

Intitulé de l'offre

Stage M2: Apprentissage fédéré hiérarchique optimisé pour un environnement hétérogène

Description de l'unité

Situé à Saclay, en Ile-de-France sud, le CEA LIST (<http://www-list.cea.fr/>) est un centre de recherche scientifique et technologique dédié au développement de logiciels, de systèmes embarqués et de capteurs pour des applications destinées à la défense, la sécurité, l'énergie, le nucléaire, l'environnement et la santé. Le CEA LIST fait partie de l'écosystème dynamique et stimulant de l'Université Paris Saclay. Il compte plus de 700 chercheurs se focalisant sur les systèmes numériques intelligents, centrés autour de l'intelligence artificielle, l'usine du futur, l'instrumentation innovante, les systèmes cyberphysiques et la santé numérique. Au sein de cet institut, le Laboratoire Instrumentation Intelligente, Distribuée et Embarquée (LIIDE) développe des fonctionnalités innovantes d'intelligence artificielle pour la mesure répartie et l'apprentissage frugal et distribué.

Sujet de stage (1000 car.)

En 2016, Google publie les principes fondateurs de l'apprentissage fédéré avec la promesse de faire de l'apprentissage statistique sans compromettre les données des clients [1]. Il s'agit d'une approche collaborative où plusieurs clients participent à l'apprentissage du modèle sans transmettre leurs données personnelles, mais uniquement les paramètres du modèle mis à jour localement puis agrégés sur le serveur central. Dans le cas des applications cross-device (IoT) où le nombre de participants est très important, le déploiement d'une solution fédérée basée sur un seul serveur apparaît en pratique peu efficace, notamment dans la gestion de l'hétérogénéité des clients et des communications clients-serveur. Dans le cadre du stage, nous allons étudier l'apprentissage fédéré hiérarchique qui apparaît comme une alternative prometteuse à l'apprentissage fédéré centralisé afin de permettre son déploiement dans le continuum Cloud-Edge-IoT.

Description de l'offre (3000 car.)

Le stage concerne l'étude de l'apprentissage fédéré hiérarchique [2] qui se base sur une couche de serveurs supplémentaire au niveau edge pour réaliser une agrégation au plus proche des IoT. Cette couche est donc située entre le serveur central du Cloud et les unités de calculs clients (voir Fig.1). Habituellement, les ressources de calculs sont allouées/connectées au serveur edge qui est le plus proche géographiquement. Nous pensons qu'une stratégie d'allocation moins naïve qui prend en compte l'hétérogénéité des clients IoT (hétérogénéité de la distribution de leurs données mais aussi de leur capacité de calculs et de leur moyen de communication) peut permettre de diminuer significativement les coûts de communication et le temps de latence pour construire des modèles à précision équivalente [3].

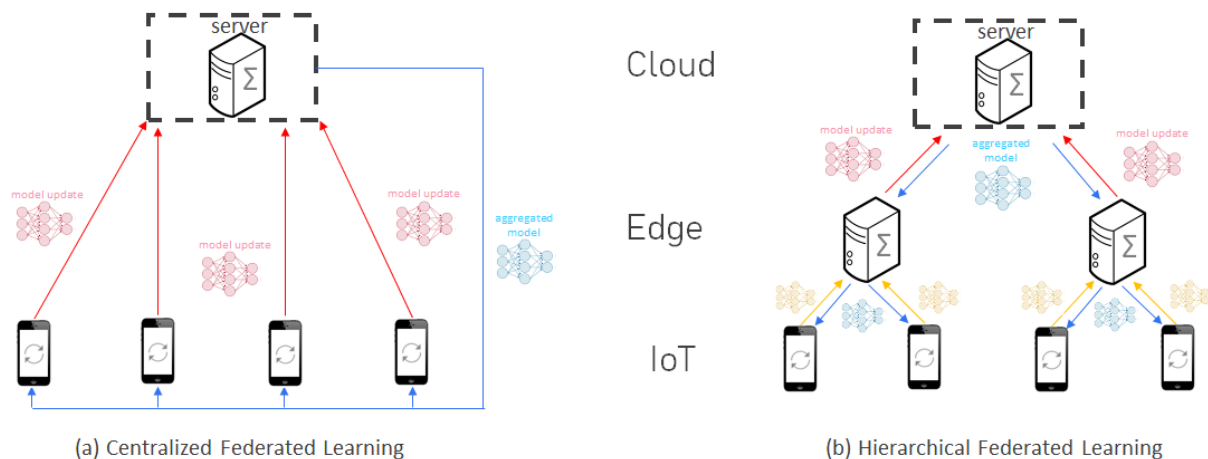


Fig-1 : apprentissage fédéré centralisé (a) vs. apprentissage fédéré hiérarchique (b)

Le stage se déroulera en quatre étapes. Dans un premier, le stagiaire se familiarisera avec l'apprentissage fédéré hiérarchique au travers d'une petite étude bibliographique. Les étapes 2 et 3 se dérouleront dans un environnement de simulation. Le stagiaire étudiera quelles sont les meilleures stratégies d'allocation de ressources ; dans un premier temps lorsque les clients ont des données hétérogènes (mais avec des capacités de calculs et des moyens de communication identiques) puis dans un deuxième temps lorsqu'ils ont des unités de calculs hétérogènes. Enfin, si le temps le permet, les algorithmes développés au cours des étapes 2 et 3 pourront être testés dans un environnement cloud-edge-IoT.

Le stage s'adresse à un(e) étudiant(e) du cycle ingénieur/universitaire cherchant un stage M2 et manifestant l'envie de travailler dans le milieu de la recherche. Idéalement, le/la candidat(e) suit actuellement une formation en lien avec le domaine de l'Intelligence Artificielle/Machine Learning. La connaissance des principaux algorithmes d'optimisation ainsi que la maîtrise de Python sont indispensables. Durant le stage, le/la candidat(e) sera en relation directe avec les membres de l'équipe d'apprentissage fédéré (deux ingénieurs-chercheurs, un thésard et deux post-doctorants). Nous attendons qu'il/elle soit capable d'apporter sa rigueur, son enthousiasme et sa curiosité pour la recherche. Nous offrons la possibilité à nos stagiaires de participer à la conférence JDSE (Junior conference on Data Science and Engineering) afin de les familiariser avec le déroulement d'une conférence scientifique.

[1] Google AI blog: <https://ai.googleblog.com/2017/04/federated-learning-collaborative.html>

[2] L. Liu, J. Zhang, S. H. Song and K. B. Letaief, "Client-Edge-Cloud Hierarchical Federated Learning," *In IEEE International Conference on Communications (ICC)*, pp. 1-6, Dublin, Ireland, 2020.

[3] A.A. Abdellatif, N. Mhaisen, A. Mohamed, A. Erbad, M. Guizani, Z. Dawy and, W. Nasreddine. "Communication-efficient hierarchical federated learning for IoT heterogeneous systems with imbalanced data". *Future Generation Computer Systems*, 128, pp. 406-419, 2022.