Federated Dataset Dictionary Learning For Multi-Source Domain Adaptation

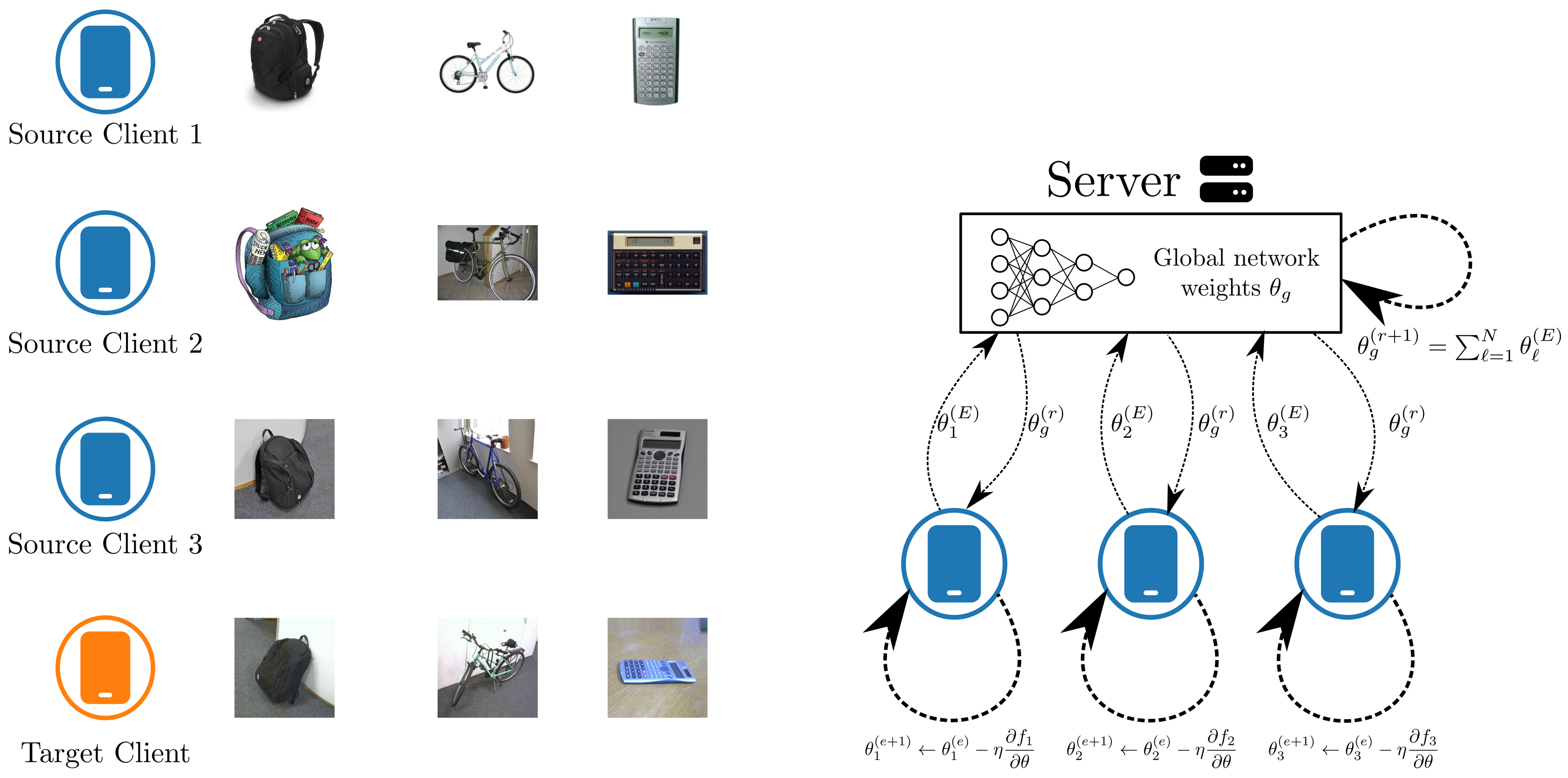
Espinoza Castellon F   Fernandes Montesuma E   Ngolè Mboula F
Mayoue A   Souloumiac A   Gouy-Pailler C

Université Paris-Saclay, CEA, List, F-91120 Palaiseau France

Abstract

In this article, we propose an approach for federated domain adaptation, a setting where distributional shift exists among clients and some have unlabeled data. The proposed framework, FedDaDiL, tackles the resulting challenge through dictionary learning of empirical distributions. In our setting, clients' distributions represent particular domains, and FedDaDiL collectively trains a federated dictionary of empirical distributions. In particular, we build upon the Dataset Dictionary Learning framework by designing collaborative communication protocols and aggregation operations. The chosen protocols keep clients' data private, thus enhancing overall privacy compared to its centralized counterpart. We empirically demonstrate that our approach successfully generates labeled data on the target domain with extensive experiments on (i) Caltech-Office, (ii) TEP, and (iii) CWRU benchmarks. Furthermore, we compare our method to its centralized counterpart and other benchmarks in federated domain adaptation.

Federated Domain Adaptation [1]



Non-i.i.d. data: Each client holds data that follows different **feature distributions**.

Challenge: how to adapt towards a target client **with unlabeled data**.

Optimal Transport [2]

Optimal Transport is a mathematical theory concerned with displacement of mass at least effort. Mathematically, it is expressed as a linear program,

$$\pi^\star = \underset{\pi \in \Pi(P,Q)}{\text{argmin}} \langle \pi, \mathbf{C} \rangle_F := \sum_{i=1}^{n}\sum_{j=1}^{m} \pi_{ij} C_{ij}.$$

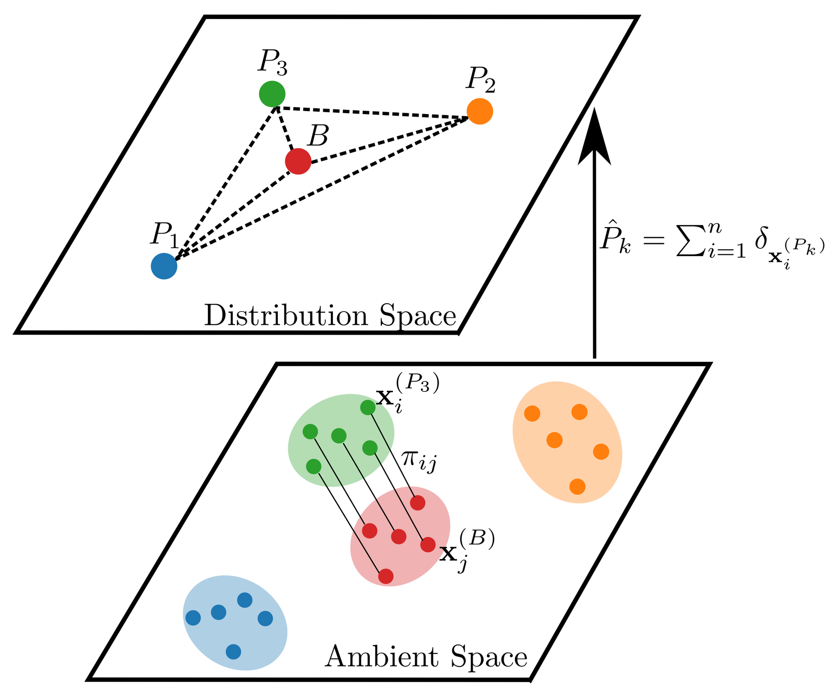
This problem defines a **distance** between probability distributions called **Wasserstein Distance**,

$$W_c(\hat{P}, \hat{Q}) = \sum_{i=1}^{n}\sum_{j=1}^{m} \pi^\star_{ij} C_{ij}.$$

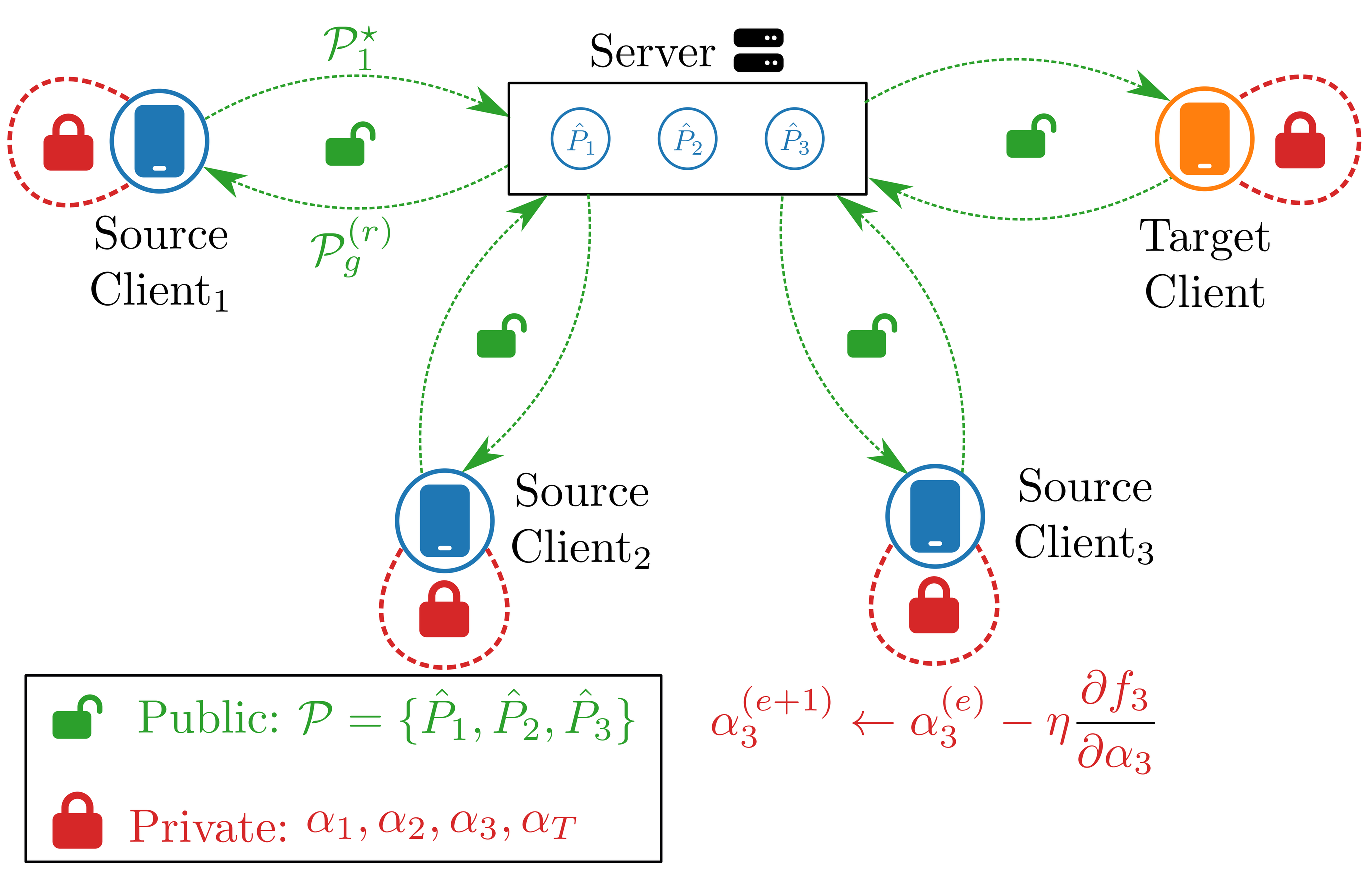
Based on OT one may define a notion of barycenter for distributions,

$$B^\star = \mathcal{B}(\alpha; \mathcal{P}) = \inf_{B} \sum_{k=1}^{K} \alpha_k W_c(P_k, B).$$

where α are called **barycentric coordinates**.



Federated Dataset Dictionary Learning [3]



Algorithm 1 FedDaDiL. The N Clients are indexed by ℓ. n_b is the batch size.

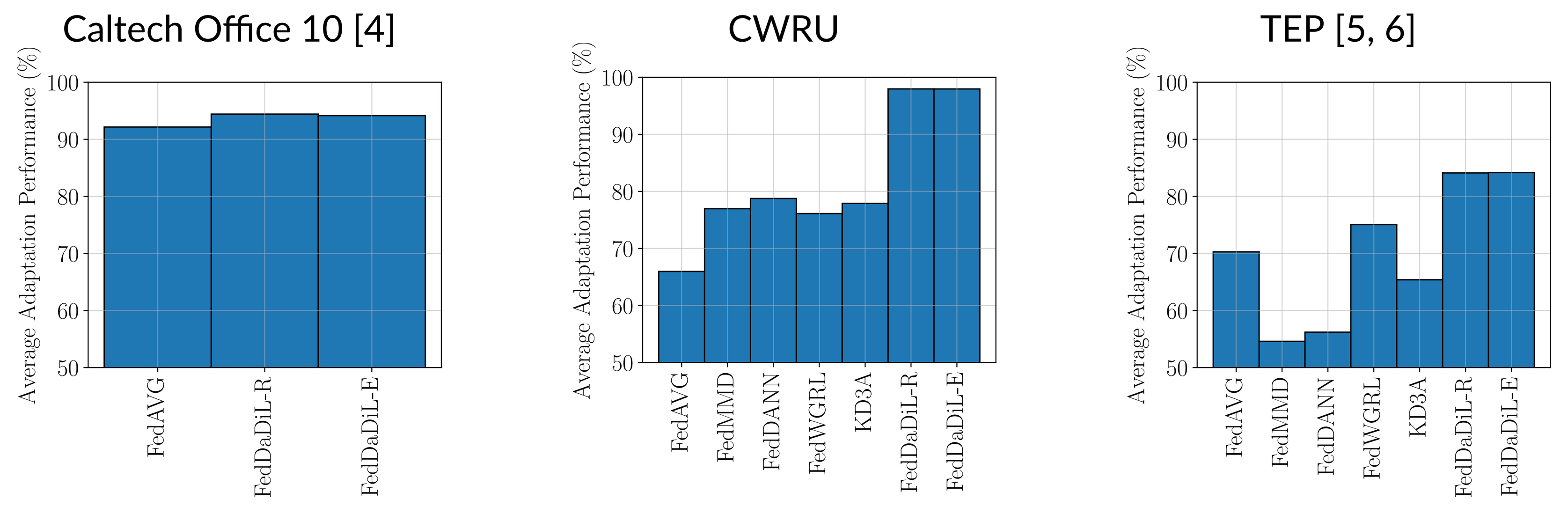
Server initializes $\mathcal{P}_g^{(0)} = \{\hat{P}_k^{(0)}\}_{k=1}^K$
client$_\ell$ initializes $\alpha_\ell^{(0)} \in \Delta_K, \forall \ell = 1 \cdots, N$
for each round $r = 1 \cdots, R$ **do**
  Sample clients $\mathcal{C} \subset \{1, \cdots, N\}$
  Communicate $\mathcal{P}_g^{(r)}, \forall \ell \in \mathcal{C}$
  for client $\ell \in \mathcal{C}$ **do**
    Initialize local dictionary $\mathcal{P}_\ell^{(0)} \leftarrow \mathcal{P}_g^{(r)}$
    $\mathcal{P}_\ell^{(r)} =$ ClientUpdate$(\mathcal{P}_\ell^{(0)}, \alpha_\ell^{(r)})$
    client$_\ell$ sends $\mathcal{P}_\ell^{(r)}$ to server.
  end for
  $\mathcal{P}_g^{(r+1)} \leftarrow$ ServerAggregate$(\{\mathcal{P}_\ell^{(r)}\}_{\ell \in \mathcal{C}})$
end for
Output: Dictionary $\mathcal{P}^\star$ and weights $\mathcal{A}^\star$.

Algorithm 2 ClientUpdate.

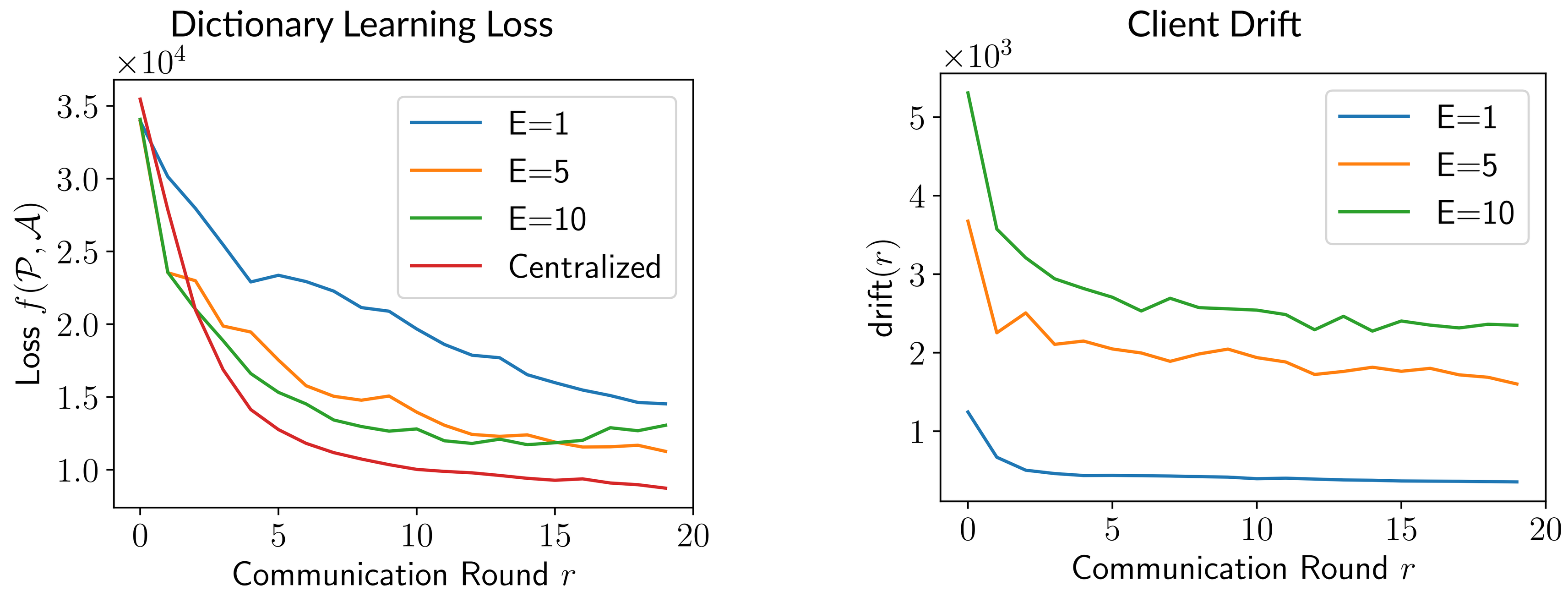
Input: Local atom $\mathcal{P}$. Set of weights $\alpha_\ell \in \Delta_K$. E is the number of local epochs. Learning rate η.
for local epoch $e = 1, \cdots, E$ **do**
  for batch $b = 1, \cdots, B$ **do**
    Compute $f_\ell(\alpha_\ell; \mathcal{P})$
    $\mathbf{x}_i^{(P_k)} \leftarrow \mathbf{x}_i^{(P_k)} - \eta \partial f_\ell / \partial \mathbf{x}_i^{(P_k)}(\alpha_\ell, \mathcal{P})$
    $\mathbf{y}_i^{(P_k)} \leftarrow \mathbf{y}_i^{(P_k)} - \eta \partial f_\ell / \partial \mathbf{y}_i^{(P_k)}(\alpha_\ell, \mathcal{P})$
    $\alpha_\ell \leftarrow \text{proj}_{\Delta_K}(\alpha_\ell - \eta \partial f_\ell / \partial \alpha_\ell(\alpha_\ell, \mathcal{P}))$
  end for
end for
Client sets $\alpha_\ell^{(r+1)} \leftarrow \alpha_\ell^\star$.
Output: $\mathcal{P}_\ell^\star$.

Empirical Results

Comparison with State-of-the-Art



Federated Dictionary Learning



- FedDaDiL **converges**
- Convergence speeds up with local iterations E.

- As training evolves, **dictionary versions** become closer.
- Dictionary version diversity increases with E.

Conclusion

We propose a new **federated domain adaptation** method,

- Domain adaptation is done through **dictionary learning** on clients' data distributions
- Server atoms are **public** while clients' barycentric coordinates are **private**
- We improve performance over previous federated algorithms.

Future Works

- Formalize the privacy properties of FedDaDiL under differentially private optimal transport [7, 8].
- Peer-to-peer federated domain adaptation.
- Consider other kinds of distributional shift (label or concept shift).

References

[1] Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, and Blaise Aguera y Arcas, "Communication-efficient learning of deep networks from decentralized data," in Artificial intelligence and statistics. PMLR, 2017, pp. 1273–1282.

[2] Eduardo Fernandes Montesuma, Fred Ngolè Mboula, and Antoine Souloumiac, "Recent advances in optimal transport for machine learning," arXiv preprint arXiv:2306.16156, 2023.

[3] Eduardo Fernandes Montesuma, Fred Mboula, and Antoine Souloumiac. Multi-source domain adaptation through dataset dictionary learning in wasserstein space. In 26th European Conference on Artificial Intelligence, pages 1739–1745, 2023.

[4] Boqing Gong, Yuan Shi, Fei Sha, and Kristen Grauman, "Geodesic flow kernel for unsupervised domain adaptation," in 2012 IEEE conference on computer vision and pattern recognition. IEEE, 2012, pp. 2066–2073.

[5] Reinartz, C., Kulahci, M., Ravn, O., 2021. An extended tennessee eastman simulation dataset for fault-detection and decision support systems. Computers & Chemical Engineering 149

[6] Eduardo Fernandes Montesuma, Michela Mulas, Fred Ngolè Mboula, Francesco Corona, and Antoine Souloumiac, "Multi-source domain adaptation for cross-domain fault diagnosis of chemical processes," arXiv preprint arXiv:2308.11247, 2023

[7] Dwork, C. (2006, July). Differential privacy. In International colloquium on automata, languages, and programming (pp. 1-12). Berlin, Heidelberg: Springer Berlin Heidelberg.

[8] Lê Tien, N., Habrard, A., & Sebban, M. (2019, August). Differentially Private Optimal Transport: Application to Domain Adaptation. In IJCAI (pp. 2852-2858).

Acknowledgements

We thank **G-Research** for the travel grant that covered part of our participation to ICASSP 2024.

Learn more about DaDiL!

Main Paper (arXiv)   Fabiola   Eduardo