Personalized Federated Learning towards Communication Efficiency, Robustness and Fairness

Shiyun Lin[1,2], Yuze Han[1], Xiang Li[1], Zhihua Zhang[1,2]

[1] School of Mathematical Sciences, Peking University
[2] Center for Statistical Science, Peking University

October 21, 2022

Introduction

- Personalized federated learning faces many challenges:
  - Expensive communication costs
  - Training-time adversarial attacks
  - Performance unfairness across devices

Introduction

- Personalized federated learning faces many challenges:
  - Expensive communication costs
  - Training-time adversarial attacks
  - Performance unfairness across devices

> Can we balance different constraints of interest (i.e., communication efficiency, robustness and fairness) simultaneously?

Introduction

- Personalized federated learning faces many challenges:
  - Expensive communication costs
  - Training-time adversarial attacks
  - Performance unfairness across devices

> Can we balance different constraints of interest (i.e., communication efficiency, robustness and fairness) simultaneously?

- The answer is YES!

Introduction

- Personalized federated learning faces many challenges:
  - Expensive communication costs
  - Training-time adversarial attacks
  - Performance unfairness across devices

> Can we balance different constraints of interest (i.e., communication efficiency, robustness and fairness) simultaneously?

- The answer is YES!
- We propose a personalized FL method named as `lp-proj` based on L^p-regularization and low-dimensional random projection. Multiple benefits of the proposed objective are explored from both theoretical and empirical perspectives.

Infimal Convolution

- Conventional federated learning:

$$\min_{\mathbf{w}\in\mathbb{R}^d} F(\mathbf{w}) := G\left\{F_1(\mathbf{w}), \cdots, F_N(\mathbf{w})\right\}. \tag{1}$$

Infimal Convolution

- Conventional federated learning:

$$\min_{\mathbf{w}\in\mathbb{R}^d} F(\mathbf{w}) := G\left\{F_1(\mathbf{w}), \cdots, F_N(\mathbf{w})\right\}. \tag{1}$$

- We introduce a local model $\mathbf{x}_k$ for each client k to fit the local data.

Infimal Convolution

- Conventional federated learning:

$$\min_{\mathbf{w}\in\mathbb{R}^d} F(\mathbf{w}) := G\left\{F_1(\mathbf{w}), \cdots, F_N(\mathbf{w})\right\}. \tag{1}$$

- We introduce a local model $\mathbf{x}_k$ for each client k to fit the local data.
- Infimal Convolution: bridges local models and global model.

$$\begin{aligned} F_k(\mathbf{w}) &= \{f_k \otimes \lambda g\}(\mathbf{w}) := \min_{\mathbf{x}_k\in\mathbb{R}^d} f_k(\mathbf{x}_k) + \lambda g(\mathbf{w} - \mathbf{x}_k), \\ f_k(\mathbf{x}_k) &= \mathbb{E}_{\xi_k}\left[\tilde{f}_k(\mathbf{x}_k; \xi_k)\right]. \end{aligned} \tag{2}$$

- g is the smoothing kernel, which is designed to characterize the relationship between local models and global model.

Subspace Regularization

- Concerning that communication cost is critical in the application of FL.

Subspace Regularization

- Concerning that communication cost is critical in the application of FL.
- High-dimensional data usually has low-dimensional representation.

Subspace Regularization

- Concerning that communication cost is critical in the application of FL.
- High-dimensional data usually has low-dimensional representation.
- Random projection would preserve similarity of data vectors.

Subspace Regularization

- Concerning that communication cost is critical in the application of FL.
- High-dimensional data usually has low-dimensional representation.
- Random projection would preserve similarity of data vectors.
- We propose to **regularize the projection of local models in a shared low-dimensional space**.

$$g(\mathbf{w}-\mathbf{x}_k)=\frac{1}{p}\left\|\boldsymbol{P}(\mathbf{w}-\mathbf{x}_k)\right\|_p^p=\frac{1}{p}\left\|\tilde{\mathbf{w}}-\boldsymbol{P}\mathbf{x}_k\right\|. \tag{3}$$

  - $\boldsymbol{P}$ is a $d_{\text{sub}} \times d$ random matrix generated initially and fixed during training.
  - $d_{\text{sub}} \ll d$ is the dimension of the shared-and-fixed random subspace.

Subspace Regularization

- Concerning that communication cost is critical in the application of FL.
- High-dimensional data usually has low-dimensional representation.
- Random projection would preserve similarity of data vectors.
- We propose to **regularize the projection of local models in a shared low-dimensional space**.

$$g(\mathbf{w}-\mathbf{x}_k)=\frac{1}{p}\left\|\boldsymbol{P}(\mathbf{w}-\mathbf{x}_k)\right\|_p^p=\frac{1}{p}\left\|\tilde{\mathbf{w}}-\boldsymbol{P}\mathbf{x}_k\right\|. \tag{3}$$

  - $\boldsymbol{P}$ is a $d_{\mathrm{sub}}\times d$ random matrix generated initially and fixed during training.
  - $d_{\mathrm{sub}}\ll d$ is the dimension of the shared-and-fixed random subspace.

- Combining (1), (2) and (3), our personalized FL method is formulated as a bi-level problem. We introduce the algorithm `lp-proj`, which alternatively minimizes the local and global objectives with gradient descent.

Benefits: Communication Efficiency, Robustness and Fairness

(Intuitions are provided here. For formal theoretical analysis, please refer to Section 4 in our paper.)

- **Communication Efficiency**: The global model $\tilde{\mathbf{w}}$ is restricted to lie in a fixed low-dimensional subspace. $\Longrightarrow$ Only $\tilde{\mathbf{w}}$ of dimension d_{sub}, instead of the full model $\mathbf{x}_k$ of dimension d, is communicated each round.

Benefits: Communication Efficiency, Robustness and Fairness

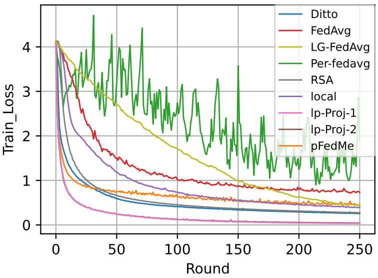
(Intuitions are provided here. For formal theoretical analysis, please refer to Section 4 in our paper.)

- **Communication Efficiency**: The global model $\tilde{\mathbf{w}}$ is restricted to lie in a fixed low-dimensional subspace. $\Longrightarrow$ Only $\tilde{\mathbf{w}}$ of dimension d_{sub}, instead of the full model $\mathbf{x}_k$ of dimension d, is communicated each round.

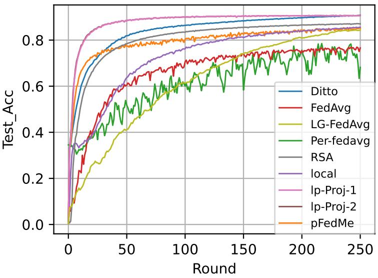
- **Robustness and Fairness**:
  - (Near) consensus of model parameters in the low-dimensional subspace leaves flexibility towards personalization and better generalization to the local data distribution.
  - L^p-norm regularization is equivalent to launching an uncertainty set to the model parameter. $\Longrightarrow$ We can enhance accuracy by searching for a model adaptive to the local data distribution in the uncertainty set.

- **Personalization Accuracy Performance**:

EMNIST, Train Loss



EMNIST, Test Acc



- `lp-proj` has comparable or even superior performance than other SOTA methods. Moreover, the training process is more stable as the loss and accuracy curves have less fluctuation.

Numerical Experiments - Communication Efficiency

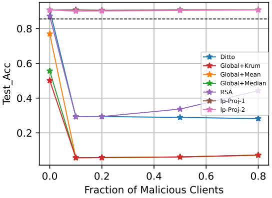
- **Communication Efficiency**:

Method	EMNIST			
	Bytes Budget	Test Acc	Target Acc	Used Bytes
FedAvg	4236900	$\star$	0.7	445851400
Sketch	4236900	$\star$	0.7	$\star$
lp-proj-1	4236900	**0.906**	0.7	**174720**
lp-proj-2	4236900	**0.906**	0.7	196560
LBGM	4236900	$\star$	0.7	769902776
QSGD	4236900	$\star$	0.7	673302175
DGC	4236900	$\star$	0.7	$\star$it
LG-FedAvg	4236900	0.071	0.7	230786010

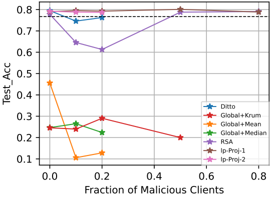
- Given a communication budget of bytes, `lp-proj` obtains $\sim$ 83.5% test accuracy improvement on `EMNIST`.
- Given a target test accuracy, the communication cost is saved by `1320x` on `EMNIST` compared with the best competing method.

- **Robustness**:

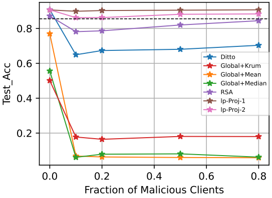
`EMNIST`, same-value



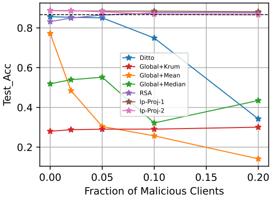
`CIFAR`, sign-flipping



`EMNIST`, Gaussian



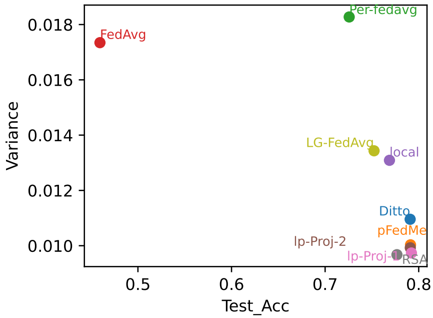
`Synthetic(0,0)`, data-poisoning



- `lp-proj` is resistant to standard malicious attacks.
- Random projection helps alleviate the attacks applied in the original space, while the L^p-norm helps eliminate outliers further.

- **Fairness**:

CIFAR,
accuracy-fairness trade-off



- `lp-proj` provides accurate and fair solutions that are comparable to other SOTA methods.
- On `CIFAR`, `lp-proj-1` achieves the highest test accuracy of 79.22% with the lowest variance of 0.0097 among all the competitors.

Conclusion

- We propose a simple yet powerful personalized FL approach based on infimal convolution and subspace projection.
- We present convergence results for smooth objectives with square regularizers.
- Theoretical analysis and numerical experiments show that our approach could promote communication efficiency, robustness and performance fairness.
- **Code Implementation**: `https://github.com/desternylin/perfed`

Conclusion

- We propose a simple yet powerful personalized FL approach based on infimal convolution and subspace projection.
- We present convergence results for smooth objectives with square regularizers.
- Theoretical analysis and numerical experiments show that our approach could promote communication efficiency, robustness and performance fairness.
- **Code Implementation**: `https://github.com/desternylin/perfed`

Thanks for listening!