

Aprententatge Automàtic per a Xarxes (ML4Net)

Seminar 5 - Federated Learning

Francesc Wilhelmi & Boris Bellalta

School of Engineering
Universitat Pompeu Fabra



**Universitat
Pompeu Fabra**
Barcelona

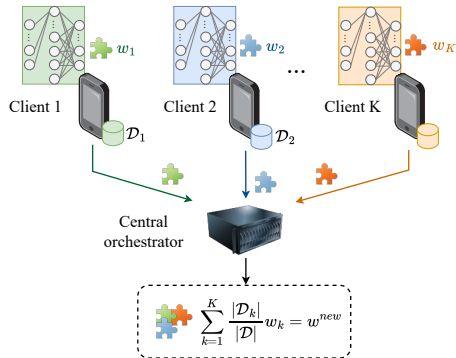
Federated Learning

Key Idea (Iterative Process):

1. Central server sends a global model $\omega(t)$ to selected clients.
2. Clients train $\omega(t)$ on their local data D_k , obtaining local update $\omega_k(t)$.
3. Clients send $\omega_k(t)$ back to central server.
4. Central server aggregates local updates to create a new global model $\omega(t+1)$.

Objective: $\min_{\omega} (F(\omega) = \sum_{k=1}^K \frac{|\mathcal{D}_k|}{|\mathcal{D}|} F_k(\omega))$

- K : total number of clients.
- $|\mathcal{D}_k|$: number of data samples on client k .
- $|\mathcal{D}| = \sum_{k=1}^K |\mathcal{D}_k|$: total number of data samples across all clients.
- $F_k(\omega)$: local loss function on client k 's data \mathcal{D}_k .

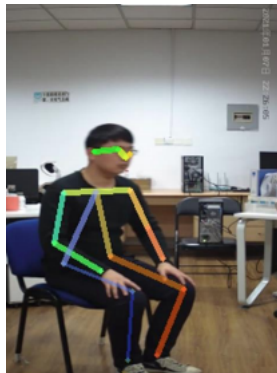


Dataset (I)

How were the data generated?^a

- Real Wi-Fi measurements were taken when people were doing specific poses.
 - Walk, sit down, run, stand up, bend...
- The features are the Wi-Fi measurements (the CSI, in this case).
- The labels (poses) were taken using cameras (Alphapose).

^aZhou, Y., Xu, C., Zhao, L., Zhu, A., Hu, F., & Li, Y. (2022). CSI-former: Pay more attention to pose estimation with WiFi. *Entropy*, 25(1), 20.



Dataset (II)

- Train features (`client_k_features.csv`, inside `client_datasets` folder): CSI measurements in the training dataset.
 - A random number of samples with a flattened CSI matrix of $30 \times 3 \times 3$.
- Train labels (`client_k_labels.csv`, inside `client_datasets` folder): actual pose in the training dataset.
 - A random number of samples with an integer between 1 and 12 (corresponding to poses {'wave', 'push', 'crouch', 'sitdown', 'bend', etc.})
- Test features (`test_features.csv`): CSI measurements in the test dataset.
 - 500 samples \times 270 (a flattened CSI matrix of $30 \times 3 \times 3$)
- Test labels (`test_labels.csv`): actual pose in the test dataset.
 - 500 samples of an integer between 1 and 12 (corresponding to poses {'wave', 'push', 'crouch', 'sitdown', 'bend', etc.})

In this seminar, we will use a simplified version of a dataset that can be found at:
<https://github.com/NjtechCVLab/Wi-PoseDataset?tab=readme-ov-file>

Dataset (III)

- Data is distributed among $N = 10$ clients (e.g., `client_1_features.csv`, `client_1_labels.csv`)
- Each client has data from $[2, 8]$ classes.
- Between 10 and 100 samples are selected per class.
- The test dataset is apart and contains 500 samples (randomly selected among all the classes).

