Titre

[FR] Apprentissage fédéré: collaboration et personnalisation

[EN] Federated learning: collaboration vs. personalization

Présentation du laboratoire : CEA LIST / LS2D

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 LS2D (Laboratoire Sciences des Données et de la Décision) travaille sur les algorithmes et méthodologies de l'intelligence artificielle et du traitement du signal. Les recherches et avancées technologiques du laboratoire sont guidées par des applications variées, pour lesquelles les spécificités et contraintes sur les données ou l'environnement d'exécution nécessitent une conception fine des IA et de leur intégration comme briques unitaires de systèmes complexes.

Résumé

[FR] 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. Au lieu de centraliser les données, seuls les paramètres du modèle sont agrégés sur le serveur central ce qui permet de préserver la confidentialité des données et de limiter les coûts de communication.

L'apprentissage d'un réseau de neurones profond nécessite un ensemble de données indépendantes et identiquement distribuées (iid) afin de garantir que le gradient stochastique est une estimation non biaisée du gradient entier. Contrairement à un apprentissage centralisé, il est impossible dans le cadre de l'apprentissage fédéré d'assurer que les données locales de chaque utilisateur soient toujours iid. Le principal objectif de la thèse consistera à explorer des pistes de recherche afin d'adapter l'apprentissage des modèles locaux en fonction de la distance entre la distribution des données locales et celles de l'ensemble de la population. Cette personnalisation des modèles devrait homogénéiser les performances « at edge » indépendamment de la distribution des données locales.

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

[EN] In 2016, Google introduced the founding principles of federated learning [1] which opened up a brand new computing paradigm for AI. Until now, most of deep machine learning approaches adopt a centralized way to train a model. It requires the data to be stored in a datacenter. This is practically what giant AI companies have been doing over the years. However, this centralized approach is privacy-intrusive as users of the service have to send their data to the service provider which manages the datacenter. Federated learning is a collaborative process which leaves the training data distributed

on the client devices and learns a shared model by aggregating locally-computed updates. As the data remains in its original location, the privacy is improved and the cost communication also decreases.

The independent and identically distributed (IID) sampling of the training data is a key point to train a machine learning model. It ensures that the stochastic gradient is an unbiased estimate of the full gradient. But, in a decentralized learning process, it is unrealistic to assume that the local data on each edge device is always IID. “Non-iidness” can be the cause of a significant decrease of the model accuracy. To improve the performance of federated learning whatever the local distribution of data, we investigate a way of personalizing the models at edge which permits each node to fine-tune its model locally while continuing to train the shared model.

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

Description du sujet de thèse

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.

Traditionnellement, l’apprentissage automatique en IA nécessite la collecte de données clients. Ces données sont stockées sur un serveur central où se déroule l’entrainement du modèle. Cette architecture centralisée, utilisée par les fournisseurs de services en IA, pose néanmoins deux problèmes majeurs. D’une part, elle compromet la vie privée des clients qui doivent céder leurs données et, d’autre part, elle crée des coûts de transfert exorbitants dans un contexte d’explosion du volume des données à traiter. Une solution pour les éditeurs de services consiste à abandonner la collecte centralisée de données. Pour cela, ils s’orientent vers un nouveau paradigme pour l’apprentissage des modèles : l’apprentissage fédéré.

Il s’agit d’un apprentissage collaboratif entre tous les utilisateurs d’un service qui s’appuie sur une architecture décentralisée. Le modèle est tout d’abord initialisé (aléatoirement ou en utilisant des données publiques) sur un serveur. Il est ensuite déployé sur chaque terminal client où il est amélioré localement à partir des données utilisateur. Les mises à jour du modèle sont alors chiffrées et agrégées au niveau du serveur pour obtenir un modèle optimisé qui sera à son tour déployé et amélioré sur les terminaux clients. Ainsi, chaque utilisateur participe à l’apprentissage du modèle sans transmettre au fournisseur de service ses données personnelles mais uniquement les paramètres du modèle mis à jour. Cette approche présente l’avantage de préserver la confidentialité des données et de limiter les coûts de communication.

L’apprentissage fédéré est particulièrement adapté aux scénarios de mobilité ou d’internet des objets, où les données sont sensibles et les utilisateurs variés (e.g. Gboard [2]). Il devrait également inciter la collaboration entre sociétés qui jusque-là étaient réticentes à transmettre leurs données à des tiers pour des raisons de confidentialité (e.g. la fondation Substra [3] dans le domaine de la santé). Enfin, l’engouement autour de l’apprentissage fédéré est illustré par le fait que plusieurs start-ups créées ces dernières années proposent des outils et des solutions autour de cette technologie (Snips, Owkin, S20.ai…).

Notre sujet de thèse aborde deux verrous propres à l’apprentissage fédéré : (1) Comment garantir l’homogénéité des performances pour les utilisateurs d’un même service dans le cas où les données utilisateurs ne sont pas iid ? (2) Comment se prémunir des attaques venues de l’intérieur, i.e. lancées par un ou plusieurs utilisateurs du service ?

1) L'apprentissage d'un réseau de neurones profond nécessite un ensemble de données iid afin de garantir que le gradient stochastique est une estimation non biaisée du gradient entier. Contrairement à un apprentissage centralisé, il est impossible dans le cadre de l'apprentissage fédéré d'assurer que les données locales de chaque utilisateur soient toujours iid. Ceci se traduit par une diminution des performances globales du modèle (par rapport à une architecture centralisée où on aurait pu construire des lots de données iid) et une grande hétérogénéité des performances sur les données locales des utilisateurs. L'objectif de cette tâche est de trouver un compromis entre collaboration (où tous les utilisateurs participent à l'apprentissage du modèle partagé) et personnalisation (où chaque utilisateur participe uniquement à l'amélioration de son modèle). En pratique, durant les premières itérations d'apprentissage, tous les utilisateurs collaborent à l'apprentissage du modèle partagé mais à partir de l'itération S, chacun aura désormais la possibilité de personnaliser son modèle. Le choix de S n'est pas du tout trivial et pourrait se baser sur la distance entre la distribution des données locales et celle de l'ensemble de la population, ou sur l'écart entre performances locales et performances globales. Cette personnalisation des modèles devrait homogénéiser les performances « at edge » indépendamment de la distribution des données locales.

2) De nombreux papiers de recherche présentent des solutions pour sécuriser les transferts de données entre le serveur central et les terminaux clients (differential privacy, secure aggregation, homomorphic encryption). L'objectif est de garantir que les données clients ne pourraient pas être reconstruites même si les paramètres des modèles étaient interceptés. Dans notre cas, nous aborderons les attaques internes, i.e. réalisées par un ou plusieurs utilisateurs du service. Ces attaques peuvent avoir plusieurs objectifs : détériorer les performances du modèle partagé et/ou récupérer les données personnelles d'autres clients. Dans un premier temps, nous nous intéresserons à la détérioration des performances du modèle partagé, qui est intiment liée à la thématique de personnalisation des modèles au travers notamment de l'étude statistique de la distribution des données.

Les cas d'application envisagés concernent des données de consommations électriques, des données de véhicules autonomes ou des données d'assistants vocaux.

[1] H.B. McMahan, E. Moore, D. Ramage and B. Aguera y Arcas, "Federated Learning of Deep Networks using Model Averaging", Feb. 2016

[2] https://ai.googleblog.com/2017/04/federated-learning-collaborative.html

[3] https://www.substra.ai

Profil recherché

La thèse s'adresse à un(e) étudiant(e) du cycle ingénieur/universitaire disposant d'un Master 2 dans l'un des domaines suivants : IA, Machine Learning, statistiques.

Contacts

Aurélien MAYOUE [encadrant] (aurelien.mayoue@cea.fr)

Cédric GOUY-PAILLER [directeur de thèse] (cedric.gouy-pailler@cea.fr)