Chair of Applied Computer Science X

Prof. Dr. Ruben Mayer 07/2025

UNIVERSITÄT BAYREUTH

Bachelor/Master Thesis

Non-uniform Data Pre-processing for Federated Learning

Background

Federated Learning (FL) is a technique for privacy-enhancing machine learning collaboration. Clients train a shared model locally, and communicate their updates via a central server, which has no data access. FL is typically benchmarked with a centralized dataset that is artificially distributed to the clients. However, in a real FL system, clients would gather data locally, not from a global repository. This implies that they may implement their own individual preprocessing strategies. For example, there may be missing data, e.g., NULL values in SQL data, which they deal with in different ways.

Task definition

Based on a comprehensive literature review, conceptualize and perform a benchmarking of the FL system: For each client, implement local data loaders and pre-processing. Evaluate the system on various scenarios to determine its robustness and quality.

Literature review: Perform a literature review, investigating FL, metrics, data pre-processing, and existing approaches relevant to the problem.

Implementation: This complex task requires knowledge of machine/federated learning with python (desired: flower), SQL, docker. Self-organization and a structured approach are essential. Implement and benchmark FL in python (with flower). Using SQL data, implement local data loading and pre-processing. Evaluate your implementations on established metrics and perform a benchmarking of the system. Document literature review, concept, approach, and implementation in your thesis in German or English.

Required Knowledge and Skills

Necessary: Python (flower), Konzepte der Programmierung, docker
Desired: Machine Learning, SQL, bash

Supervisor

Simon Mertel, M.Sc.
Feel free to reach out via Teams or e-mail (simon.mertel@uni-bayreuth.de)