NTK-DFL: Enhancing Decentralized Federated Learning in Heterogeneous Settings via Neural Tangent Kernel

NC STATE UNIVERSITY

Gabriel Thompson*, Kai Yue, Chau-Wai Wong, Huaiyu Dai

ICML International Conference On Machine Learning

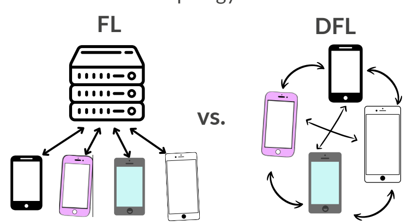
Introduction

Federated Learning (FL)

- Goal: train machine learning models across multiple clients
- Clients **do not share** training data. Instead, model updates are sent
- Privacy-preserving, yet centralized

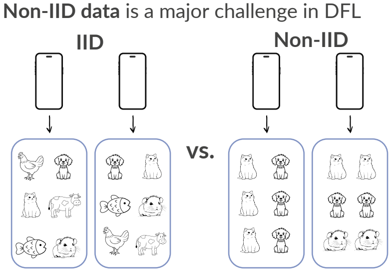
Decentralized Federated Learning (DFL)

- Training done **without** a central server
- Clients communicate on a sparse, decentralized topology



Background

- **Non-IID data** is a major challenge in DFL



Neural Tangent Kernel (NTK) & Weight Evolution

- NTK provides linearized solution to evolved function + weights at an arbitrary timestep
- NTK calculated from Jacobian matrices

$$\frac{d\mathbf{f}}{dt} = -\eta \mathbf{H} \nabla_{\mathbf{f}} \mathcal{L}$$

Proposed NTK-DFL Approach

NTK-DFL: Trades traditional gradient-based evolution for NTK evolution in decentralized FL

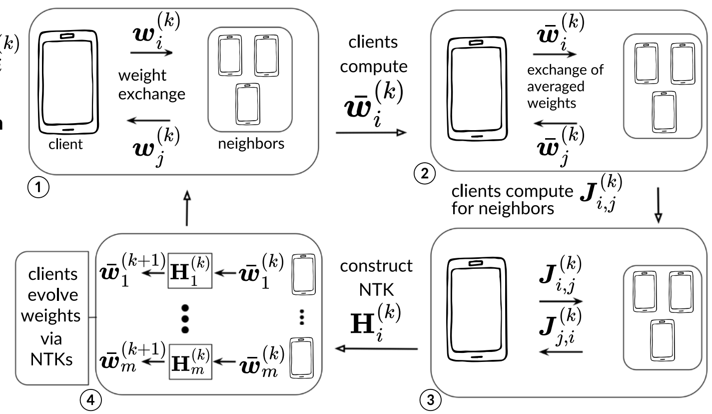
- Allows for clients to send more **expressive updates**
- **Decreases** the **number of communication rounds** needed between clients

1) **Model Averaging**
Clients average weights $\boldsymbol{w}_i^{(k)}$ with neighbors

2) **Jacobian Communication**
Clients compute and share Jacobians $\boldsymbol{J}_{i,j}^{(k)}$

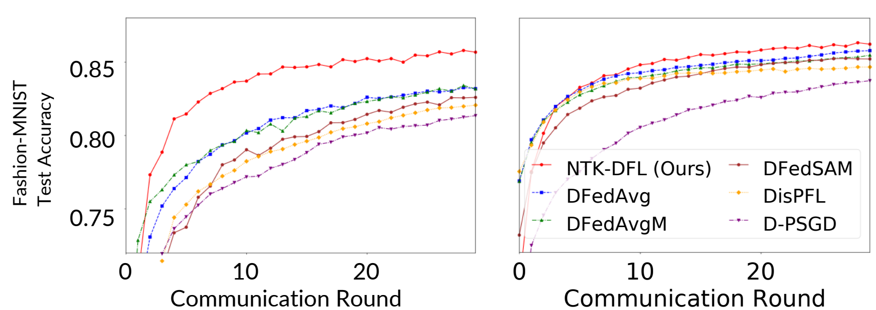
3) **Kernel Construction**
- Clients build NTK $\mathbf{H}_i^{(k)}$ from neighbor's data
- Analytical evaluation, weight prediction

4) **Model Selection**
Optimal weights chosen



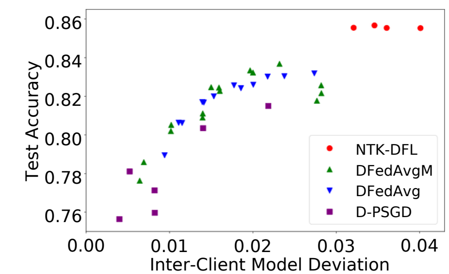
Experimental Results: Fewer Communication Rounds

- NTK-DFL **reduces communication rounds** needed for convergence
- Particularly effective with **non-IID** data
  - **Figure:** Test accuracy vs. communication round of various methods (Fashion-MNIST)
  - **Left:** Highly **non-IID data** distribution among clients
  - **Right: IID** data distribution among clients

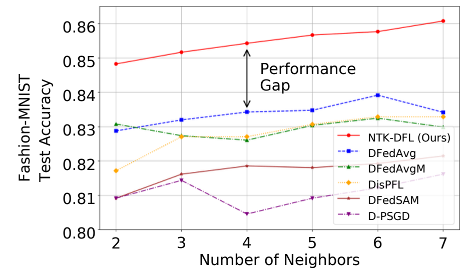


Experimental Results

- NTK updates encourage a **useful inter-client model deviation**
- **Synergetic with DFL model averaging,** improving generalization



- NTK-DFL test accuracy is resilient across varying **topological sparsity**



Conclusion

- NTK-DFL **improves convergence** through **expressive NTK updates**
- Especially effective in **non-IID settings**
- Discovered a **useful synergy** with DFL model averaging, improving generalization

Check out **our paper** here!