A multi-objective optimization framework for decentralized learning with coordination constraints

Roberto Morales Ponce (UD, Spain)

Supported by CoDeFeL ERC project.

December 2, 2025

European Commission

European Research Council
Established by the European Commission

Deusto
Universidad de Deusto
Deustuko Unibertsitatea
University of Deusto

Outline

Introduction

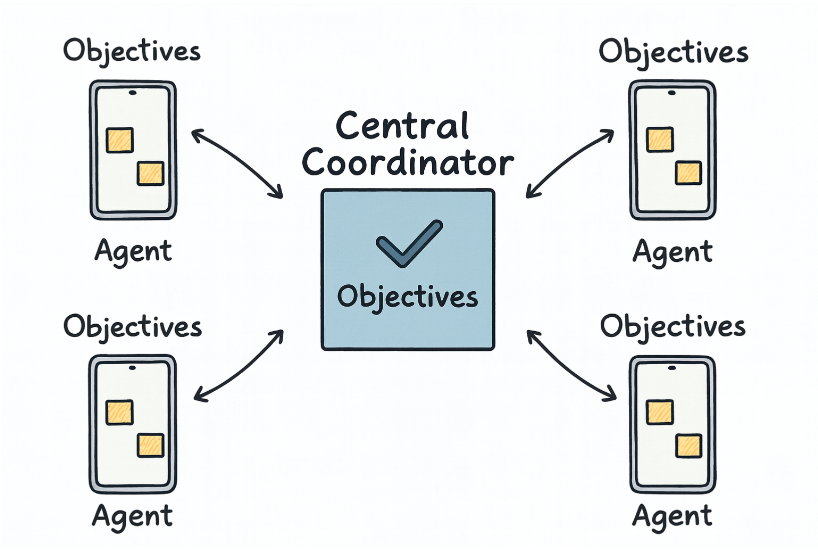
Motivation: Why Decentralized Learning?

- Modern data is **distributed by nature** across devices, institutions, or geographic regions.
- Centralized training is often **infeasible** due to privacy constraints, communication costs, or storage limitations.
- Decentralized learning offers a **paradigm** where agents keep their data locally and collaborate only through model updates.
- This framework is especially **relevant** in healthcare, finance, and mobile-device ecosystems.

Challenges in Classical Federated Learning(FL)

- Standard FL focuses mainly on global performance, often ignoring agent-specific objectives.
- Real systems involve heterogeneous agents:
  1. different local datasets
  2. different priorities
  3. different performance metrics
- A single aggregated objective may cause unfairness or poor local performance.

Learning Task



Mathematical formulation

Setting

Let $m \in \mathbb{N}$. Each agent $i \in [M] := \{1, \ldots, M\}$ has a dataset

$$D^i := \{(x_l^i, y_l^i) \in \mathbb{R}^{d_1} \times \mathbb{R}^{d_2}\}_{l=1}^{n_i},$$

over which it trains a ML model $f_\Theta : \mathbb{R}^{d_1} \to \mathbb{R}^{d_2}$, where $\Theta \in \mathcal{U}$. For simplicity, we assume that

$$n_1 = n_2 = \ldots = n_M.$$

We consider:

- **Agent-specific objectives**: $\{C_i\}_{i=1}^{M}$,
- **Central coordinator criteria:** $\{S_j\}_{j=1}^{N}$, $N \in \mathbb{N}$.

Learning Task

We consider the following Multi Objective Optimization (MOO) problem:

$$\min_{\Theta \in \mathcal{U}} \left(C_1(\Theta), \ldots, C_M(\Theta), S_1(\Theta), \ldots, S_N(\Theta) \right).$$

- Under certain conditions, the **existence** of (Global) Pareto Optimal Solutions are granted.
- Pareto Optimal Solutions are typically **not unique**.
- We focus on the use of of **scalarization** to select among the possible efficient solutions.

Scalarization

Let $\lambda \in [0,1)$. We consider the scalar problem

$$\min_{\Theta \in \mathcal{U}} \left\{ \frac{1-\lambda}{M} \sum_{i=1}^{M} C_i(\Theta) + \frac{\lambda}{N} \sum_{j=1}^{N} S_j(\Theta) \right\}.$$

The parameter $\lambda \in [0,1)$ serves as a weight that modulates the balance between local agent objectives and the coordinator's preferences.

We shall rewrite the problem in the equivalent form:

$$\min_{\Theta \in \mathcal{U}} \frac{1}{M} \sum_{i=1}^{M} F_i(\Theta),$$

where

$$F_i(\Theta) = (1-\lambda) C_i(\Theta) + \frac{\lambda M}{N} \sum_{j=1}^{N} S_j(\Theta), \quad \forall i \in [M].$$

Notation

- $\Theta^0 \in \mathcal{U}$: initial guess for the parameters
- $T \in \mathbb{N}^*$: total number of iterations
- $\tau \in \mathbb{N}^*$: total number of epochs for the agents' updates
- $\eta > 0$: learning rate for the agents' updates
- $\lambda \in [0, 1)$: agents-coordinator trade-off.

Algorithm (Inspired in FedAvg)

for $t \in [T]$ **do**
    for $i \in [M]$ **do**
        Set $\Theta_i^{t-1,0} \leftarrow \Theta^0$.
        for $k \in [\tau]$ **do**
            Compute $g_i^{t-1,k-1}$: s.g. of $C_i(\Theta_i^{t-1,k-1})$.
            Compute $h_i^{t-1,k-1}$: s.g. of $\sum_{j=1}^{N} S_j(\Theta_i^{t-1,k-1})$.
            Update:

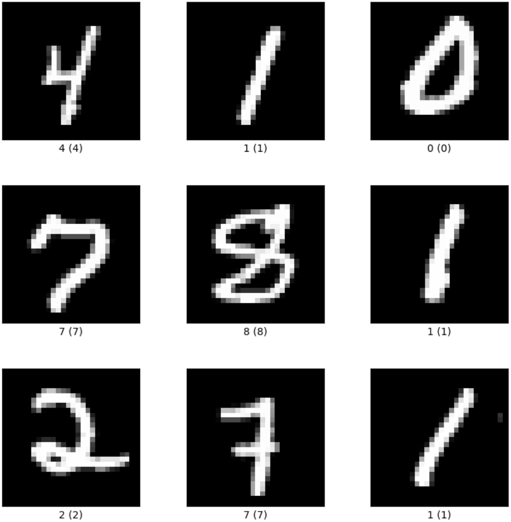
$$\Theta_i^{t-1,k} \leftarrow \Theta_i^{t-1,k-1} - \eta \left(g_i^{t-1,k-1} + \alpha h_i^{t-1,k-1} \right).$$

    Set $\Theta^t = \sum_{i=1}^{M} \Theta_i^{t-1,\tau}$

- Under suitable assumptions, we can quantify how far the algorithm's output is from a weak Pareto solution.
- For more details, see the preprint: Biccari, U. & Morales, R. (2025). A Multi-Objective Optimization framework for Decentralized Learning with coordination constraints. arXiv preprint arXiv:2507.13983.

Experimental evaluation

MNIST dataset



- Each agent trains a **Convolutional Neural Network**,
- The learning rate is fixed to $\eta = 0.001$.
- We use 50 rounds.
- We use 1 epoch.

IID data distribution

In this experiment, we consider

- 5 agents.
- Each agent receives 8000 images for training, 2000 for local validation, and 2000 for local test.
- The data is drawn randomly across all digit classes.

We suppose the coordinator has the following single objective

$$S_1(\Theta) := 10^2 \|\Theta\|_2^2, \quad \Theta \in \mathcal{U}.$$

Validation accuracy

	$t = 1$	$t = 10$	$t = 20$	$t = 30$	$t = 40$	$t = 50$
Agent 1	0.3900	0.9367	0.9367	0.9429	0.9371	0.9396
Agent 2	0.3804	0.9317	0.9346	0.9371	0.9317	0.9325
Agent 3	0.3825	0.9342	0.9379	0.9392	0.9417	0.9383
Agent 4	0.3842	0.9267	0.9279	0.9329	0.9313	0.9317
Agent 5	0.3842	0.9371	0.9383	0.9396	0.9387	0.9363

Table: Validation accuracy of each agent after 1, 10, 20, 30, 40 and 50 rounds, with $\lambda = 0.87$ and IID data distribution.

Validation F1 score

	$t = 1$	$t = 10$	$t = 20$	$t = 30$	$t = 40$	$t = 50$
Agent 1	0.3966	0.9359	0.9362	0.9422	0.9364	0.9390
Agent 2	0.3912	0.9307	0.9337	0.9361	0.9307	0.9348
Agent 3	0.3867	0.9331	0.9372	0.9384	0.9410	0.9377
Agent 4	0.3929	0.9264	0.9280	0.9325	0.9383	0.9316
Agent 5	0.3887	0.9366	0.9381	0.9391	0.9383	0.9358

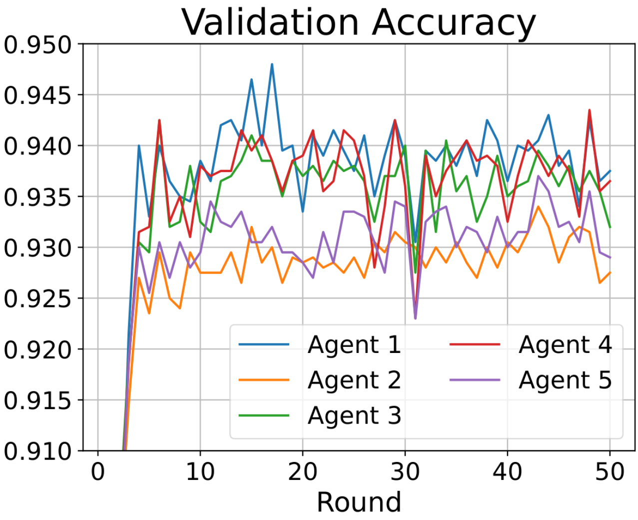
Table: Validation F1 score of each agent after 1, 10, 20, 30, 40 and 50 rounds, with $\lambda = 0.87$ and IID data distribution.

The influence of λ

λ	0.00	0.25	0.50	0.65	0.75	0.87
Acc.	0.9837	0.9846	0.9728	0.9657	0.9538	0.9409
F1	0.9836	0.9845	0.9727	0.9655	0.9535	0.9404

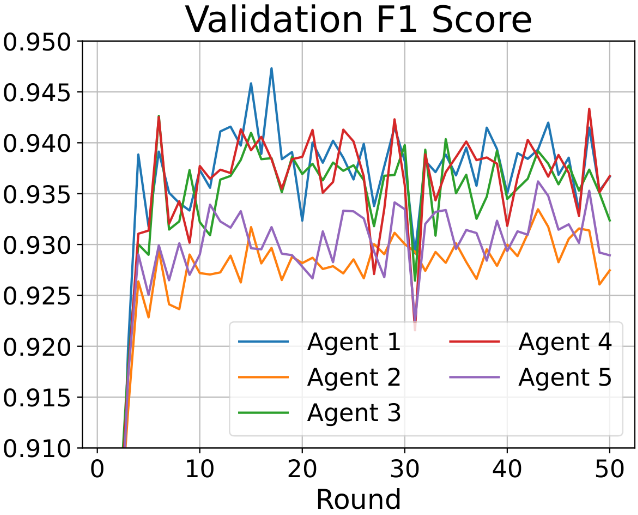
Table: Test accuracy and F1 score for different values of λ after 50 rounds with IID data distribution.

IID case



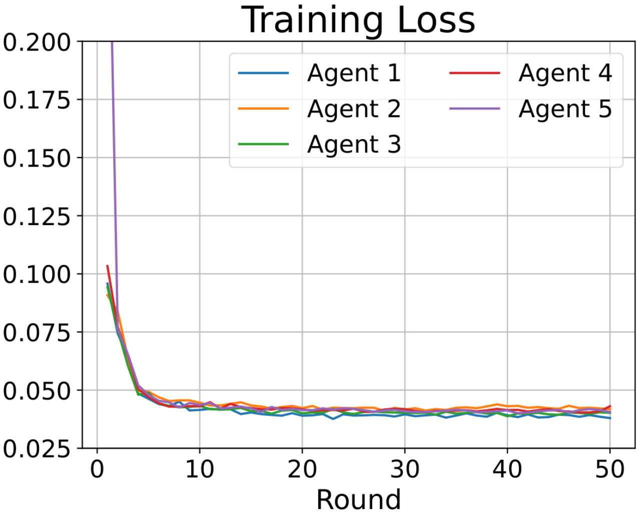
Validation Accuracy
0.950
0.945
0.940
0.935
0.930
0.925
0.920
0.915
0.910
0
10
20
30
40
50
Round
Agent 1
Agent 2
Agent 3
Agent 4
Agent 5

IID case



Validation F1 Score
0.950
0.945
0.940
0.935
0.930
0.925
0.920
0.915
0.910
0
10
20
30
40
50
Round
Agent 1
Agent 2
Agent 3
Agent 4
Agent 5

IID case


Training Loss
0.200
0.175
0.150
0.125
0.100
0.075
0.050
0.025
0
10
20
30
40
50
Round
Agent 1
Agent 2
Agent 3
Agent 4
Agent 5

The non-IID case

In this consider the case where

- Agent 1: digits 2 and 8;
- Agent 2: digits 4 and 9;
- Agent 3: digits 1 and 6;
- Agent 4: digits 3 and 7;
- Agent 5: digits 0 and 5.

Each agent receives 8000 images for training, 2000 for local validation, and 2000 for local test.
The coordinator's objective in this case is given by

$$S_1(\Theta) = 10^7 \|\Theta\|_2^2, \quad \forall \Theta \in \mathcal{U}.$$

Validation accuracy

	$t = 1$	$t = 10$	$t = 20$	$t = 30$	$t = 40$	$t = 50$
Agent 1	0.4790	0.6450	0.6805	0.6330	0.6950	0.7670
Agent 2	0.0000	0.2735	0.5040	0.7295	0.7875	0.7575
Agent 3	0.0000	0.7390	0.8295	0.9195	0.9120	0.9015
Agent 4	0.0000	0.2720	0.3875	0.4705	0.3915	0.4390
Agent 5	0.0000	0.5610	0.6570	0.7215	0.7105	0.7160

Table: Validation accuracy of each agent after 1, 10, 20, 30, 40 and 50 rounds, with $\lambda = 0.87$ and non-IID data distribution.

Validation F1 score

	$t = 1$	$t = 10$	$t = 20$	$t = 30$	$t = 40$	$t = 50$
Agent 1	0.3239	0.1531	0.1601	0.1531	0.1633	0.1736
Agent 2	0.0000	0.0783	0.1211	0.1683	0.1745	0.1692
Agent 3	0.0000	0.1884	0.2015	0.2129	0.2726	0.2371
Agent 4	0.0000	0.0854	0.1227	0.1228	0.1126	0.1209
Agent 5	0000	0.1392	0.1550	0.1649	0.1630	0.1620

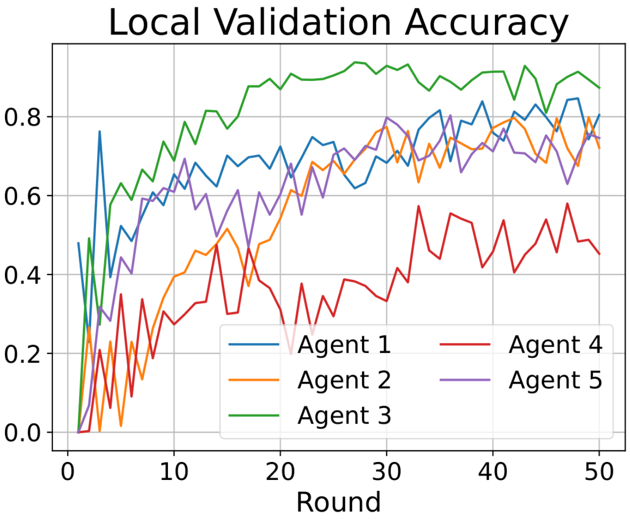
Table: Validation F1 score of each agent after 1, 10, 20, 30, 40 and 50 rounds, with $\lambda = 0.87$ and non-IID data distribution.

The influence of λ

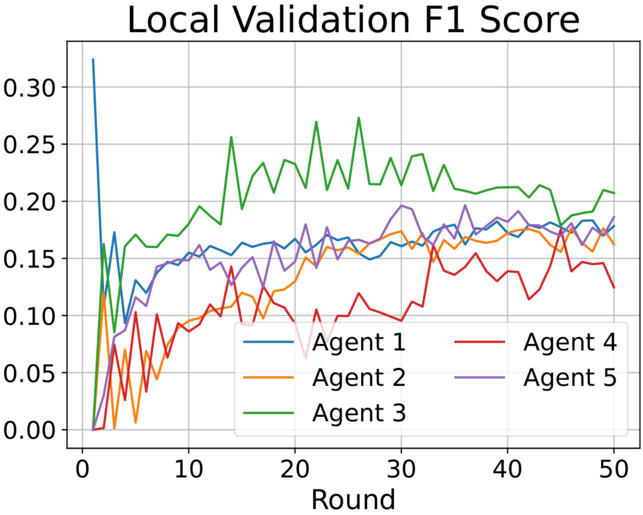
λ	0.00	0.25	0.50	0.65	0.75	0.87
Acc.	0.6600	0.6943	0.7771	0.7212	0.7505	0.8056
F1	0.6389	0.6820	0.7755	0.6919	0.7366	0.7999

Table: Test accuracy and F1 score for different values of λ after 50 rounds with non-IID data distribution.

Non-IID case

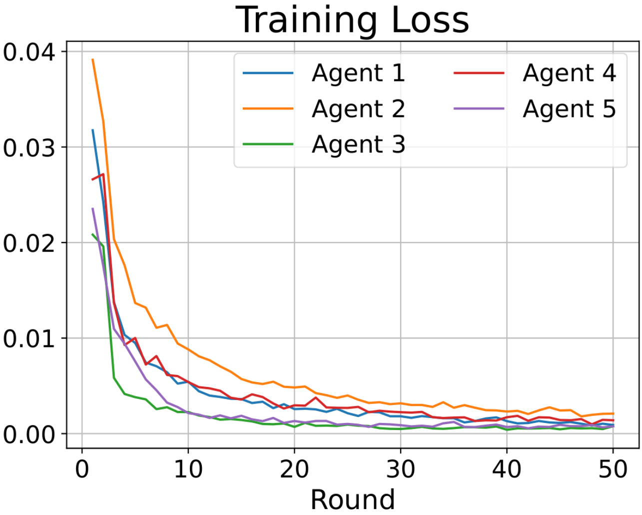

Local Validation Accuracy
0.0
0.2
0.4
0.6
0.8
0
10
20
30
40
50
Round
Agent 1
Agent 2
Agent 3
Agent 4
Agent 5

Non-IID case



Local Validation F1 Score
0.30
0.25
0.20
0.15
0.10
0.05
0.00
0
10
20
30
40
50
Round
Agent 1
Agent 2
Agent 3
Agent 4
Agent 5

Non-IID case



Training Loss
0.04
0.03
0.02
0.01
0.00
0
10
20
30
40
50
Round
Agent 1
Agent 2
Agent 3
Agent 4
Agent 5

Thanks for your attention!

The CoDeFeL project (ERC-2022-ADG) has received funding from the European Union's Horizon ERC Grants programme under grant agreement No. 101096251. The views and opinions expressed are solely those of the author(s) and do not necessarily reflect those of the European Research Council Executive Agency (ERCEA), the European Union or the granting authority who cannot be held responsible for them.

A preprint version of the article can be found in https://arxiv.org/pdf/2507.13983.

European Research Council

Established by the European Commission

European Commission

Deusto

Universidad de Deusto
Deustuko Unibertsitatea
University of Deusto