Proposition de stage de recherche - Année 2024/2025

Amélioration de la performance des modèles d'apprentissage fédéré à l'Aide de métaheuristiques

Lieu : LAMADE - Université Paris Dauphine – PSL
Place du Maréchal de Lattre de Tassigny - 75775 PARIS Cedex 16
Proposé par : Sonia Guehis (sonia.guehis@lamsade.dauphine.fr), Ines Alya (ialaya@parisnanterre.fr)
Sana Ben Hamida (sana.mrabet@dauphine.psl.eu)

Contexte :

L'apprentissage fédéré (Federated Learning, FL) est une approche de l'apprentissage machine décentralisé où les modèles sont entraînés localement sur plusieurs appareils sans partager les données brutes, préservant ainsi la confidentialité et réduisant les coûts de communication. Ce paradigme est de plus en plus utilisé dans des domaines sensibles comme la santé, la finance ou encore les appareils connectés. Cependant, il pose des défis uniques, notamment :

- L'hétérogénéité des données entre les participants,
- L'optimisation efficace des modèles sur des ensembles de données distribués et déséquilibrés,
- La communication coûteuse entre les participants et le serveur central.

Objectif :

Ce sujet vise à développer des approches innovantes pour améliorer la performance des modèles d'apprentissage fédéré. Il s'agira d'explorer des méthodes hybrides mêlant l'intégration de méta-heuristiques (ex. : colonies de fourmis, algorithmes génétiques) pour :

- Identifier dynamiquement les sous-ensembles d'appareils participant à chaque cycle, maximisant ainsi la performance globale du modèle tout en respectant les contraintes du système.
- Optimiser la sélection dynamique des modèles locaux contribuant efficacement à la performance globale.
- Optimiser les stratégies d'agrégation des poids pour gérer le déséquilibre des données et améliorer la convergence.

Contributions attendues :

1. Conception d'une stratégie hybride combinant gestion de la diversité des données et techniques d'optimisation pour améliorer la performance globale.
2. Validation expérimentale sur des benchmarks standards d'apprentissage fédéré (par exemple, FEMNIST, CIFAR-10, etc.) en comparant avec des approches de l'état de l'art.

Compétences requises :

- Connaissances en machine learning (TensorFlow, PyTorch, ou frameworks similaires).
- Bases en optimisation ou en méta-heuristiques.
- Familiarité avec des concepts en traitement de données (gestion des biais, équilibrage des données).

Résultats attendus :

- Une analyse comparative entre les approches proposées et les solutions existantes.
- Un prototype fonctionnel validant l'approche

Références

1. **A. Rauniyar et al.**, "Federated Learning for Medical Applications: A Taxonomy, Current Trends, Challenges, and Future Research Directions," in IEEE Internet of Things Journal, vol. 11, no. 5, pp. 7374-7398, 1 March1, 2024, doi: 10.1109/JIOT.2023.3329061.
2. **Li, T., Sahu, A. K., Talwalkar, A., & Smith, V. (2020).** Federated learning: Challenges, methods, and future directions. IEEE Signal Processing Magazine, 37(3), 50-60. DOI
3. **Supriya, Y.; Gadekallu, T.R.** Particle Swarm-Based Federated Learning Approach for Early Detection of Forest Fires. Sustainability 2023, 15, 964. https://doi.org/10.3390/su15020964
4. **Essam H. Houssein, Awny Sayed**, Boosted federated learning based on improved Particle Swarm Optimization for healthcare IoT devices, Computers in Biology and Medicine,Volume 163,2023. https://doi.org/10.1016/j.compbiomed.2023.107195.
5. **Yang, Q., Liu, Y., Chen, T., & Tong, Y. (2019).** Federated machine learning: Concept and applications. ACM Transactions on Intelligent Systems and Technology (TIST), 10(2), 1-19. https://doi.org/10.1145/3298981
6. **Kulkarni, V., Kulkarni, M. and Pant, A**. Survey of personalization techniques for federated learning. IEEE, City, 2020.
7. **Zhang, C., Xie, Y., Bai, H., Yu, B., Li, W. and Gao, Y**. A survey on federated learning. Knowledge-Based Systems, 216 (2021), 106775.
8. **Souhila, Badra & Ouchani, Samir & Malki, Mimoun**. (2022). Genetic Algorithm Based Aggregation for Federated Learning in Industrial Cyber Physical Systems. 10.1007/978-3-031-18409-3_2.

English version

Improving the Performance of Federated Learning Models Using Metaheuristics

Context:

Federated Learning (FL) is a decentralized machine learning approach where models are trained locally on multiple devices without sharing raw data, thus preserving privacy and reducing communication costs. This paradigm is increasingly used in sensitive domains such as healthcare, finance, and connected devices. However, it presents specific challenges, including:

- The heterogeneity of data across participants,
- Efficient optimization of models on distributed and imbalanced datasets,
- High communication costs between participants and the central server.

Objective:

This project aims to develop innovative approaches to improve the performance of Federated Learning models. The focus will be on exploring hybrid methods integrating metaheuristics (e.g., ant colony optimization, genetic algorithms) to:

- Dynamically identify subsets of devices participating in each training round, maximizing overall model performance while adhering to system constraints,
- Optimize the dynamic selection of local models that contribute effectively to the global performance,
- Enhance aggregation strategies to handle data imbalance and improve convergence.

Expected Contributions:

1. Design of a hybrid strategy combining data diversity management and optimization techniques to enhance overall performance.
2. Experimental validation on standard Federated Learning benchmarks (e.g., FEMNIST, CIFAR-10) with comparisons to state-of-the-art approaches.

Required Skills:

- Knowledge of machine learning frameworks (TensorFlow, PyTorch, or similar).
- Foundations in optimization or metaheuristics.
- Familiarity with data processing concepts (bias handling, data balancing).

Expected Outcomes:

- A comparative analysis between the proposed approaches and existing solutions.

- A functional prototype validating the proposed approach.

References:

1. **A. Rauniyar et al.**, "Federated Learning for Medical Applications: A Taxonomy, Current Trends, Challenges, and Future Research Directions," in IEEE Internet of Things Journal, vol. 11, no. 5, pp. 7374-7398, 1 March1, 2024, doi: 10.1109/JIOT.2023.3329061.
2. **Li, T., Sahu, A. K., Talwalkar, A., & Smith, V. (2020).** Federated learning: Challenges, methods, and future directions. IEEE Signal Processing Magazine, 37(3), 50-60. DOI
3. **Supriya, Y.; Gadekallu, T.R.** Particle Swarm-Based Federated Learning Approach for Early Detection of Forest Fires. Sustainability 2023, 15, 964. https://doi.org/10.3390/su15020964
4. **Essam H. Houssein, Awny Sayed**, Boosted federated learning based on improved Particle Swarm Optimization for healthcare IoT devices, Computers in Biology and Medicine,Volume 163,2023. https://doi.org/10.1016/j.compbiomed.2023.107195.
5. **Yang, Q., Liu, Y., Chen, T., & Tong, Y. (2019).** Federated machine learning: Concept and applications. ACM Transactions on Intelligent Systems and Technology (TIST), 10(2), 1-19. https://doi.org/10.1145/3298981
6. **Kulkarni, V., Kulkarni, M. and Pant, A**. Survey of personalization techniques for federated learning. IEEE, City, 2020.
7. **Zhang, C., Xie, Y., Bai, H., Yu, B., Li, W. and Gao, Y**. A survey on federated learning. Knowledge-Based Systems, 216 (2021), 106775.
8. **Souhila, Badra & Ouchani, Samir & Malki, Mimoun**. (2022). Genetic Algorithm Based Aggregation for Federated Learning in Industrial Cyber Physical Systems. 10.1007/978-3-031-18409-3_2.