NEURAL INFORMATION PROCESSING SYSTEMS

Every Parameter Matters: Ensuring the Convergence of Federated Learning with Dynamic Heterogeneous Models Reduction

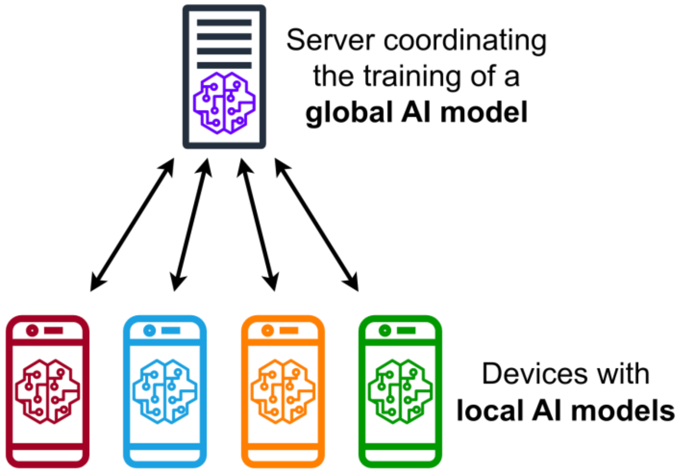
Hanhan Zhou [1], Tian Lan [1], Guru Venkataramani [1], Wenbo Ding [2]

[1] The George Washington University

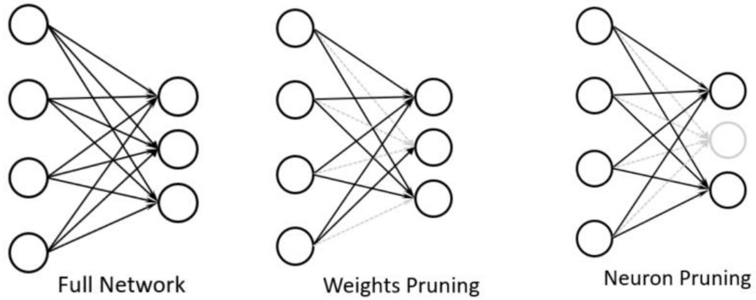
[2] Tsinghua-Berkeley Shenzhen Institute

Background

- Federated Learning



- Heterogeneous Federated Learning



Main Problem

- Given a heterogeneous FL algorithm that trains a shared global model through a sequence of time-varying and client-dependent local models, what conditions can guarantee its convergence?

- How do the trained models compare to that of standard FL?

Contributions

$$\theta_{q,n,T}$$

- We establish sufficient conditions for FL algorithms with heterogeneous local models to converge to a neighborhood of a stationary point of standard FL.

- We formulate the problem to allow any model reduction strategy and identify two key factors that impact the convergence:
  - pruning-induced noise
  - minimum coverage index

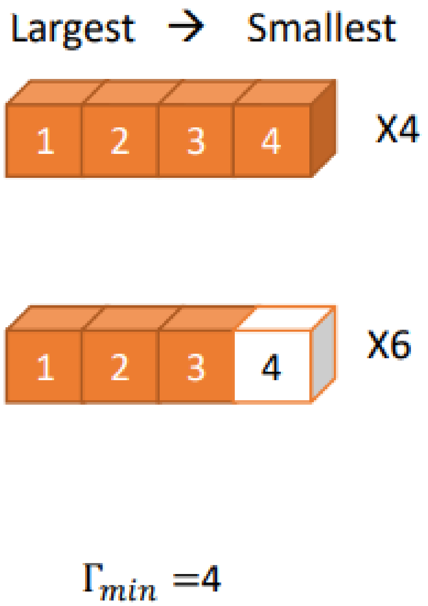
- The results are numerically validated.

Convergence Analysis

- Key Notion: minimum covering index

$$\Gamma_{\min} = \min_{q,i} |\mathcal{N}_q^{(i)}|$$

Since $|\mathcal{N}_q^{(i)}|$ is the number of heterogeneous local models containing the ith parameter, $\Gamma_{\min}$ measures the minimum occurrence of the parameter in the local models in all rounds.



(a)

Convergence Analysis

- Key Notion: minimum covering index

$$\Gamma_{\min} = \min_{q,i} |\mathcal{N}_q^{(i)}|$$

Since $|\mathcal{N}_q^{(i)}|$ is the number of heterogeneous local models containing the ith parameter, $\Gamma_{\min}$ measures the minimum occurrence of the parameter in the local models in all rounds.

Theorem 1. *Under Assumptions 1-4 and for arbitrary masks satisfying $\Gamma_{\min} \geq 1$, heterogeneous FL converges to a small neighborhood of a stationary point of standard FL as follows:*

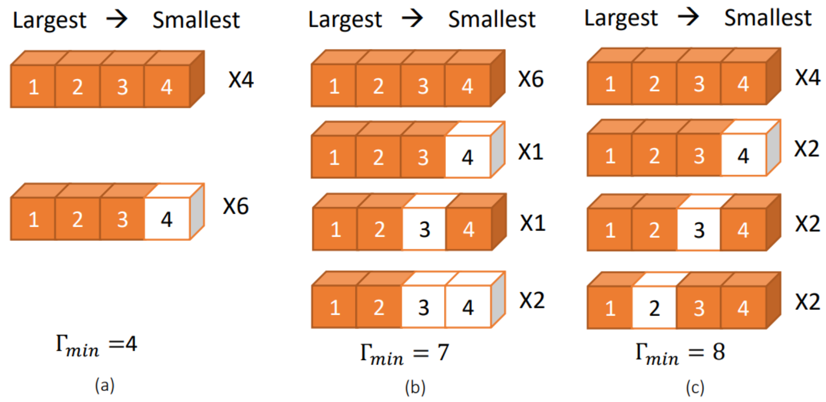
$$\frac{1}{Q}\sum_{q=1}^{Q}\mathbb{E}||\nabla F(\theta_q)||^2 \leq \frac{G_0}{\sqrt{TQ}} + \frac{V_0}{Q} + \frac{I_0}{\Gamma_{\min}} \cdot \frac{\delta^2}{Q}\sum_{q=1}^{Q}\mathbb{E}||\theta_q||^2$$

where $V_0 = 3L^2NG/\Gamma_{\min}$, $I_0 = 3L^2N$, and $G_0 = 4\mathbb{E}[F(\theta_0)] + 6LN\sigma^2/\Gamma_{\min}^2$, are constants depending on the initial model parameters and the gradient noise.

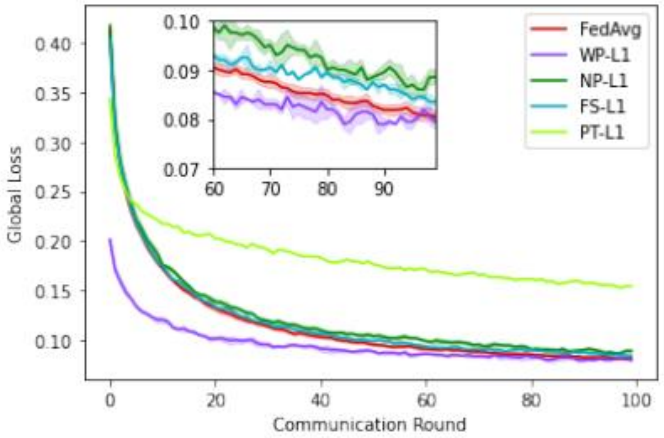
Insights

- Every Parameter Matters
  - The analysis shows that as long as each global parameter appears in at least one local model per communication round, heterogeneous federated learning can converge.
- More coverage leads to faster convergence

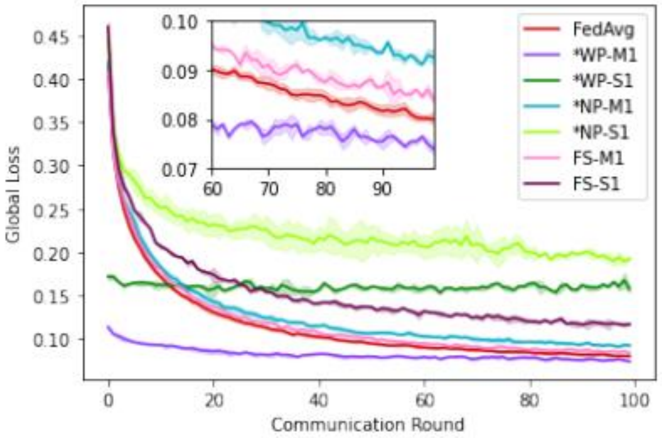
- Instead of pruning greedily for local heterogenous models, the minimum coverage index should also be considered



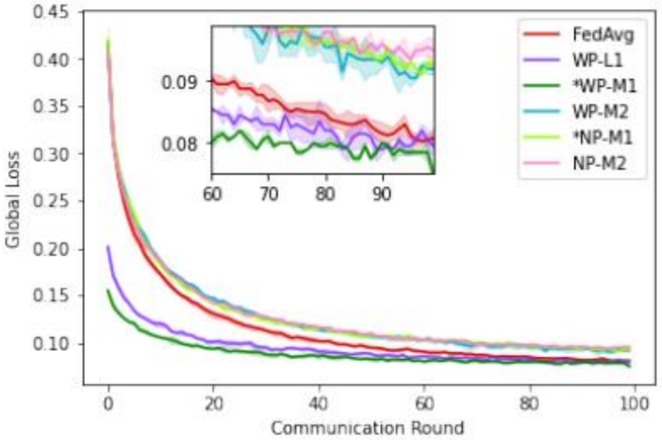
Experiments



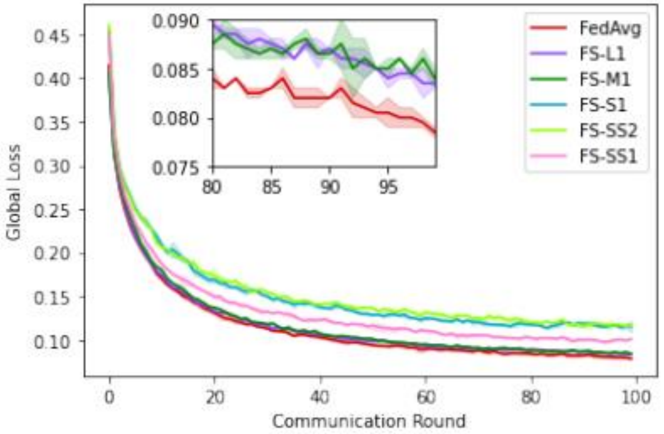
(a) Different Pruning Techniques



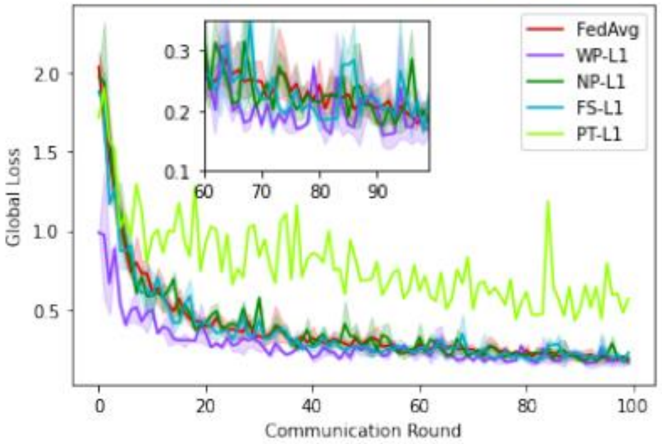
(b) Impact of Pruning Level



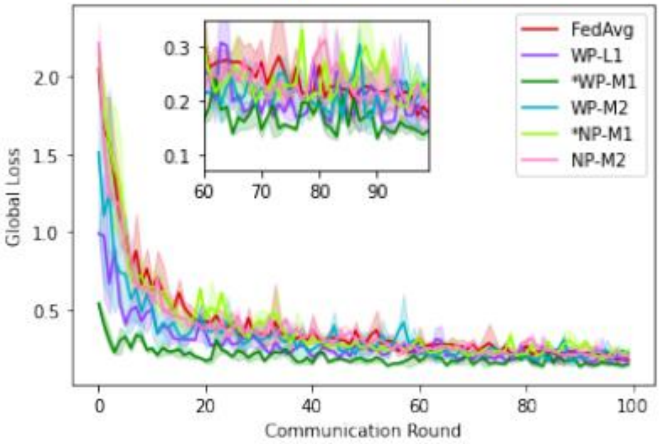
(c) Impact of Γ_{min}



(d) Pruning-induced Noise



(e) Pruning Techniques w/ non-IID



(f) Impact of Γ_{min} w/ non-IID

Conclusion

- We have provided convergence guarantees for heterogeneous federated learning algorithms employing arbitrary, dynamic local models under sufficient conditions.
- The analysis identifies model reduction noise and minimum coverage index as two key factors that impact the convergence gap.
- These insights can guide the design of optimized model reduction strategies to improve convergence.
- Experiments on image classification validate the theory and show optimized strategies guided by the analysis can improve accuracy.