results and collaborations at both European and international levels.

Context

This PhD thesis is in the framework of Inria research initiative on Federated Learning, FedMalin https://project.inria.fr/fedmalin/.

The PhD candidate will join NEO project-team https://team.inria.fr/neo/.
NEO is positioned at the intersection of Operations Research and Network Science.
By using the tools of Stochastic Operations Research, the team members model

situations arising in several application domains, involving networking in one way or the other.

The research activity will be supervised by

- Giovanni Neglia, http://www-sop.inria.fr/members/Giovanni.Neglia/index.htm
- Aurélien Bellet, http://researchers.lille.inria.fr/abellet/

Assignment

Topology Design for Decentralized Federated Learning

Context
The increasing size of data generated by smartphones and IoT devices motivated the development of Federated Learning (FL) [li20,kairouz21], a framework for on-device collaborative training of machine learning models. FL algorithms like FedAvg [mcmahan17] allow clients to train a common global model without sharing their personal data. FL reduces data collection costs and can help to mitigate data privacy issues, making it possible to train models on large datasets that would otherwise be inaccessible. FL is currently used by many big tech 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 [rieke20].

In the classic FL setting, a server coordinates the training phase. At each training round, the server sends the current model to the clients, which individually train on their local datasets and send model updates to the server, which in turn aggregates them (often through a simple averaging operation). In contrast to this client-server approach, decentralized FL algorithms (also called P2P FL algorithms) work by having each client communicate directly with a subset of the clients (its neighbours): this process alternates between model updates and weighted averaging of the neighbours' models (consensus-based optimization). Decentralized algorithms can take advantage of good pairwise connectivity, avoid the potential communication bottleneck at the server [marfoq20] as well as provide better privacy guarantees [cyffers22].

The communication graph (i.e., the graph induced by clients' pairwise communications) and the local clients' aggregation strategies play a fundamental role in determining FL algorithms' convergence speed. In particular, the communication topology has two contrasting effects on training time. First, a more connected topology leads to faster convergence in terms of number of communication rounds [nedic18]. Second, a more connected topology increases the duration of a communication round (e.g., because it may cause network congestion), motivating the use of degree-bounded topologies where every client sends and receives a small number of messages at each round [lian17]. Most of the existing literature has focused on one aspect or the other.

The classic literature on consensus-based optimization has quantified the effect of the communication topology on the number of rounds through worst-case convergence bounds in terms of the spectral gap of the consensus matrix (i.e., the matrix with the averaging weight), see [nedic18] and references there. Later papers have highlighted the convergence rate' insensitivity to the spectral gap for a large number of communication rounds and small learning rates [lian17,koloskova21,pu20].
Another line of work has shown that the effect of the topology is less important if local data distributions [neglia20] or average data distributions in each neighborhood [lebars23,dandi22] are close to the average data distribution over the whole population. In the extreme case of homogeneous local distributions, one may even prefer consensus matrices with poor spectral properties because they enable the use of larger learning rates [vogel22].
A separate line of works has studied how to design the communication topology in order to minimize the duration of one round, taking into account the variability of the computation times [neglia19] or the characteristics of Internet connections [marfoq20].

Research objectives

The goal of this PhD is to propose algorithms to design the communication topology for decentralized federated learning with the goal of minimizing the total training duration, taking into account how connectivity affect both the number of rounds required and the duration of a single round.
Several settings will be considered: in particular, one may construct the topology in a pre-processing step (prior to learning), or dynamically while learning. Dynamic topology design can be a way to tackle online decentralized learning [asadi22,marfoq23], where the topology is adjusted and refined as clients collect more data.
The candidate will also investigate how to practically quantify the similarity of local data distributions during training in order to exploit the advantage of having a neighborhood representative of the average population distribution [lebars23,dandi22].
Finally, he/she will also study to what extent the existing results can be extended to asymmetric communication links and other distributed optimization algorithms like push-sum ones [kempe03,benezit10].

References

[asadi22] M. Asadi, A. Bellet, O.-A. Maillard and M. Tommasi. Collaborative Algorithms for Online Personalized Mean Estimation. Transactions on Machine Learning Research, 2022.

[benezit10] F. Benezit, V. Blondel, P. Thiran, J. Tsitsiklis, and M. Vetterli, Weighted gossip: distributed averaging using non-doubly stochastic matrices, in Proceedings of the 2010 IEEE International Symposium on Information Theory, Jun. 2010.

[cyffers22] E. Cyffers and A. Bellet, Privacy Amplification by Decentralization, in Proceedings of The 25th International Conference on Artificial Intelligence and Statistics, PMLR, May 2022, pp. 5334–5353.

[dandi22] Y. Dandi, A. Koloskova, M. Jaggi, and S. U. Stich, “Data-heterogeneity-aware Mixing for Decentralized Learning.” arXiv, Apr. 13, 2022.

[kairouz21] P. Kairouz, et al. Advances and open problems in federated learning. Foundations and Trends in Machine Learning, 14(1-2), pp. 1-210, 2021.

[kempe03] D. Kempe, A. Dobra, and J. Gehrke, Gossip-based computation of aggregate information, in Proceedings of the 44th Annual IEEE Symposium on Foundations of Computer Science, 2003, pp. 482–491.

[koloskova21] A. Koloskova, N. Loizou, S. Boreiri, M. Jaggi, and S. U. Stich, A Unified Theory of Decentralized SGD with Changing Topology and Local Updates. arXiv, Mar. 02, 2021. doi: 10.48550/arXiv.2003.10422.

[lebars23] B. Le Bars, A. Bellet, M. Tommasi, E. Lavoie, and A.-M. Kermarrec, Refined Convergence and Topology Learning for Decentralized SGD with Heterogeneous Data. AISTATS 2023

[li20] T. Li, A. Kumar Sahu, A. Talwalkar, and V. Smith. Federated learning: Challenges, methods, and future directions. IEEE Signal Processing Magazine, 37 (3), 2020.

[lian17] X. Lian, C. Zhang, H. Zhang, C.-J. Hsieh, W. Zhang, and J. Liu, Can decentralized algorithms outperform centralized algorithms? a case study for decentralized parallel stochastic gradient descent, in Proceedings of the 31st International Conference on Neural Information Processing Systems, in NIPS’17. Red Hook, NY, USA: Curran Associates Inc., Dec. 2017, pp. 5336–5346.

[lian18] X. Lian, W. Zhang, C. Zhang, and J. Liu, Asynchronous Decentralized Parallel Stochastic Gradient Descent, in Proceedings of the 35th International Conference on Machine Learning, PMLR, Jul. 2018, pp. 3043–3052.

[marfoq20] O. Marfoq, C. Xu, G. Neglia, and R. Vidal, Throughput-Optimal Topology Design for Cross-Silo Federated Learning, in 34th Conference on Neural Information Processing Systems (NeurIPS 2020), Vancouver, Canada, Dec. 2020.

[marfoq23] O. Marfoq, G. Neglia, L. Kameni, R. Vidal. Federated Learning for Data Streams. AISTATS 2023.

[mcmahan17] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. Aguera y Arcas. Communication efficient learning of deep networks from decentralized data. In Artificial Intelligence and Statistics, PMLR, 2017.

[neglia19] G. Neglia, G. Calbi, D. Towsley, and G. Vardoyan. 2019. The Role of Network Topology for Distributed Machine Learning. In IEEE INFOCOM 2019 - IEEE Conference on Computer Communications. IEEE Press, 2350–2358.

[nedic18] A. Nedi?, A. Olshevsky, and M.G. Rabbat, Network Topology and Communication-Computation Tradeoffs in Decentralized Optimization. In: Proceedings of the IEEE 106.5 (May 2018), pp. 953–976.

[neglia20] G. Neglia, C. Xu, D. Towsley, and G. Calbi, Decentralized gradient methods: does topology matter?, in 23rd International Conference on Artificial Intelligence and Statistics (AISTATS), Palermo, Italy, Jun. 2020.

[pu20] S. Pu, A. Olshevsky, and I. Ch. Paschalidis, Asymptotic Network Independence in Distributed Stochastic Optimization for Machine Learning: Examining Distributed and Centralized Stochastic Gradient Descent. In: IEEE Signal Process. Mag. 37.3 (2020), pp. 114–122.

[rieke20] Rieke, N., Hancox, J., Li, W. et al. The future of digital health with federated learning. npj Digit. Med. 3, 119, 2020.

[tang18] H. Tang, X. Lian, M. Yan, C. Zhang, and J. Liu, “D^2: Decentralized Training over Decentralized Data,” in Proceedings of the 35th International Conference on Machine Learning, PMLR, Jul. 2018, pp. 4848–4856.

[vogels22] T. Vogels, H. Hendrikx, and M. Jaggi, Beyond spectral gap: the role of the topology in decentralized learning, in Advances in neural information processing systems, A. H. Oh, A. Agarwal, D. Belgrave, and K. Cho, Eds., 2022.

Main activities

Research.

Skills

The candidate should have a solid mathematical background (in particular on optimization) and in general be keen on using mathematics to model real problems and get insights. He should also be knowledgeable on machine learning and have good programming skills. Previous experiences with PyTorch or TensorFlow is a plus.

We expect the candidate to be fluent in 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 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
- Contribution to mutual insurance (subject to conditions)

Remuneration

Duration: 36 months
Location: Sophia Antipolis, France
Gross Salary per month: 2051€ brut per month (year 1 & 2) and 2158€ brut per month (year 3)

General Information

- **Theme/Domain :** Optimization, machine learning and statistical methods
System & Networks (BAP E)
- **Town/city :** Sophia Antipolis
- **Inria Center :** Centre Inria d'Université Côte d'Azur
- **Starting date :** 2023-10-01
- **Duration of contract :** 3 years
- **Deadline to apply :** 2025-04-30

Contacts

- **Inria Team :** NEO

- **PhD Supervisor :**
  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.

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.

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.