DATI, AI E ROBOTICA @POLITO

RICERCA, TRASFERIMENTO TECNOLOGICO E SUPPORTO ALLE AZIENDE SUI TEMI FONDAMENTALI DEI BIG DATA, INTELLIGENZA ARTIFICIALE, ROBOTICA E RIVOLUZIONE DIGITALE

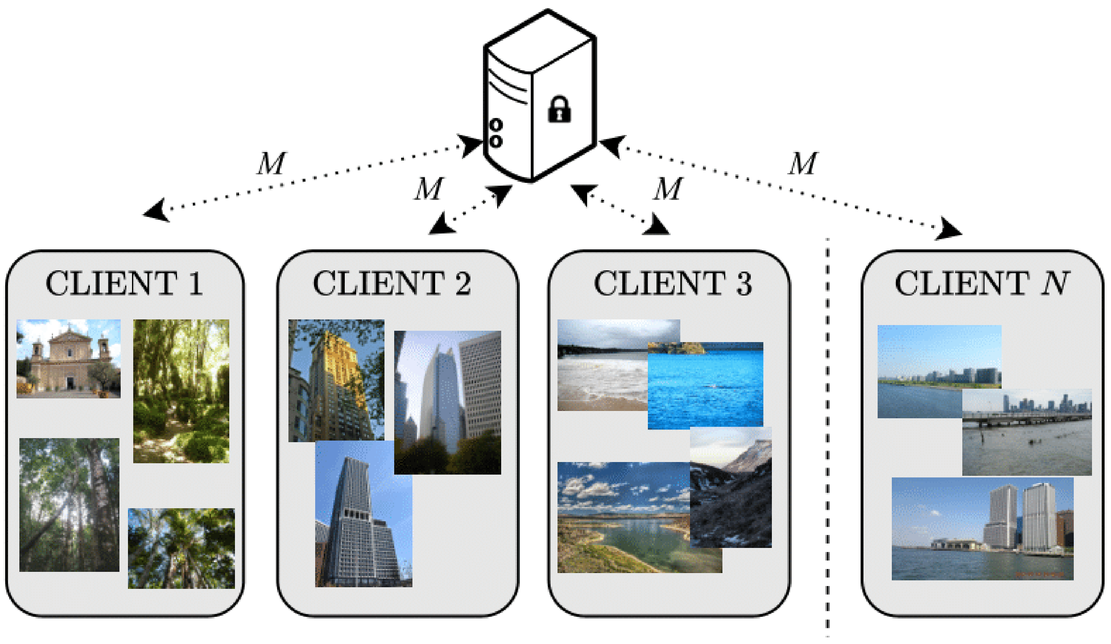
Cluster-driven Graph Federated Learning over Multiple Domains

D. Caldarola, M. Mancini, F. Galasso, M. Ciccone, E. Rodolà, B. Caputo

ARTIFICIAL INTELLIGENCE HUB

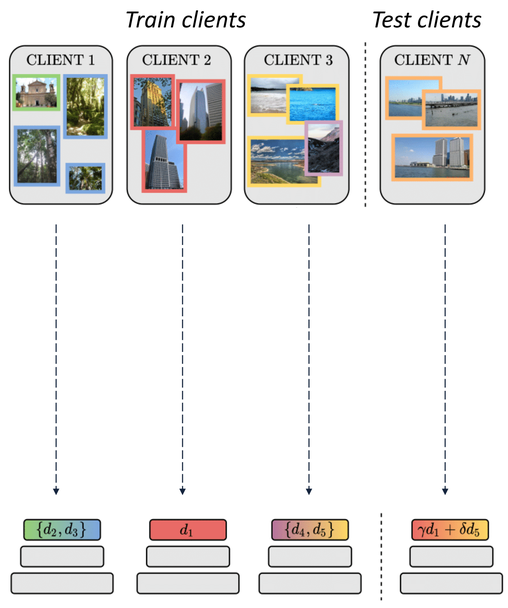
Federated Learning

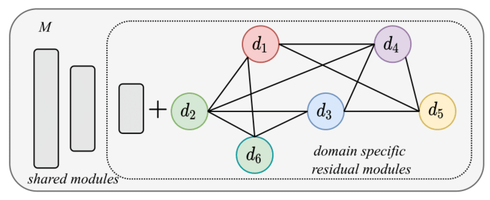
Federated Learning (FL) deals with learning a global model *M* on the server-side in **privacy-constrained scenarios**, where data are stored on edge devices, i.e. the *clients*.
Standard algorithm for models' aggregation: **FedAvg**.



FedCG

1. **Federated Clustering** of the local distributions, i.e. ***domains***. Based on pseudo-labels predicted by the teacher-student domain classifiers.
   Test time: new domains as a soft combination of the previously discovered ones.
2. **Cluster-specific** models: domain specific parameters added as residuals in the model.
3. **Graph-based** interactions among domain-specific parameters.





Statistical heterogeneity in FL

In realistic scenarios, clients may hold different data distributions, leading to poor performances of the global model. Simply averaging the updates' weights is not enough anymore.

How to deal with **non-i.i.d.** and **unbalanced** clients' data?

McMahan, Brendan, et al. "Communication-efficient learning of deep networks from decentralized data." *Artificial intelligence and statistics. PMLR*, 2017.

Kipf, Thomas N., and Max Welling. "Semi-supervised classification with graph convolutional networks." *ICLR* (2016).

Caldas, Sebastian, et al. "Leaf: A benchmark for federated settings." *arXiv preprint arXiv:1812.01097* (2018).

Results

Datasets statistics. Task: image classification.

Dataset	Clients	Total samples	Samples per client		Classes
			Mean	Stdev	
CelebA	9,343	200,288	21.44	7.63	2
FEMNIST	3,550	805,263	226.83	88.94	62

Table 1: **Datasets Statistics. Origin: LEAF benchmark.**

Ablation studies

Model	A	W	ReLU	Acc(%)
Domain-specific models	-	-	-	33.61
GCN	U	✗	✗	84.39
	H	✗	✗	82.10
	U	✓	✗	87.92
	H	✓	✗	84.25
FedCG	U	✓	✗	86.96
	H	✓	✗	88.65
	U	✓	✓	87.97
	H	✓	✓	**89.57**

Experiments

Dataset	Model	Accuracy (%)
CelebA	FedAvg	86.88
	FedCG	**89.18**
FEMNIST	FedAvg	77.81
	FedCG	**83.41**
	FedProx	75.00
	SCAFFOLD	84.20

Conclusions

We propose *FedCG*, a new algorithm for addressing **statistical heterogeneity** in FL, based on:

- an iterative **clustering** algorithm used to identify different data distributions and instantiate **domain-specific parameters**.
- **Graph Convolutional Neural Networks** for **sharing knowledge** across domains.
- the possibility to address **new distributions** at test time exploiting both the domain classifiers and the connections of the graph.

debora.caldarola@polito.it