Federated machine learning: simulation platform development and performance evaluation

This master thesis opportunity is jointly provided by Chalmers E2 and RISE, Research Institute of Sweden.

Federated learning (FL) is a new privacy-preserving machine learning paradigm where many clients (e.g. mobile devices or local servers) collaboratively train a model under the orchestration of a central server (e.g. service provider) or through peer-to-peer communication, while keeping the training data local and decentralized. While achieving high privacy, the scale and heterogeneity of decentralized data brings new challenges in research areas of machine learning. Specially, it is hard for researchers to establish a real large-scale system for research purposes at the initial phase of their study. In this regard, simulation and emulation are the tools that can help accelerate this process. Some simulation frameworks and libraries for FL have been developed, such as Tensorflow Federated, LEAF and PySyft. One of the key components of FL is the model aggregation and averaging method, where algorithms such as FedAvg and FedPro have been proposed and evaluated. However, the performance of these algorithms has been evaluated based on different simulation frameworks, such as their self-developed frameworks. Thus, it is hard to compare their performance fairly without considering the effects of simulation framework.

This thesis includes two tasks, 1) integrating the existing FL algorithms into a chosen framework (i.e. LEAF), 2) analyzing the performance of these algorithms under the same framework, 3) investigating application potential in industry areas.

The candidates, who have some basic knowledges of machine learning and programming skills with Python in Linux environment are preferred. The experiments of developing machine learning related tasks with Tensorflow or Pytorch will be a merit.

Contact info: Dr. Jun Li (ljun@chalmers.se) and Dr. Lei Chen (lei.chen@ri.se)