FedSR: A Simple and Effective Domain Generalization Method for Federated Learning


A. Tuan Nguyen


Philip Torr


Ser-Nam Lim

UNIVERSITY OF OXFORD

Introduction

- **Federated Learning** (FL) refers to the decentralized and privacy-preserving machine learning framework.
- There is often **distribution shift** among the clients' data.
- However, FL techniques often only focus on performance on the source domains/clients, **not** how the model **generalize to an unseen domain** under some distribution shifts.
- For example, if K clinical institutions in the US and UK collaborate to train a model with their decentralized data, the goal for the model is not only to perform well on their data distribution, but also to generalize to unseen target data (e.g., from a different country).

Introduction

- In this paper, we incorporate the **Domain Generalization** (DG) problem into the FL setting to tackle this generalization issue.
- A common and successful method for DG is **representation alignment**. However, existing works require sharing and comparing data among domains, which is **not allowed in FL**.
- We propose approaches for **implicit alignment**, that completely respect the the privacy aspect of FL.
- In particular, we propose to learn a **simple representation** of the data, with a L2-norm regularizer and a conditional mutual information regularizer.
- We also show that these regularizers help to implicitly aligns the representation.

Problem Setting

- Representation learning framework:
  - Representation mapping: $p(z|x)$
  - Classifier: $\hat{p}(y|z)$
- Predictive distribution: $\mathbb{E}_{p(z|x)}[\hat{p}(y|x)]$
- Loss per datapoint (x, y): $-\log \mathbb{E}_{p(z|x)}[\hat{p}(y|x)]$
- Local loss function of a client/domain i with data distribution $p_i(x, y)$

$$\mathbb{E}_{p_i(x,y)}\left[-\log \mathbb{E}_{p(z|x)}[\hat{p}(y|x)]\right]$$

- Global loss over all client:

$$\frac{1}{n\_clients}\sum_i \mathbb{E}_{p_i(x,y)}\left[-\log \mathbb{E}_{p(z|x)}[\hat{p}(y|x)]\right]$$

Approach

- The conventional loss function of FL (previous slide) only focus on performance on the source clients i's.
- To learn a generalizable representation, we propose to use common regularization techniques to restrict the complexity of the representation, hoping that it would learn essential information and ignore spurious correlation.
- We also show both **theoretically** and **empirically** that these regularizers leads to better marginal and conditional representation alignment.

Approach: L2-norm Regularizer

- We regularize the l2-norm of the representation:

$$\ell_i^{L2R} = \mathbb{E}_{p_i(x)}\left[\mathbb{E}_{p(z|x)}\left[||z||_2^2\right]\right]$$

- Connection to **marginal alignment** of the representation (details in our paper).

Approach: Conditional Mutual Information

- We minimize a tractable upper bound of the conditional mutual information $I_i(x, z|y)$:

$$\ell_i^{CMI} = \mathbb{E}_{p_i(x,y)}\big[KL[p(z|x)|r(z|y)]\big]$$

With $r(z|y)$ being a learnable variational distribution.

- Connection to **conditional alignment** of the representation (details in our paper).

Results

Quantitative:

Table 2: **PACS**. Reported numbers are from 3 runs

Models		Backbone	PACS A	PACS C	PACS P	PACS S	PACS Average
Centralized Methods	DGER [47]	Resnet18	80.70	76.40	96.65	71.77	81.38
	DIRT-GAN [31]	Resnet18	82.56	76.37	95.65	79.89	**83.62**
FL Methods	FedAVG [28]	Resnet18	77.8±0.5	72.8±0.4	91.9±0.5	78.8±0.3	80.3
	FedADG [45]	Resnet18	77.8±0.5	74.7±0.4	92.9±0.3	79.5±0.4	81.2
	FedCMI (ours)	Resnet18	80.8±0.4	73.7±0.2	92.8±0.5	79.5±0.2	81.7
	FedL2R (ours)	Resnet18	82.2±0.4	75.8±0.3	92.8±0.4	81.6±0.1	83.1
	FedSR (ours)	Resnet18	83.2±0.3	76.0±0.3	93.8±0.5	81.9±0.2	**83.7**

Results

Quantitative:

Table 3: **OfficeHome**. Reported numbers are from 3 runs

Models		Backbone	OfficeHome A	OfficeHome C	OfficeHome P	OfficeHome R	OfficeHome Average
Centralized Methods	Mixup [44]	Resnet50	64.7	54.7	77.3	79.2	**69.0**
	CORAL [38]	Resnet50	64.4	55.3	76.7	77.9	68.6
FL Methods	FedAVG [28]	Resnet50	62.2±0.9	55.6±0.9	75.7±0.2	78.2±0.2	67.9
	FedADG [45]	Resnet50	63.2±0.9	57.0±0.2	76.0±0.1	77.7±0.5	68.4
	FedCMI (ours)	Resnet50	61.8±0.5	55.5±0.9	76.3±0.1	77.4±0.1	67.8
	FedL2R (ours)	Resnet50	64.5±0.3	56.5±0.5	76.1±0.2	77.9±0.2	68.8
	FedSR (ours)	Resnet50	65.4±0.5	57.4±0.2	76.2±0.6	78.3±0.3	**69.3**

Results

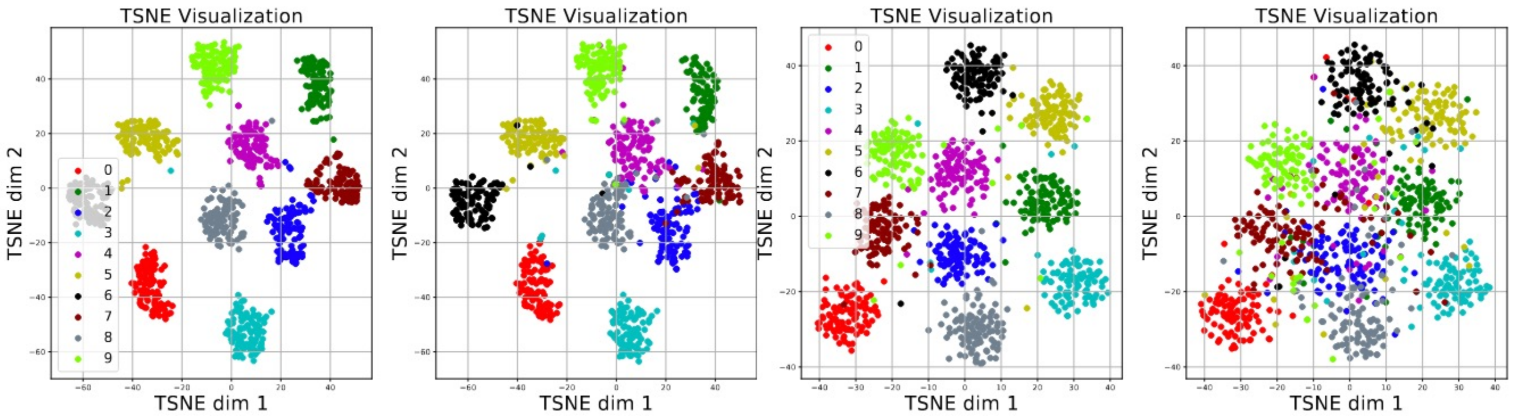
Quantitative:

Table 4: **DomainNet**. Reported numbers are from 3 runs

Models		Backbone	DomainNet C	I	P	Q	R	S	AVG
Centralized Methods	MLDG [21]	Resnet50	59.5	19.8	48.3	13.0	59.5	50.4	**41.8**
	CORAL [38]	Resnet50	58.7	20.9	47.3	13.6	60.2	50.2	**41.8**
FL Methods	FedAVG [28]	Resnet50	59.3±0.7	16.5±0.9	44.2±0.7	10.8±1.8	57.2±0.8	49.8±0.4	39.6
	FedADG [45]	Resnet50	60.9±0.6	17.2±0.2	44.3±0.2	12.4±0.2	57.6±0.9	50.3±0.8	40.4
	FedCMI (ours)	Resnet50	59.0±0.9	18.0±0.7	44.6±0.5	12.2±0.4	56.2±0.2	50.0±0.4	40.0
	FedL2R (ours)	Resnet50	60.2±0.6	18.1±0.4	44.9±0.6	11.0±0.9	57.8±0.4	51.5±0.7	40.6
	FedSR (ours)	Resnet50	61.0±0.6	18.6±0.4	45.2±0.5	13.4±0.6	57.6±0.2	51.8±0.3	**41.3**

Results

Qualitative: Our method leads to better alignment of the representation.



(a) FedSR (b) FedAVG

Figure 1: Visualization using t-SNE of the representation space of our method FedSR and the baselines FedAVG. For each method, the left subfigure corresponds to one source domain $\mathcal{M}_{15}$ and the right one corresponds to the target domain $\mathcal{M}_0$. Each color represents a digit class.