

APPRENTISSAGES FEDERES SUR LES DONNEES DE MOBILITE

POST-DOC

Contexte et problématique

Le projet VILAGIL ambitionne de créer un écosystème capable de répondre aux besoins de la mobilité pour la région Occitanie, et plus largement de contribuer au développement d'innovations dans le domaine de la ville intelligente ». Dans le cadre de ce projet, l'action « Data et Mobilité » conçoit et développe des mécanismes d'intégration automatique et d'accès fédérés aux données de mobilité. Ces données sont fournies par les partenaires non académiques du projet tels que Toulouse Metropole, Sicoval, Tisséo Collectivité ; dans certains cas ces données seront distribuées sur les différents sites des partenaires sans possibilité de les intégrer de manière centralisée. L'offre de postdoc cible donc particulièrement la problématique de l'apprentissages fédérés (« federated learning») afin de développer des algorithmes d'apprentissage automatique dirigés par ces données.

Objectifs

Vu la forte hétérogénéité des données collectées dans le projet VILAGIL (données de capteurs dans la ville images et vidéos de surveillance, données de véhicules autonomes, etc), nous souhaitons proposer une architecture d'apprentissage fédéré qui considère les clients de données en fonction de leur nature (cross-device / cross-silo), et de la nature de distribution des données (verticale / horizontale). Cette architecture devrait assurer à terme un apprentissage personnalisé pour chacun des clients et une équité par rapport à la collaboration dans l'apprentissage.

Nous envisageons d'élaborer cette architecture en combinant à la fois une première étape de clustering de clients et une deuxième étape d'apprentissage machine fédéré. Le clustering des clients est souvent appliqué dans l'objectif d'identifier un sous ensemble de clients qui auraient des modèles d'apprentissage similaires sans divulguer leurs données. Les approches [1] et [2] proposent des méthodes appliquées dans le cadre de données non-IID qui calculent les similarités qu'il y aurait entre les paramètres de modèles pour regrouper les clients. L'objectif de ses approches est de personnaliser les modèles au niveau des nœuds locaux en tenant compte de distributions hétérogènes dans les données. En s'inspirant des travaux récents de [1], [2] l'architecture proposée devrait être comparée expérimentalement par rapport à d'autres approches existantes de la littérature [3] et démontrant son intérêt.

Le postdoc aura pour mission :

- De développer une preuve de concept pour démontrer l'architecture sur des jeux de données identifiés pertinents par rapport au projet (jeux de données réels de partenaires et/ou jeux de données issus de la littérature scientifique du domaine).
- De participer à des réunions du projet VILAGIL au niveau de l'Action « Data & Mobilité » et des réunions de coordination entre les différentes actions.
- De positionner ses travaux et les contributions par rapport à la littérature scientifique du domaine de l'apprentissage fédéré, et de publier les contributions dans des journaux et conférences internationales de premier rang mondial

Références Bibliographiques

[1] Yoo, Joo Hun et al. "Personalized Federated Learning with Clustering: Non-IID Heart Rate Variability Data Application." *2021 International Conference on Information and Communication Technology Convergence (ICTC) (2021)*: 1046-1051.

[2] Huang, Yutao et al. "Personalized Cross-Silo Federated Learning on Non-IID Data." *AAAI (2021)*.

[3] Hangyu Zhu, Jinjin Xu, Shiqing Liu, Yaochu Jin. Federated learning on non-IID data: A survey, *Neurocomputing*, Volume 465, 2021, Pages 371-390, ISSN 0925-2312,

Encadrement et environnement

- **Laboratoire** : Institut de Recherche en Information de Toulouse, IRIT, CNRS/UMR5505
- **Direction scientifique** :
Imen Megdiche, Equipe SIG (Imen.Megdiche@irit.fr)
André Péninou, Equipe SIG (Andre.Peninou@irit.fr)
Olivier Teste, Equipe SIG (Olivier.Teste@irit.fr)
- **Lieu de travail** : IRIT Site Paul Sabatier, 118 Route de Narbonne, 31062 Toulouse, FRANCE
- **Durée** : 1 ans
- **Démarrage** : possible dès Décembre 2022.

Dossier de candidature

Nous recherchons un(e) candidat(e) motivé(e), force de propositions, avec un solide background en mathématiques appliquées et des connaissances confirmées en IA.

Les candidatures sont à envoyer à (imen.megdiche@irit.fr) et (olivier.teste@irit.fr) . Le dossier doit contenir :

- CV incluant les publications scientifiques
- Lettre de motivation

- Copie du dernier diplôme
- Pré-rapport et rapport de soutenance de thèse
- Une expérience en federated learning serait un atout.