

Titre

Stage M2 : apprentissage fédéré incrémental



Mots clés

Machine Learning ; Apprentissage fédéré ; Apprentissage incrémental ; Oubli catastrophique.

Présentation du service : CEA LIST / SID

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 - le plus grand pôle scientifique français comptant 60 000 étudiants. 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 SID (Service d'Intelligence des Données) est composé de 50 ingénieurs-chercheurs qui travaillent au développement et au transfert industriel de technologies de pointe en IA.

Description du projet

En 2016, Google publie les principes fondateurs de l'apprentissage fédéré [1] avec la promesse de créer des IA sans compromettre les données des utilisateurs. Cette méthode est en train de changer le paradigme actuel de l'IA centralisée, où construire de meilleurs modèles se résume souvent à collecter toujours plus de données personnelles et les centraliser sur un serveur. L'apprentissage fédéré est une approche collaborative où tous les utilisateurs d'un service participent à l'apprentissage du modèle sans transmettre leurs données personnelles mais uniquement les paramètres du modèle mis à jour localement (voir Fig.1). Au lieu de centraliser les données, seuls les paramètres du modèle sont agrégés sur le serveur central ce qui permet d'améliorer la confidentialité des données et de limiter les coûts de communication.

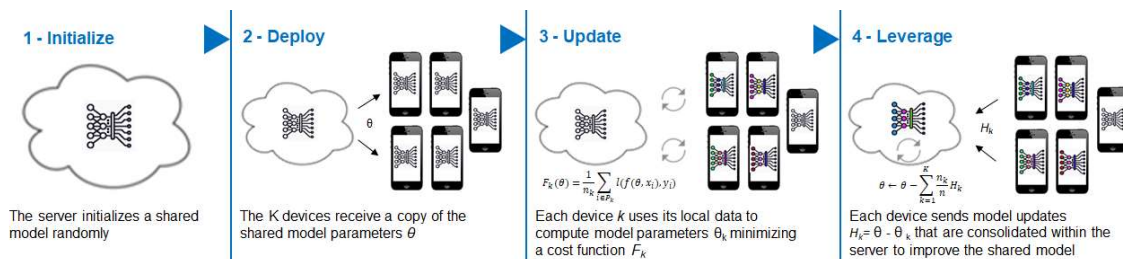


Fig-1 : illustration de l'apprentissage fédéré

En pratique, les solutions d'apprentissage fédéré ont surtout été mises en œuvre pour des applications cross-silo de collaboration entre entreprises avec des bases de données locales figées dans le temps. Or cette hypothèse devient peu vraisemblable pour des applications cross-devices IoT où les acteurs reçoivent des données en flux avec peu de capacité de stockage (voir Fig.2).

L'objectif du stage est de mettre en œuvre des solutions d'apprentissage incrémental dans un scénario d'apprentissage fédéré, i.e. apprendre un modèle de façon collaborative à partir de données en flux décentralisées. Le principal challenge de l'étude sera de proposer des algorithmes robustes à l'oubli catastrophique qui a tendance à limiter les capacités d'un réseau de neurones à apprendre des tâches successives. Le déroulement du stage sera scindé en deux parties. Dans un premier temps, une étude bibliographique sera menée

sur les quelques algorithmes proposés dans la littérature autour de l'apprentissage fédéré incrémental [2,3,4,5]. Dans un deuxième temps, des solutions prometteuses issues de la littérature ou novatrices seront mises en œuvre dans un environnement de simulation.

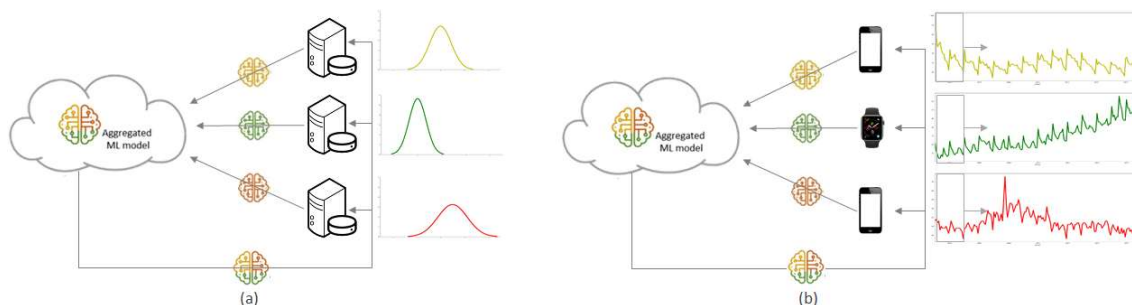


Fig-2 : illustration (a) application cross-silo (b) application cross-devices

- [1] Google AI blog: <https://ai.googleblog.com/2017/04/federated-learning-collaborative.html>
- [2] J. Yoon *et al.* Federated Continual Learning with Weighted Inter-Client Transfer. arXiv, 2020.
- [3] J. Dong *et al.* Federated Class-Incremental Learning. arXiv, 2022.
- [4] Y. Ma *et al.* Continual Federated Learning Based on Knowledge Distillation. IJCAI, 2022.
- [5] T. Park *et al.* Tackling Dynamics in Federated Incremental Learning with Variational Embedding Rehearsal. arXiv, 2021.

Profil recherché

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 algorithmes d'optimisation en Machine Learning 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 (2 ingénieurs-chercheurs, 2 thésards et 2 post-doctorants). Nous attendons qu'il/elle soit capable d'apporter ses idées innovantes, son enthousiasme, sa rigueur et qu'il/elle fasse preuve d'un esprit d'équipe prononcé.

La durée du stage est de 6 mois minimum et pourra se poursuivre par une thèse de doctorat. Le stage est rémunéré.

Informations administratives et contacts

Nature du contrat de travail : stage

Type du contrat de travail : Droit privé

Délai administratif pour début de contrat : environ 3 mois

Date de prise de fonction envisagée : entre mars et mai 2023

Envoyer CV et lettre de motivation à :

Aurélien Mayoue (aurelien.mayoue@cea.fr)