technische universität dortmund

Fraunhofer ISST

Thesis Proposal

Decentralized Federated Learning: Enhancing IFCA for Efficient Clustered Model Training

Jonas Kirch

October 9, 2024

Reviewer:
Prof. Dr. Stefan Harmeling
M.Sc. Sebastian Becker

TU Dortmund University
Chair of Artificial Intelligence

Fraunhofer Institute for Software and Systems Engineering
Department Mobility and Smart Cities

1 Introduction

Edge- and distributed computing have become a crucial component in many data-demanding applications, such as real-time analytics in IoT devices, autonomous vehicles or smart cities, only naming a few of them. In most of these contexts, data is stored locally on the devices of end users which significantly enhances the difficulty of applying machine learning based on user data due to data privacy restrictions. Since 2016, Federated Learning (FL) (McMahan et al., 2017) [1], a distributed solution to machine learning introduced by Google researchers, has emerged as a promising approach that allows models to be trained collaboratively across multiple devices while keeping user data decentralized and private, thereby addressing privacy concerns. Over the past years, researchers have optimized FL in different contexts and applications, e.g. in those mentioned above.

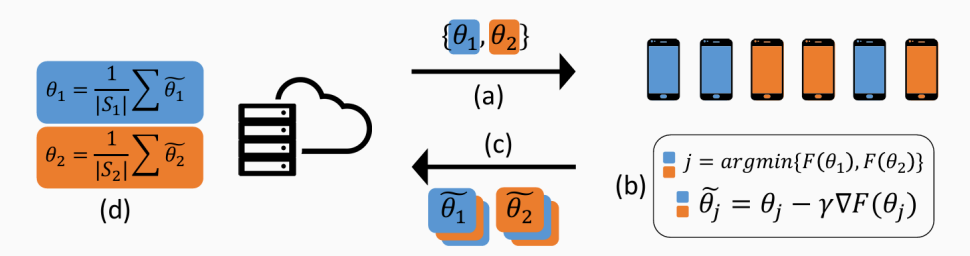

Figure 1: An overview of IFCA [2]. (a) The server broadcast models. (b) Worker machines identify their cluster memberships and run local updates. (c) The worker machines send back the local models to the server. (d) Average the models within the same estimated cluster Sj.

This thesis is intended to look for ways to optimize the in Figure 1 illustrated Iterative Federated Clustering Algorithm (IFCA) (Ghosh et al. 2021) [2], especially in time efficiency and evaluation. The main idea is to use decentralized FL within the clusters and picking a representative model at the end of training. The model chosen for each inference is going to be determined by a confidence score, similar to the evaluation for test machines in IFCA.

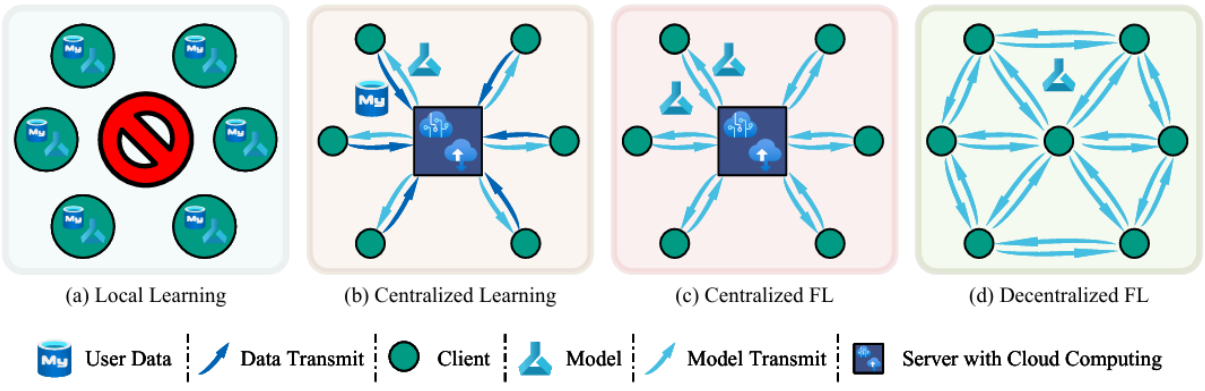

Figure 2: Illustration of different Federated Learning communication schemas [4]

The Decentralized Federated Learning: A Survey and Perspective (Yuan et al. 2024) [4] paper will assist in selecting the communication concepts and schemas for the clusters. Communication is a key element in decentralized Federated Learning and will have a significant impact on this initiative. Figure 2 shows the possible communication schemas for Federated Learning, with (d) being the most important one for this thesis.

2 Literature

The *Iterative Federated Clustering Algorithm* (IFCA) (Ghosh et al. 2021) [2] is one solution to Federated Learning with heterogeneous data among devices in a network of devices, also known as non-i.i.d. data. In this clustered federated learning approach, clusters of homogeneous data are identified by comparing the models after each training round. Models that show similar characteristics are assigned to their respective clusters. At the end of the training process, a cluster-specific model is aggregated from the individual models within each cluster, resulting in tailored models that serve to the unique data distributions of their respective groups. These refined models can then be used for inference on new data effectively.

As I intend to take a serverless, decentralized approach to Clustered Federated Learning, I am going to draw upon the research carried out in the paper *Decentralized Federated Learning: A Survey and Perspective* (Yuan et al. 2024) [4] where researchers introduce different concepts and algorithms for Decentralized Federated Learning. Here, I will discuss which concept or algorithm is best suited for IFCA and the thesis.

3 Methodology

The following section depicts the process by which this project aims to implement and test the approach proposed above. The hypothesis provides an expectation of the results, and the goals describe what this project aims to achieve.

3.1 Hypothesis

A decentralized approach to the IFCA will achieve comparable or improved performance and efficiency compared to the traditional centralized IFCA.

3.2 Goals

- More efficient, decentralized or serverless Clustered Federated Learning with similar performance to centralized IFCA.
- A complete implementation of decentralized IFCA, with confidence based inference for new data.

3.3 Methods

This research project will be implemented in Python3 using Machine Learning libraries, mainly Pytorch and Numpy. Performance evaluation will be carried out following the methods of references [2] and [3] to determine the extent to which the decentralized IFCA performance compares to that of traditional IFCA. Ultimately, the confidence-based inference discussed will be included to gain further understanding of the test outcomes.

4 References

[1] McMahan and Ramage. Federated learning: Collaborative machine learning without centralized training data. https://arxiv.org/abs/1602.05629, 2016

[2] Ghosh, Chung, Yin, Ramchandran: An Efficient Framework for Clustered Federated Learning. https://arxiv.org/abs/2006.04088, 2020

[3] Jichan Chung: Code for the IFCA-Paper. https://github.com/jichan3751/ifca, 2020

[4] Yuan, Wang, Sun, Yu, Brinton: Decentralized Federated Learning: A Survey and Perspective. https://arxiv.org/abs/2306.01603, 2023