Resolving the Tug-of-War: A Separation of Communication and Learning in Federated Learning

Junyi Li[1], Heng Huang[1]

[1]University of Maryland College Park

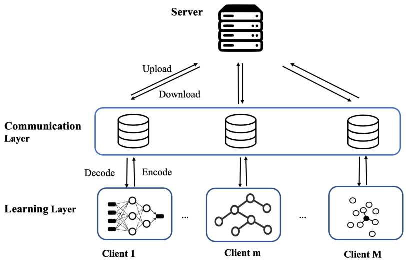
January 15, 2024

The Conflicts between Learning and Communication in FL

- **Communication**: shared parameter space among clients, low-dimensional space to reduce communication cost;
- **Learning**: different parameter space to incorporate system and data heterogeneity; high-dimensional space for better performance;

FedSep: Separating Communication and Learning in FL

The FedSep Framework:



Optimization Objective:

$$\min_{x \in \mathbb{R}^p} h(x) := \frac{1}{M} \sum_{m=1}^{M} h^{(m)}(x) := \frac{1}{M} \sum_{m=1}^{M} f^{(m)}(y_x^{(m)}),$$
$$y_x^{(m)} = \arg\min_{y^{(m)} \in \mathbb{R}^{d^{(m)}}} g^{(m)}(x, y^{(m)}) \quad (1)$$

FedSep Algorithm

Algorithm 1 Separating Communication and Learning in FL (**FedSep**)

for $t = 1$ **to** T **do**
  Randomly sample a subset $\mathcal{M}_t$ of clients;
  for $m \in \mathcal{M}_t$ in parallel **do**
    Decode stage: estimate $y_x^{(m)} = Dec^{(m)}\{x\}$;
    Learning stage: optimize $f^{(m)}(y)$;
    Encode stage: encode the update of the learning layer back to the communication layer;
  end for
  $x_{t+1} = x_t - \frac{1}{|\mathcal{M}_t|} \sum_{m \in \mathcal{M}_t} \eta_g \Delta \hat{x}_t^{(m)}$
end for

Convergence Theorem

Theorem

Suppose we choose the learning rates as $\gamma = \min(\frac{1}{2L}, (\frac{1}{C_\gamma T})^{1/2})$,

$\eta = \min\left(1, \left(\frac{8Ib_x M\bar{L}h(x_1)}{TG_2^2}\right)^{1/2}, \left(\frac{4\bar{L}h(x_1)}{C_\eta I^2 T}\right)^{1/3}\right)$ *and* $\eta_g = \frac{1}{2I\bar{L}}$, *then we have:*

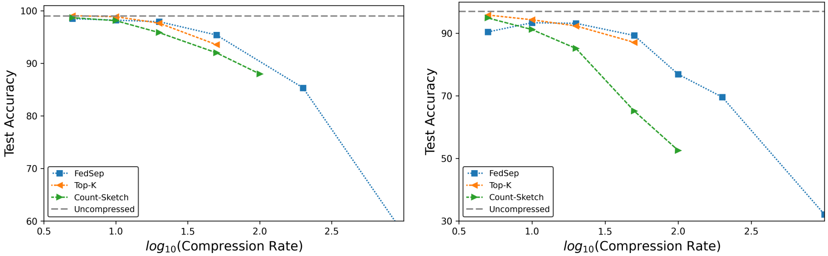
$$\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\big(\|\nabla h(x_t)\|^2 + \frac{1}{2I}\sum_{i=1}^{I}\|\mathbb{E}_\xi[\bar{\Delta}\hat{x}_{t,i}]\|^2\big)$$
$$= O\left(\frac{\kappa^3}{T} + \left(\frac{\kappa^5}{T}\right)^{1/2} + \left(\frac{\kappa^6}{T^2}\right)^{1/3} + \tilde{G}\right)$$

where $\tilde{G} = \kappa^2(1-\tau\mu)^{2(Q+1)} + \kappa^4(1-\mu\gamma)^{I_{dec}}$, C_η *and* C_γ *are some constants.*

Communication-efficient Federated Learning

- **Objective**:

$$\min_{\omega \in \mathbb{R}^p} \frac{1}{M} \sum_{m=1}^{M} \mathcal{L}(\theta_\omega^{(m)}; \mathcal{D}_{tr}^{(m)}) \ s.t.\ \theta_\omega^{(m)} = \arg\min_{\theta \in \mathbb{R}^d} \frac{1}{2} \left\| S^{(m)}\theta - \omega \right\|_2^2 + \beta \left\| \theta \right\|_1$$



- Test Accuracy w.r.t Communication Rate for FedSep and other baseline methods for MNIST Dataset. The left plot shows results under the I.I.D case, and the right plot shows results for the Non-I.I.D case. The local learning steps are set as $I = 5$.

Model-Heterogeneous Federated Learning

- **Objective**:

$$\min_{\omega \in \mathbb{R}^p} \frac{1}{M} \sum_{m=1}^{M} \mathcal{L}(\theta_\omega^{(m)}; \mathcal{D}_{tr}^{(m)})$$

$$s.t.\ \theta_\omega^{(m)} = a_\omega^{(m)} \odot \omega,\ \ a_\omega^{(m)} = \arg\min_{a \in \{0,1\}^p} \mathcal{L}(a \odot \omega; \mathcal{D}_{val}^{(m)}) + \beta \mathcal{R}(T(a), p^{(m)} T_{tol})$$

Experimental Results

Table 1: **Test accuracy comparison between FedSep with other model-heterogeneous FL baseline methods.** High data heterogeneity represents $K = 2$ for CIFAR-10 and $K = 20$ for CIFAR-100; Lower data heterogeneity represents $K = 5$ for CIFAR-10 and $K = 50$ for CIFAR-100.

	Method	High Data Heterogeneity		Low Data Heterogeneity	
		CIFAR-10	CIFAR-100	CIFAR-10	CIFAR-100
KD-based	FedDF [38]	73.81 (± 0.42)	31.87 (± 0.46)	76.55 (± 0.32)	37.87 (± 0.31)
	DS-FL [24]	65.27 (± 0.53)	29.12 (± 0.51)	68.44 (± 0.47)	33.56 (± 0.55)
	Fed-ET [10]	**78.66 (± 0.31)**	**35.78 (± 0.45)**	**81.13 (± 0.28)**	**41.58 (± 0.36)**
PT-based	HeteroFL [12]	63.90 (± 2.74)	52.38 (± 0.80)	73.19 (± 1.71)	57.44 (± 0.42)
	Federated Dropout [6]	46.64 (± 3.05)	45.07 (± 0.07)	76.20 (± 2.53)	46.40 (± 0.21)
	ZeroFL [47]	64.61 (± 2.18)	51.39 (± 0.45)	83.31 (± 0.78)	53.62 (± 0.51)
	FedDST [5]	67.65 (± 1.27)	54.21 (± 0.34)	84.57 (± 0.28)	54.97 (± 0.44)
	Flash [2]	67.08 (± 1.46)	54.92 (± 0.29)	84.61 (± 0.37)	55.04 (± 0.32)
	FedRolex [1]	69.44 (± 1.50)	56.57 (± 0.15)	84.45 (± 0.36)	58.73 (± 0.33)
	FedSep (Ours)	**71.13 (± 0.94)**	**58.16 (± 0.25)**	**84.61 (± 0.37)**	**61.41 (± 0.29)**
	Homogeneous (smallest)	38.82 (± 0.88)	12.69 (± 0.50)	46.86 (± 0.54)	19.70 (± 0.34)
	Homogeneous (largest)	**75.74 (± 0.42)**	**60.89 (± 0.60)**	**84.48 (± 0.58)**	**62.51 (± 0.20)**