Fairness in Decentralized Learning

Machine Learning is becoming ubiquitous in our everyday lives. It is now used in digital assistants, for medical diagnosis, for autonomous vehicles, ... Its success can be explained by the good performances of learned models, sometimes reaching human-level capabilities. However, simply being accurate is not sufficient if these models are to be largely deployed. Hence, the notion of trustworthiness has to be considered as soon as human are involved in the loop. Among all the existing trustworthiness notions, fairness is especially important. Hence, a model used for medical diagnosis should not be biased against sub-groups of the population. Similarly, a model designed to predict whether someone should receive a loan should give the same opportunity of receiving one to every creditworthy person regardless of their gender, ethnicity, or any other sensitive attribute.

Fairness has been extensively studied in the centralized case, that is when all the data is available in a single place. However, it has received far less attention in the decentralized setting (Kairouz et al., 2019; Li et al., 2020), that is when the data is owned by multiple entities that would like to collaborate to learn efficient models but do not wish to share their data. In this context, fairness can be defined at two different levels. On the one hand, at the data owners level where the objective is either to obtain a global model whose performances are similar among the different involved entities (Mohri et al., 2019; Li et al., 2019) or personalized models for each entity that reflect their involvement (Lyu et al., 2020; Zhang et al., 2020b). On the other hand, at the data level where the goal is to use the limited information provided by each entity to learn a central model that does not unjustly discriminate against sub-groups of the population (Du et al., 2020; Zhang et al., 2020a; Abay et al., 2020).

Objectives: The goal of this 6 months internship is to study the latter setting, that is fairness at the data level in the context of Decentralized Learning. The main objectives are (i) to review some of the existing literature in the field, (ii) to design new algorithms to learn fair models in a decentralized context and (iii) to derive theoretical guarantees on the fairness and utility levels of the obtained models.

Requirements: Successful candidates should have a solid background in Machine Learning with an interest in Fairness and Federated Learning. Some knowledge in Statistical Learning Theory would be a plus but is not mandatory. A good understanding of the Python programming language is expected. Finally, proficiency in English is required as it will be one of the main working language.

Keywords: Machine Learning, Fairness, Decentralized Learning, Learning Theory

Contact: The recruited student will be based in the INRIA Lille - Nord Europe research center and will be supervised by Michaël Perrot. The interested students should send an e-mail (in english) with a CV and a Motivation Letter to *michael.perrot@inria.fr*.

References

Abay, A., Zhou, Y., Baracaldo, N., Rajamoni, S., Chuba, E., and Ludwig, H. (2020). Mitigating bias in federated learning. *arXiv preprint arXiv:2012.02447*.

Du, W., Xu, D., Wu, X., and Tong, H. (2020). Fairness-aware agnostic federated learning. *arXiv preprint arXiv:2010.05057*.

Kairouz, P., McMahan, H. B., Avent, B., Bellet, A., Bennis, M., Bhagoji, A. N., Bonawitz, K., Charles, Z., Cormode, G., Cummings, R., et al. (2019). Advances and open problems in federated learning. *arXiv preprint arXiv:1912.04977*.

Li, T., Sahu, A. K., Talwalkar, A., and Smith, V. (2020). Federated learning: Challenges, methods, and future directions. *IEEE Signal Processing Magazine*, 37(3):50–60.

Li, T., Sanjabi, M., Beirami, A., and Smith, V. (2019). Fair resource allocation in federated learning. *arXiv preprint arXiv:1905.10497*.

Lyu, L., Xu, X., Wang, Q., and Yu, H. (2020). Collaborative fairness in federated learning. In *Federated Learning*, pages 189–204. Springer.

Mohri, M., Sivek, G., and Suresh, A. T. (2019). Agnostic federated learning. In *International Conference on Machine Learning*, pages 4615–4625. PMLR.

Zhang, D. Y., Kou, Z., and Wang, D. (2020a). Fairfl: A fair federated learning approach to reducing demographic bias in privacy-sensitive classification models. In *2020 IEEE International Conference on Big Data (Big Data)*, pages 1051–1060. IEEE.

Zhang, J., Li, C., Robles-Kelly, A., and Kankanhalli, M. (2020b). Hierarchically fair federated learning. *arXiv preprint arXiv:2004.10386*.