

Apprentissage fédératif pour la prédiction du churn : une évaluation

Sébastien Godard *, Nicolas Voisine**
Tanguy Urvoy**, Vincent Lemaire**

*Université de Bretagne Sud, **Orange Labs Lannion

Résumé. Les smartphones sont omniprésents dans notre quotidien. Ils constituent une ressource informatique à portée de la main avec un accès direct à une quantité considérable d'informations personnelles. Ils représentent une source de données très précieuse pour les opérateurs de télécommunication, mais la nature très décentralisée de ces données et les attentes évidentes des clients en matière de respect de la vie privée requièrent de nouvelles approches en apprentissage statistique. L'apprentissage ubiquitaire (ou *ubiquitous datamining*) qui intègre des terminaux avec la capacité de traiter localement leurs propres données, est une alternative intéressante à la centralisation de masse en traitement de données. L'apprentissage fédératif, (ou *federated learning*) est une réalisation de l'ubiquitous datamining, qui permet de déployer certains modèles d'apprentissage automatique sur des terminaux "autonomes" tels les smartphones. Cet article propose une évaluation détaillée de ce type d'apprentissage distribué dans le cadre de la prédiction de l'attrition (ou *churn*) sur des données issues d'un opérateur télécom.

1 Introduction

Omniprésents dans notre quotidien, les smartphones représentent une véritable ressource informatique à portée de la main. Pour de nombreuses personnes, il s'agit des appareils informatiques les plus utilisés. Fréquemment sollicités, ils ont accès à une quantité considérable de données issues de leurs puissants capteurs (appareils photos, microphones, GPS, etc.), ainsi qu'à des objets connectés auxquels ils sont reliés. Cette riche source de données est potentiellement porteuse de nombreuses promesses pour les opérateurs télécoms et leurs clients : elle permet de contrôler au plus près la qualité des services et leur ergonomie, elle permet aussi d'envisager des services innovants par apprentissage statistique. Mais ces données sont aussi, et surtout, des données très personnelles et très sensibles.

L'entrée en vigueur du RGPD (règlement général sur la protection des données European Union, 2016) redéfinit complètement la politique de gestion des données. Il n'est plus question de recueillir massivement les données de clients sans objectif précis de service. Le RGPD fixe le cadre juridique relatif à la protection des données personnelles au sein de l'Union Européenne. Responsabilisant davantage les entreprises, le RGPD donne de nouvelles obligations aux opérateurs de services quant à la gestion des données, rendant notamment leur centralisa-