

Optimisation de l'Apprentissage Fédéré en Présence de Données Hétérogènes et Déséquilibrées

Sonia Guehis, Ines Alaya, Sana Ben Hamida

February 10, 2025

1 Sujet:

L'apprentissage fédéré (FL) [1], [2] permet d'entraîner des modèles de machine learning de manière décentralisée, sans partager directement les données brutes des participants. Ce paradigme est particulièrement intéressant pour les domaines où la confidentialité est primordiale (santé, finance, IoT, biodiversité marine, etc). Cependant, les données utilisées dans FL posent plusieurs défis majeurs (Non-IID Data:Non-Independent and Identically Distributed), notamment l'hétérogénéité des données, le déséquilibre entre les volumes et les classe de données, le biais, la disponibilité et la variabilité temporelle des données. Ce projet vise à optimiser la convergence, la robustesse et la généralisation des modèles fédérés face aux défis des données en agissant à deux niveaux. D'une part, il s'agira d'agir directement sur les données pour harmoniser les contributions locales en appliquant des méthodes de ré-échantillonnage intelligent, d'augmentation de données ou de pondération des échantillons. Des méthodes d'optimisation peuvent être utilisées pour minimiser les divergences entre les distributions locales et la distribution globale. D'autre part, le projet explorera l'optimisation des stratégies d'agrégation et de pondération des modèles en tenant compte de la qualité et de la diversité des données locales, en intégrant des métaheuristiques telles que les algorithmes génétiques ou l'optimisation par essaim particulaire pour ajuster dynamiquement les contributions.

2 Motivation du Sujet

Le sujet proposé par les trois collègues Ines, Sana et Sonia appartenant respectivement au pôle 1 et pôle 3 traite à la fois des problématiques d'hétérogénéité de format et de volumétrie des données dans le contexte d'apprentissage fédéré, tout en explorant les méthodes d'optimisation et intégrant les métaheuristiques. Ce sujet a donc pour avantage de fédérer les travaux interpoles au sein du Lam-sade.

3 Mots clés:

Apprentissage Fédérés, Machine Learning, Approche data-centric, Méthodes d'Optimisation, métaheuristiques.

4 Direction de thèse:

- Sonia Guehis: MCF à Paris Nanterre et chercheuse associée au Lamsade, Pôle3;
- Ines Alaya: MCF à Paris Nanterre et chercheuse associée au Lamsade, Pôle1;
- Sana Ben Hamida: MCF à Paris Nanterre, HDR et chercheuse associée au Lamsade, pôle3.

References

- [1] Betul Yurdem et al. “Federated learning: Overview, strategies, applications, tools and future directions”. In: *Heliyon* 10.19 (2024), e38137. ISSN: 2405-8440. DOI: <https://doi.org/10.1016/j.heliyon.2024.e38137>. URL: <https://www.sciencedirect.com/science/article/pii/S2405844024141680>.
- [2] Qiang Yang et al. “Federated Machine Learning: Concept and Applications”. In: *CoRR* abs/1902.04885 (2019). arXiv: 1902.04885. URL: <http://arxiv.org/abs/1902.04885>.