SemiFL: Semi-Supervised Federated Learning for Unlabeled Clients with Alternate Training

Enmao Diao[1] Jie Ding[2] Vahid Tarokh[1]

[1]Duke University [2]University of Minnesota-Twin Cities

Overview

We propose a new Federated Learning (FL) framework SemiFL to address the problem of Semi-Supervised Federated Learning (SSFL). We discover that it is challenging to directly combine the state-of-the-art Semi-Supervised Learning (SSL) methods with the communication efficient federated learning methods such as FedAvg to allow local clients to train multiple epochs [1]. The key ingredient that enables SemiFL to allow unlabeled clients to train multiple local epochs is that *we alternate the training of a labeled server and unlabeled clients* to ensure that the quality of pseudo-labeling is maintained during training.

- We propose SemiFL in which clients have completely unlabeled data and can train multiple local epochs to reduce communication costs, while the server has a small amount of labeled data.
- We develop a theoretical analysis on strong data augmentation for SSL methods. We provide a theoretical understanding of the success of data augmentation-based SSL methods to illustrate the bottleneck of a vanilla combination of communication-efficient FL with SSL.
- Our proposed method achieves 30% improvement over the existing SSFL methods and performs competitively with the state-of-the-art FL methods and SSL method.

Motivation

Question: *How a server that hosts a labeled dataset can leverage clients with unlabeled data for a supervised learning task in the Federated Learning setting?*

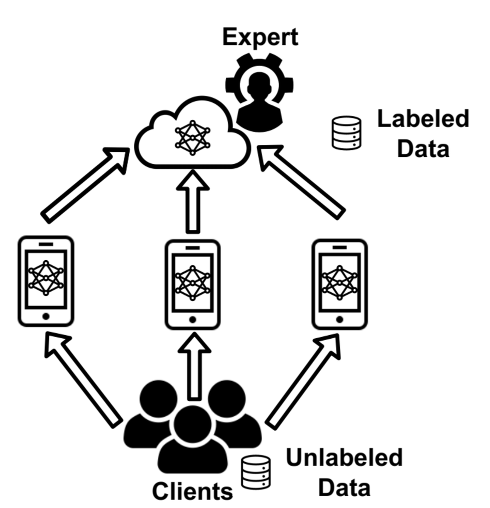

Figure 1. A resourceful server with labeled data can significantly improve its learning performance by working with distributed clients with unlabeled data without data sharing.

Problem: *Existing SSFL methods cannot outperform the case of training with only labeled data [2,3]. It is not straightforward how we can combine the SSL method in a communication-efficient FL scenario where we train multiple local epochs.*

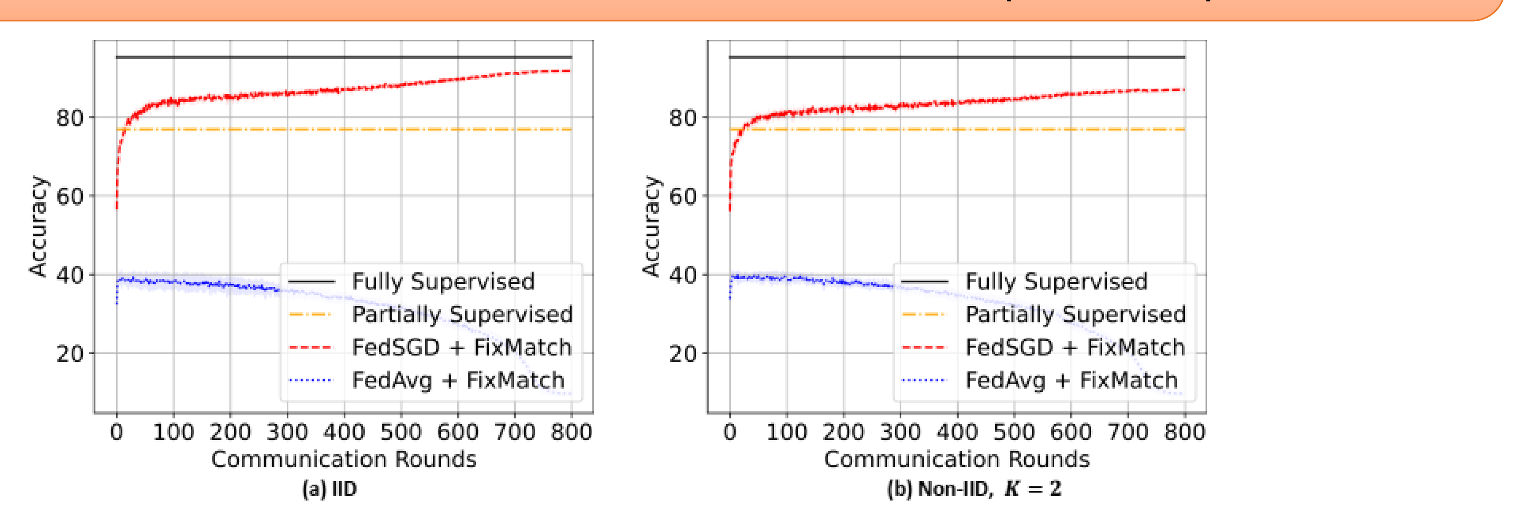

Figure 2. Results of CIFAR10 dataset with (a) IID and (b) Non-IID. The "Fully Supervised" and "Partially Supervised" refer to training a centralized model with full and 4000 labeled data, respectively.

Paper

Code

Method

Theoretical Analysis of Strong Data augmentation for SSL

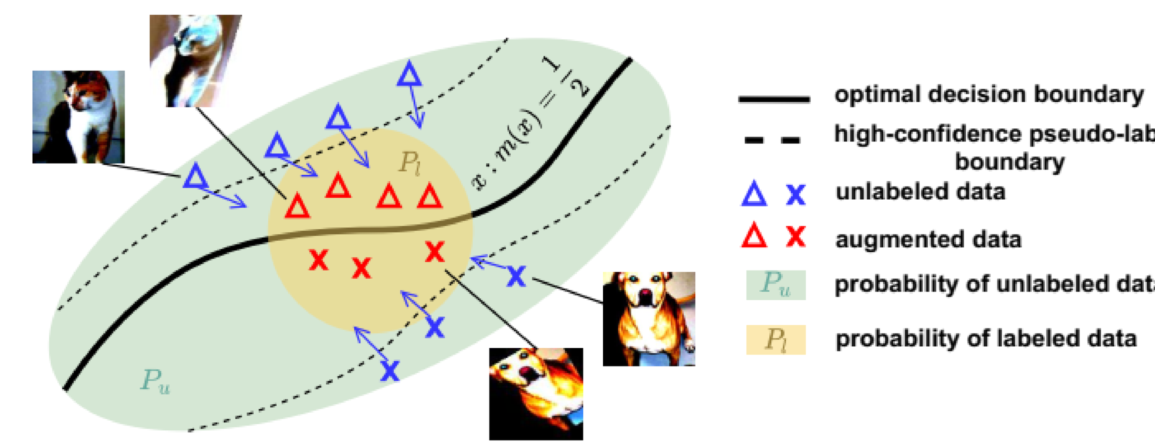

Figure 3. Illustration of the strong data augmentation-based SSL. The above ideas are theoretically formalized.

Alternate Training

- **Fine-tune global model with labeled data** At each round, the server will retrain the global model with the labeled data. In this way, the server can provide a comparable or better model than the previous round for the active clients in the next round to generate pseudo-labels.
- **Generate pseudo-labels with global model** We will label the unlabeled data once the active clients immediately receive the global model from the server. This way, pseudo-labels' quality will not degrade during the local training.

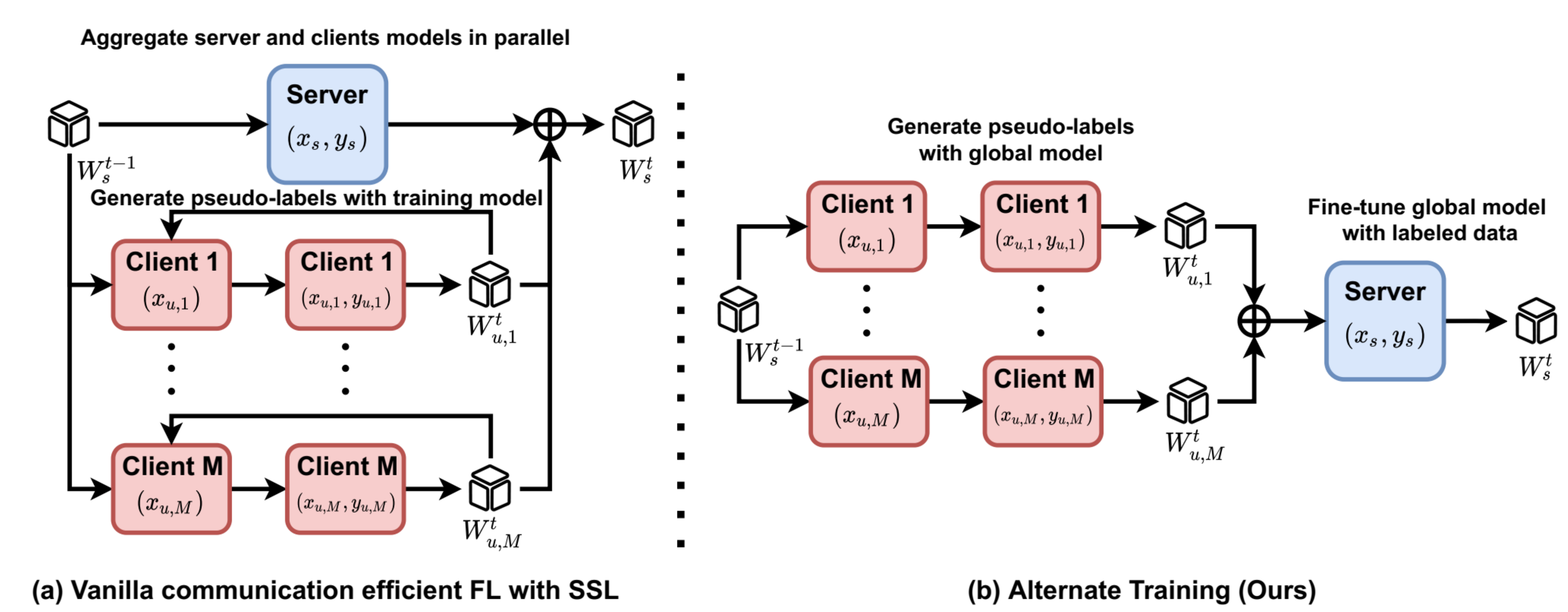

Figure 4. An illustration of (a) vanilla combination of communication efficient FL and SSL, and (b) Alternate Training (Ours). (a) The vanilla combination trains and aggregates server and client models in parallel and generates pseudo-labels with the training models for every batch of unlabeled data. (b) Alternate Training fine-tunes the aggregated global model with labeled data and generates pseudo-labels only once upon receiving the global model from the server.

The SemiFL Algorithm

At each iteration t, the server will first update the model with the standard supervised loss L_s for local epochs E with data batch (x_b, y_b) randomly split from the supervised dataset. Client m defines the "fix" loss L_{fix} [4] and "mix"' loss L_{mix} [5] and performs gradient descent steps.

- **Supervised loss** $\alpha(\cdot)$ represents a weak data augmentation, such as random horizontal flipping and random cropping, that maps one image to another.

$$L_s = \ell(f(\alpha(x_b), W_s), y_b), \quad W_s = W_s - \eta \nabla_W L_s$$

- **Semi-Supervised loss** $\mathcal{A}(\cdot)$ represents a strong data augmentation mapping and $\lambda > 0$ is a hyperparameter set to be one in our experiments. After training for E local epochs, client m transmits $W_{u,m}$ to the server.

$$L_{\text{fix}} = \ell(f(\mathcal{A}(x_b^{\text{fix}}), W_{u,m}), y_b^{\text{fix}}),$$

$$L_{\text{mix}} = \lambda_{\text{mix}} \cdot \ell(f(\alpha(x_{\text{mix}}), W_{u,m}), y_b^{\text{fix}}) + (1 - \lambda_{\text{mix}}) \cdot \ell(f(\alpha(x_{\text{mix}}), W_{u,m}), y_b^{\text{mix}}))$$

$$W_{u,m} = W_{u,m} - \eta \nabla_W (L_{\text{fix}} + \lambda \cdot L_{\text{mix}})$$

Experiments

- **Datasets:** CIFAR10, SVHN and CIFAR100
- Number of clients $M = 100$

Table 1. Comparison of SemiFL with the Baselines, SSL, FL, and SSFL methods. SemiFL improves the performance of the labeled server, SemiFL significantly outperforms the existing SSFL methods, and performs close to the state-of-the-art FL and SSL methods.

Dataset			CIFAR10		SVHN		CIFAR100	
Number of Supervised			250	4000	250	1000	2500	10000
Baseline		Fully Supervised	95.3(0.1)		97.3(0.0)		79.3(0.1)	
		Partially Supervised	42.4(1.8)	76.9(0.2)	77.1(2.9)	90.4(0.5)	27.2(0.7)	59.3(0.1)
SSL		Π-Model [13]	45.7(4.0)	86.0(0.4)	81.0(1.9)	92.5(0.4)	42.8(0.5)	62.1(0.1)
		Pseudo-Labeling [44]	50.2(0.4)	83.9(0.3)	79.8(1.1)	90.1(0.6)	42.6(0.5)	63.8(0.2)
		Mean Teacher [44]	67.7(2.3)	90.8(0.2)	96.4(0.1)	96.6(0.1)	46.1(0.6)	64.2(0.2)
		MixMatch [23]	89.0(0.9)	93.6(0.1)	96.0(0.2)	96.5(0.3)	60.1(0.4)	71.7(0.3)
		UDA [22]	91.2(1.1)	95.1(0.2)	94.3(2.8)	97.5(0.2)	66.9(0.2)	75.5(0.3)
		ReMixMatch [24]	94.6(0.1)	95.3(0.1)	97.1(0.5)	97.4(0.1)	72.6(0.3)	77.0(0.6)
		FixMatch [32]	94.9(0.7)	95.7(0.1)	97.5(0.4)	97.7(0.1)	71.7(0.1)	77.4(0.1)
Non-IID, $K = 2$	FL	HeteroFL [9]	51.5(3.6)		72.3(4.4)		3.1(0.3)	
	SSFL	FedMatch [27]	41.3(1.1)	58.3(1.0)	58.2(3.1)	84.3(1.0)	17.7(0.5)	30.5(0.8)
		FedRGD [28]	32.7(3.6)	48.9(1.4)	21.2(2.2)	21.6(2.3)	13.8(1.4)	26.5(3.0)
		SemiFL	**60.0(0.9)**	**85.3(0.3)**	**87.5(1.1)**	**92.2(0.8)**	**35.2(0.3)**	**62.1(0.4)**
Non-IID, Dir(0.1)	FL	HeteroFL [9]	85.0(0.6)		95.8(0.1)		74.0(0.4)	
	SSFL	FedMatch [27]	41.6(1.0)	58.9(0.7)	58.4(3.4)	84.3(0.6)	17.5(0.5)	30.8(0.6)
		FedRGD [28]	31.5(2.9)	45.2(0.8)	20.0(4.0)	23.8(3.4)	13.4(1.3)	23.6(2.6)
		SemiFL	**63.0(0.6)**	**84.5(0.4)**	**91.2(0.3)**	**93.0(0.5)**	**49.0(1.0)**	**68.0(0.2)**
Non-IID, Dir(0.3)	FL	HeteroFL [9]	91.6(0.1)		96.8(0.0)		76.9(0.1)	
	SSFL	FedMatch [27]	41.2(1.1)	58.4(0.6)	59.1(2.8)	84.0(1.1)	17.8(0.4)	31.1(0.5)
		FedRGD [28]	32.5(3.0)	46.9(1.6)	24.8(5.1)	22.0(3.9)	13.1(2.0)	23.8(1.9)
		SemiFL	**71.9(1.2)**	**88.9(0.3)**	**94.0(0.5)**	**95.2(0.2)**	**54.9(1.4)**	**70.0(0.3)**
IID	FL	HeteroFL [9]	94.3(0.1)		97.5(0.0)		77.8(0.2)	
	SSFL	FedMatch [27]	41.7(1.1)	58.6(0.5)	58.6(3.0)	84.3(0.9)	17.6(0.3)	31.3(1.0)
		FedRGD [28]	33.2(1.9)	47.8(1.7)	21.3(6.5)	20.7(1.1)	13.3(1.4)	23.8(2.6)
		SemiFL	**88.2(0.3)**	**93.1(0.1)**	**96.8(0.3)**	**96.9(0.1)**	**61.3(1.2)**	**72.1(0.2)**

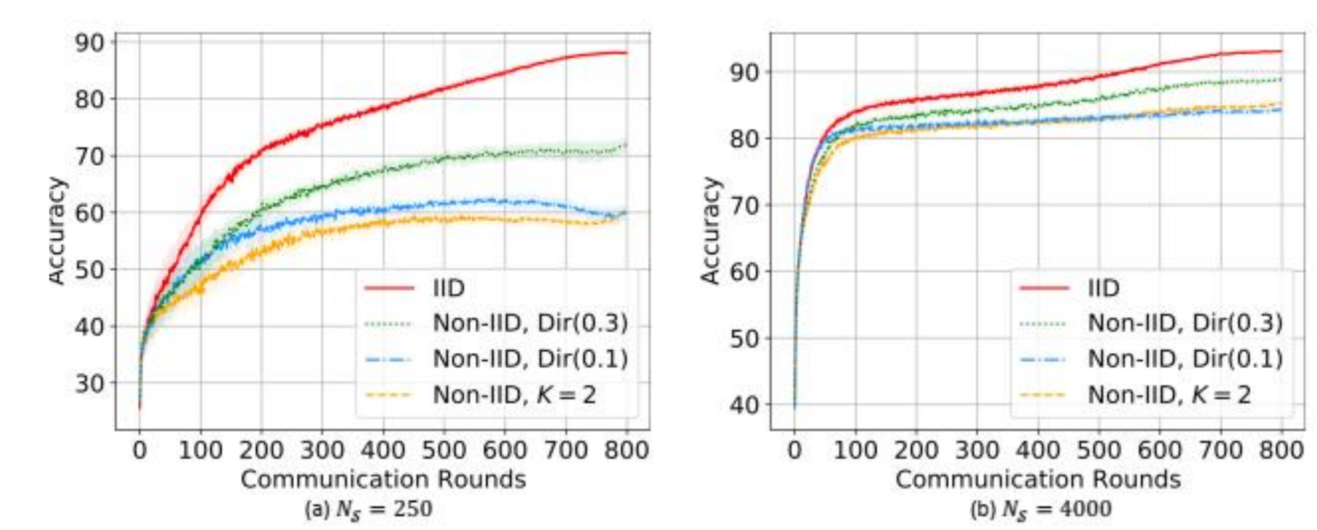

Figure 5. Results of CIFAR10 dataset with (a) $N_{\mathcal{S}} = 250$ and (b) $N_{\mathcal{S}} = 4000$.

Table 2. Ablation study on each component of alternative training with CIFAR10 dataset. The combination of "Fine-tune global model with labeled data" and "Generate pseudo-labels with global model" significantly improves the performance.

Method	Fine-tune global model with labeled data	Generate pseudo-labels with global model	Accuracy	
			Non-IID, $K = 2$	IID
Fully Supervised	N/A		95.33	
Partially Supervised			76.92	
FedAvg + FixMatch	✗	✗	41.01	40.26
	✗	✓	48.89	47.03
SemiFL	✓	✗	80.42	81.70
	✓	✓	**85.34**	**93.10**

References

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

[2] Jeong, Wonyong, et al. " Federated semi-supervised learning with inter-client consistency & disjoint learning. " In *International Conference on Learning Representations*, 2021.

[3] Zhang, Zhengming, et al. "Improving semi-supervised federated learning by reducing the gradient diversity of models." *2021 IEEE International Conference on Big Data (Big Data)*. IEEE, 2021.

[4] Sohn, Kihyuk, et al. "Fixmatch: Simplifying semi-supervised learning with consistency and confidence." *Advances in neural information processing systems* 33 (2020): 596-608.

[5] Berthelot, David, et al. "Remixmatch: Semi-supervised learning with distribution alignment and augmentation anchoring." *arXiv preprint arXiv:1911.09785* (2019).