



UMEÅ UNIVERSITET

Advancing Federated Learning: Algorithms and Use-Cases

Sourasekhar Banerjee

Akademisk avhandling

som med vederbörligt tillstånd av Rektor vid Umeå universitet för avläggande av filosofie doktorsexamen framläggs till offentligt försvar i Hörsal HUM.D.210, Humanisthuset, den 23 September 2024, kl. 13:00.

Avhandlingen kommer att försvaras på engelska.

Fakultetsopponent: Associate Professor Salman Toor (Docent),
Institutionen för informationsteknologi, Uppsala Universitet,
Uppsala, Sweden

Organization
Umeå University
Dept. of Computing Science

Document type
Doctoral thesis

Date of publication
02 September 2024

Author
Sourasekhar Banerjee

Title
Advancing Federated Learning: Algorithms and Use-Cases

Abstract

Federated Learning (FL) is a distributed machine learning paradigm that enables the training of models across numerous clients or organizations without requiring the transfer of local data. This method addresses concerns about data privacy and ownership by keeping raw data on the client itself and only sharing model updates with a central server. Despite its benefits, federated learning faces unique challenges, such as data heterogeneity, computation and communication overheads, and the need for personalized models. Thereby results in reduced model performance, lower efficiency, and longer training times.

This thesis investigates these issues from theoretical, empirical, and practical application perspectives with four-fold contributions, such as federated feature selection, adaptive client selection, model personalization, and socio-cognitive applications. Firstly, we addressed the data heterogeneity problems for federated feature selection in horizontal FL by developing algorithms based on mutual information and multi-objective optimization. Secondly, we tackled system heterogeneity issues that involved variations in computation, storage, and communication capabilities among clients. We proposed a solution that ranks clients with multi-objective optimization for efficient, fair, and adaptive participation in model training. Thirdly, we addressed the issue of client drift caused by data heterogeneity in hierarchical federated learning with a personalized federated learning approach. Lastly, we focused on two key applications that benefit from the FL framework but suffer from data heterogeneity issues. The first application attempts to predict the level of autobiographic memory recall of events associated with the lifelog image by developing clustered personalized FL algorithms, which help in selecting effective lifelog image cues for cognitive interventions for the clients. The second application is the development of a personal image privacy advisor for each client. Along with data heterogeneity, the privacy advisor faces data scarcity issues. We developed a daisy chain-enabled clustered personalized FL algorithm, which predicts whether an image should be shared, kept private, or recommended for sharing by a third party.

Our findings reveal that the proposed methods significantly outperformed the current state-of-the-art FL algorithms. Our methods deliver superior performance, earlier convergence, and training efficiency.

Keywords

Federated Learning, Federated Feature Selection, Statistical Heterogeneity, System Heterogeneity, Model Personalization, and Socio-Cognitive Applications

Language
English

ISBN
Print: 978-91-8070-463-2
PDF: 978-91-8070-464-9

ISSN
0348-0542

Number of pages
84 + 6 papers