FedRF-Adapt: Robust and Communication-Efficient Federated Domain Adaptation via Random Features

Workshop on Timely and Private Machine Learning over Networks, ICASSP 2024

Yuanjie Wang, Zhanbo Feng, **Zhenyu Liao**

School of Electronic Information and Communications
Huazhong University of Science and Technology, Wuhan, China

April 14, 2024

Motivation

- ML (foundation) models are **giant**, and needs be trained in a distributed/decentralized manner
- data on each client can be **non-i.i.d.**, leading to domain shift and poor generalization
- federated domain adaptation (**FDA**) is great, but is generally **computational/communicational inefficient** in minimizing, e.g., the Maximum Mean Discrepancy (MMD) distance:

$$\mathrm{MMD}(\mathbf{X}_S, \mathbf{X}_T) = \left\| \frac{1}{n_S} \sum_{i=1}^{n_S} \phi(\mathbf{x}_i) - \frac{1}{n_T} \sum_{j=1}^{n_T} \phi(\mathbf{x}_j) \right\|_{\mathcal{H}}^2 = \boldsymbol{\ell}^\mathsf{T} \mathbf{K} \boldsymbol{\ell}, \tag{1}$$

  between source $\mathbf{X}_S$ and target dataset $\mathbf{X}_T$, with "label" vector $\ell_i = \frac{1}{n_S} 1_{\mathbf{x}_i \in \mathbf{X}_S} - \frac{1}{n_T} 1_{\mathbf{x}_i \in \mathbf{X}_T}$, on some RKHS $\mathcal{H}$ via the kernel trick $\langle \phi(\mathbf{x}_i), \phi(\mathbf{x}_j) \rangle_{\mathcal{H}} = K(\mathbf{x}_i, \mathbf{x}_j)$ [SS18]
- with $\mathbf{K} \in \mathbb{R}^{n \times n}$, computational/communicational cost of MMD-based FDA **inevitably** grows, at least **quadratically**, with $n = n_S + n_T$

Main take-away

With randomness, performance MMD-based FDA within $\boxed{N \sim \log(n)}$ communication cost!

Our approach: random features-based MMD

Random Fourier features (RFFs), [RR08]

For data $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_n] \in \mathbb{R}^{p \times n}$ of size n, RFF matrix $\boldsymbol{\Sigma} = \frac{1}{\sqrt{N}} \begin{bmatrix} \cos(\boldsymbol{\Omega}\mathbf{X}) \\ \sin(\boldsymbol{\Omega}\mathbf{X}) \end{bmatrix} \in \mathbb{R}^{2N \times n}$ of $\mathbf{X}$, with N the number of random features, Gaussian random matrix $\boldsymbol{\Omega} \in \mathbb{R}^{N \times p}$.

RFFs approximation of Gaussian kernels, [Tro15, Section 6.5]

For random Fourier features $\boldsymbol{\Sigma} \in \mathbb{R}^{2N \times n}$ of data $\mathbf{X} \in \mathbb{R}^{p \times n}$, there exists $C > 0$ independent of N and n that $\mathbb{E}\|\boldsymbol{\Sigma}^\mathsf{T}\boldsymbol{\Sigma} - \mathbf{K}\|_2 \leq C(\sqrt{\frac{n\log(n)}{N}}\|\mathbf{K}\|_2 + \frac{n\log(n)}{N})$, with $\mathbf{K}$ Gaussian kernel matrix of $\mathbf{X}$.

RFFs approximation of MMD distance

Let $\boxed{\mathsf{RF\text{-}MMD}(\mathbf{X}_S, \mathbf{X}_T) = \boldsymbol{\ell}^\mathsf{T}\boldsymbol{\Sigma}^\mathsf{T}\boldsymbol{\Sigma}\boldsymbol{\ell} = \|\boldsymbol{\Sigma}\boldsymbol{\ell}\|_2^2}$, then, for MMD distance defined in (1) with $n_S, n_T = \Theta(n)$,

$$\mathbb{E}[|\mathsf{RF\text{-}MMD}(\mathbf{X}_S, \mathbf{X}_T) - \mathsf{MMD}(\mathbf{X}_S, \mathbf{X}_T)|] \leq \varepsilon, \tag{2}$$

holds for $\boxed{N \geq C\log(n)/(\dim(\mathbf{K})\varepsilon^2)}$, with $\dim(\mathbf{K}) \equiv \operatorname{tr}\mathbf{K}/\|\mathbf{K}\|_2$ the intrinsic dimension of $\mathbf{K}$.

Consequence of RF-MDD in FDA: FedRF-Adapt

RFFs approximation of MMD distance

Let $\boxed{\text{RF-MMD}(\mathbf{X}_S, \mathbf{X}_T) = \boldsymbol{\ell}^\mathsf{T} \boldsymbol{\Sigma}^\mathsf{T} \boldsymbol{\Sigma} \boldsymbol{\ell} = \|\boldsymbol{\Sigma}\boldsymbol{\ell}\|_2^2}$, then, for MMD distance defined in (1) with $n_s, n_T = \Theta(n)$,

$$\mathbb{E}[|\text{RF-MMD}(\mathbf{X}_S, \mathbf{X}_T) - \text{MMD}(\mathbf{X}_S, \mathbf{X}_T)|] \leq \varepsilon, \tag{3}$$

holds for $\boxed{N \geq C \log(n) / (\dim(\mathbf{K}) \varepsilon^2)}$, with $\dim(\mathbf{K}) \equiv \operatorname{tr} \mathbf{K} / \|\mathbf{K}\|_2$ the intrinsic dimension of $\mathbf{K}$.

- $\boldsymbol{\Sigma}\boldsymbol{\ell} \in \mathbb{R}^{2N}$ is **small** with $N \sim \log(n)$;
- in multi-source FDA, exchange only **highly compressed** and **randomized** messages $\boldsymbol{\Sigma}\boldsymbol{\ell}\mathbb{R}^{2N}$!
- $\Rightarrow$ **FedRF-Adapt**: communication-efficient and robust (to network condition) FDA with added privacy

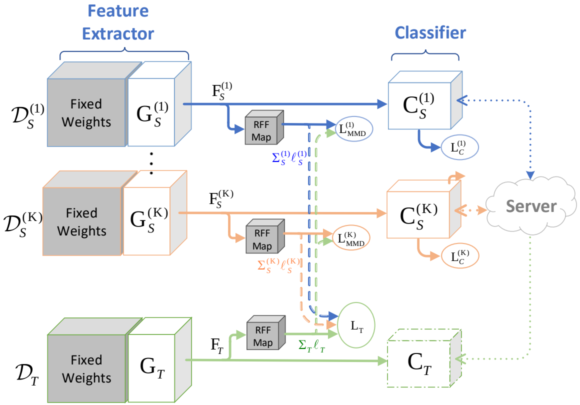


Figure: Illustration of the proposed FedRF-Adapt protocol

1. **local domain alignment** with RF-MMD by exchanging $\boldsymbol{\Sigma}\boldsymbol{\ell} \in \mathbb{R}^{2N}$
2. **global classifier aggregation** via FedAvg [McM+17]

Numerical results

Table: Classification accuracy (%) on Office-Caltech10 and Digit-Five. Best performance shown in **boldface**.

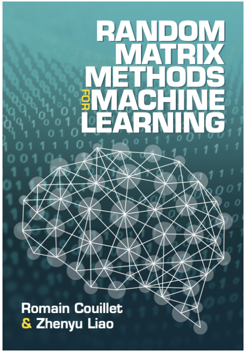
Methods	C,D,W→A	A,D,W→C	A,C,W→D	A,C,D→W	Avg
ResNet101 [He+16]	81.9	87.9	85.7	86.9	85.6
AdaBN [Li+16]	82.2	88.2	85.9	87.4	85.7
AutoDIAL [Mar+17]	83.3	87.7	85.6	87.1	85.9
f-DAN [Lon+15]	82.7	88.1	86.5	86.5	85.9
f-DANN [GL15]	83.5	88.5	85.9	87.1	86.3
FADA [Pen+19]	84.2	88.7	87.1	88.1	87.1
FedRF-Adapt (I)	92.6	85.3	97.6	**97.0**	93.1
FedRF-Adapt (II)	93.4	84.8	97.7	96.9	93.2
FedRF-Adapt (III)	92.7	82.8	96.5	96.2	92.1
FedRF-TCA [Fen+23] (III)	**94.5**	**98.6**	**98.8**	90.0	**95.5**

Methods	→mt	→mm	→up	→sv	→sy	Avg
f-DAN [Lon+15]	86.4	57.5	90.8	45.3	58.4	67.7
f-DANN [GL15]	86.1	59.5	89.7	44.3	53.4	66.6
FADA [Pen+19]	91.4	62.5	91.7	**50.5**	**71.8**	73.6
FedRF-Adapt (III)	**98.5**	75.5	**95.7**	46.0	50.4	73.2
FedRF-TCA [Fen+23] (III)	97.4	64.3	89.5	41.9	44.4	67.5

(a) Digit-Five
(b) Office-Caltech10

- (I): all clients aggregate the classifier in each communication round;
- (II): only a random subset $\mathcal{S}_t$ of source clients are involved;
- (III): as for (II) with classifier aggregation interval $T_C = 100$
- check our paper for more numerical results!

Randomness for ML and data science



- book "*Random Matrix Methods for Machine Learning*"
- by Romain Couillet and **Zhenyu Liao**
- Cambridge University Press, 2022
- a pre-production version of the book and exercise solutions at `https://zhenyu-liao.github.io/book/`
- MATLAB and Python codes to reproduce all figures at `https://github.com/Zhenyu-LIAO/RMT4ML`

References:

- Ali Rahimi and Benjamin Recht. "Random Features for Large-Scale Kernel Machines". In: *Advances in Neural Information Processing Systems*. Vol. 20. NIPS'08. Curran Associates, Inc., 2008, pp. 1177–1184
- Zhanbo Feng et al. "Robust and Communication-Efficient Federated Domain Adaptation via Random Features". In: *arXiv preprint arXiv:2311.04686* (2023)

Thank you!

Thank you! Q & A?