Neural Tangent Kernel Empowered Federated Learning

Kai Yue[1] Richeng Jin[1] Ryan Pilgrim[2]
Chau-Wai Wong[1] Dror Baron[1] Huaiyu Dai[1]

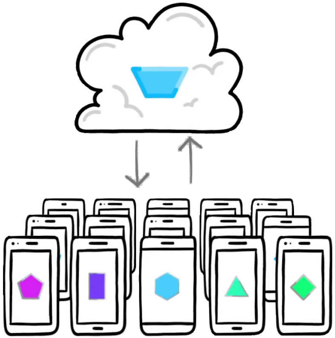
[1]NC State University, USA [2]Independent Scholar

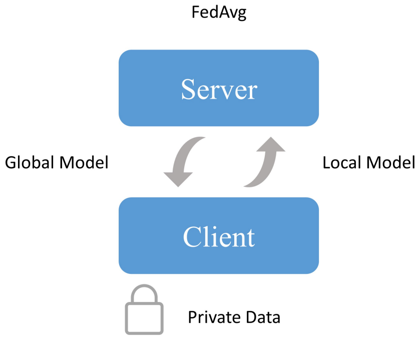
ICML 2022

June 27, 2022

Federated Learning (FL)

- Clients with private data jointly solve a machine learning task
- Raw data stored locally & not exchanged





[1] https://ai.googleblog.com/2017/04/federated-learning-collaborative.html

FL Challenge: Client Non-IID Data

- Deviate from independent & identically distributed (IID)

FL Challenge: Client Non-IID Data

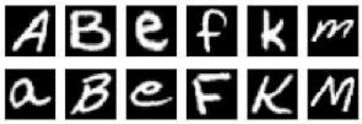
- Deviate from independent & identically distributed (IID)

- **Example:** feature skew

FL Challenge: Client Non-IID Data

- Deviate from independent & identically distributed (IID)

- **Example:** feature skew & label skew



FL Challenge: Client Non-IID Data

- Deviate from independent & identically distributed (IID)

- **Example:** feature skew & label skew

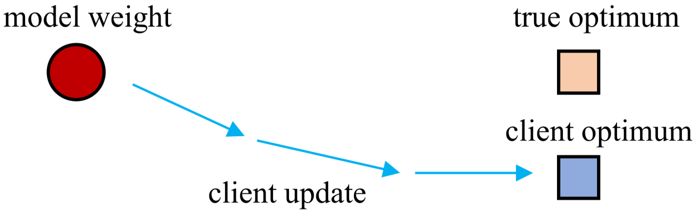
- Non-IID data can significantly lower model performance

Client Update vs. Server Update

- **Goal:** find ***true optimum*** that generalizes well

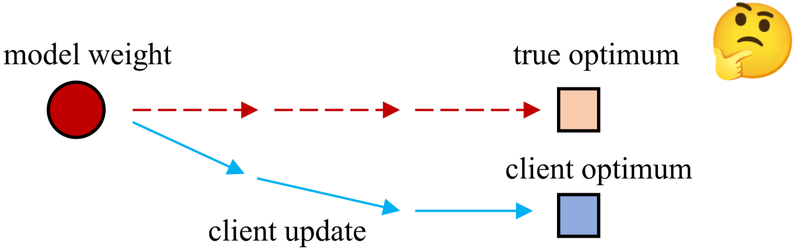
Client Update vs. Server Update

- **Goal:** find ***true optimum*** that generalizes well
- **Issue:** *Multi-step* ***client*** *update* leads to a ***client optimum***



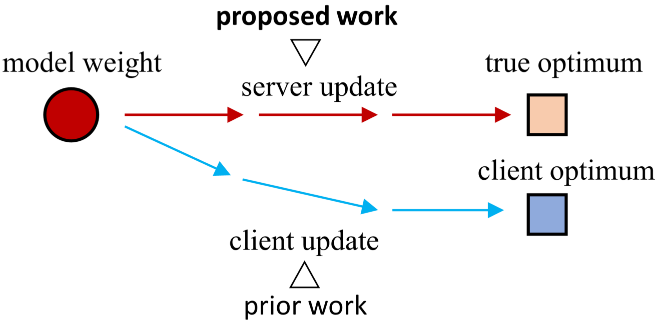
Client Update vs. Server Update

- **Goal:** find ***true optimum*** that generalizes well
- **Issue:** *Multi-step* ***client*** *update* leads to a ***client optimum***
- **Research Question:**
  Is it possible to shift *multi-step update* to server?



Client Update or Server Update

- Is it possible to shift *multi-step update* to server?
- Proposed NTK-FL enables *multi-step* ***server*** *update*



NTK: Neural Tangent Kernel

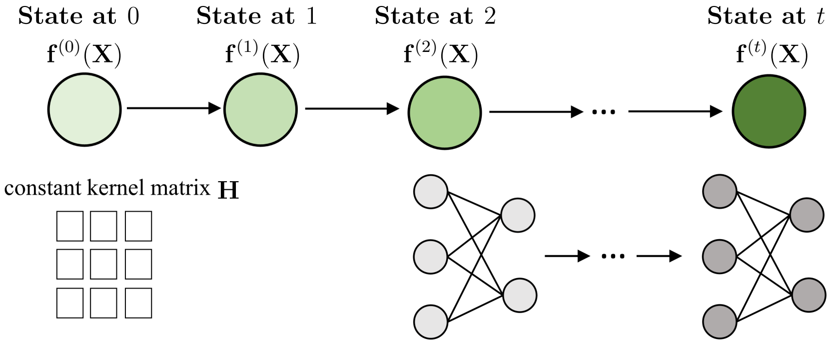
- Approximate training dynamics with a differential equation (DE)

$$\frac{\mathrm{d}\mathbf{f}}{\mathrm{d}t} = \eta\,\mathbf{H}\left[\mathbf{Y} - \mathbf{f}^{(t)}(\mathbf{X})\right]$$

[2] Arthur Jacot, Franck Gabriel, and Clément Hongler. "Neural tangent kernel: Convergence and generalization in neural networks." NeurIPS 2018.

NTK: Neural Tangent Kernel

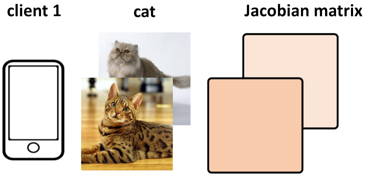
- Approximate training dynamics with a differential equation (DE)
- The state evolution can be captured by DE solution



[2] Arthur Jacot, Franck Gabriel, and Clément Hongler. "Neural tangent kernel: Convergence and generalization in neural networks." NeurIPS 2018.

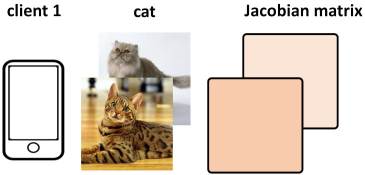
NTK-FL: Client Calculation

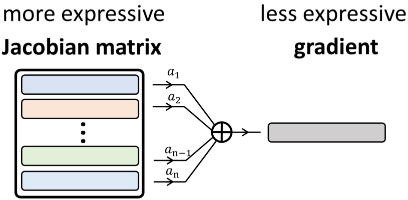
- clients calculate Jacobian matrices
  - **without** local update & avoid a local optimum



NTK-FL: Client Calculation

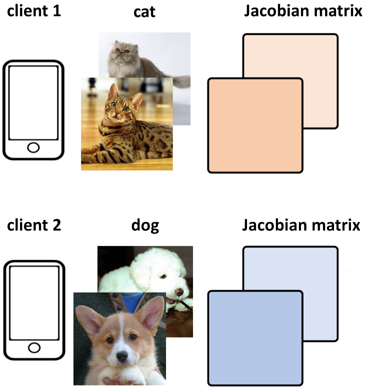
- clients calculate Jacobian matrices
  - *more expressive* & preserve client information





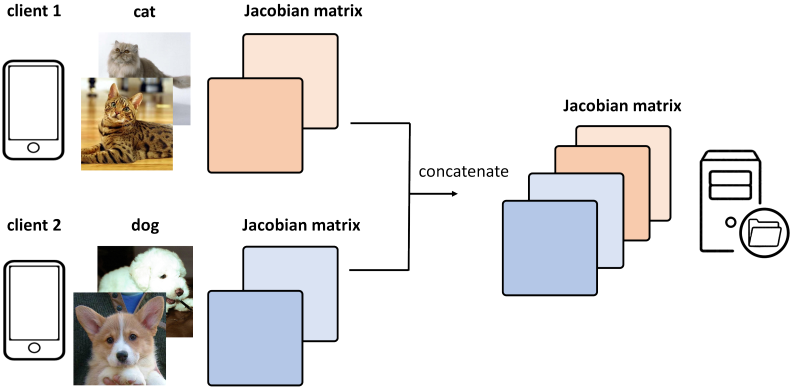
NTK-FL: Client Calculation

- clients calculate Jacobian matrices
- server concatenates the Jacobian matrices



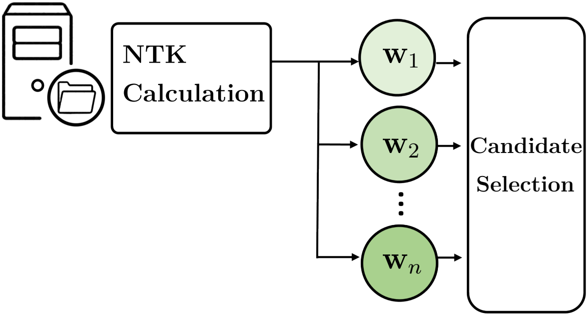
NTK-FL: Client Calculation

- clients calculate Jacobian matrices
- server concatenates the Jacobian matrices



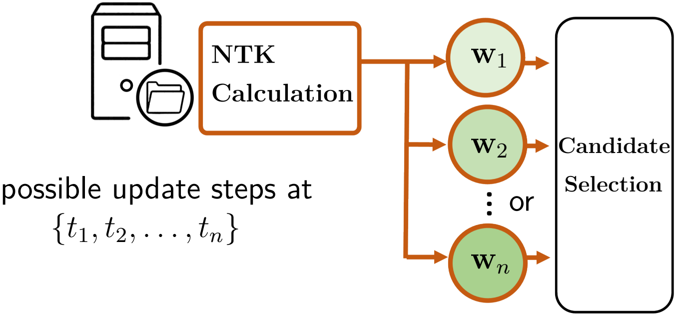
NTK-FL: Server Calculation

- Obtain different model weights via NTK evolution



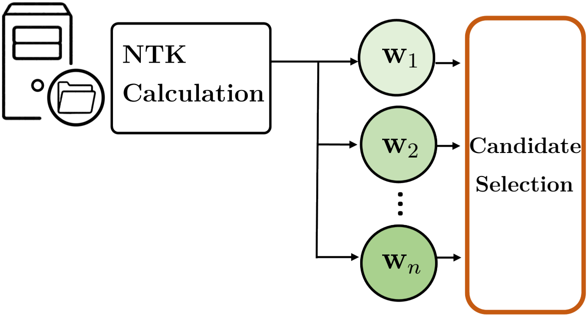
NTK-FL: Server Calculation

- Obtain different model weights via NTK evolution
  - multi-step update is shifted to the server



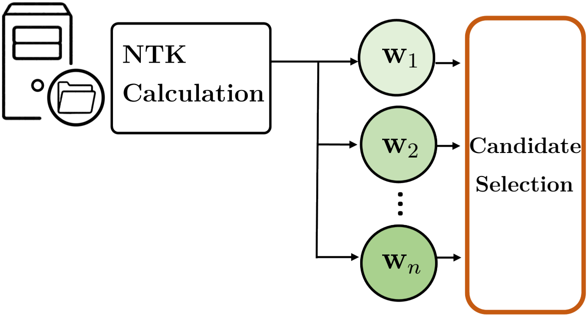
NTK-FL: Server Calculation

- Select the weight $\mathbf{w}_j$ that gives the lowest loss



NTK-FL: Server Calculation

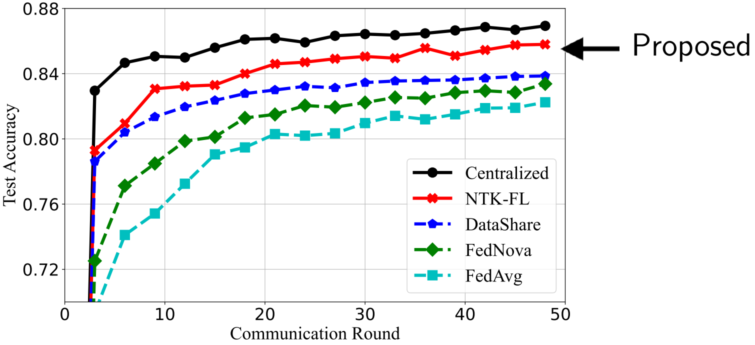
- Select the weight $\mathbf{w}_j$ that gives the lowest loss
  - **dynamic** update steps in different communication rounds



Experiments: Non-IID Fashion-MNIST

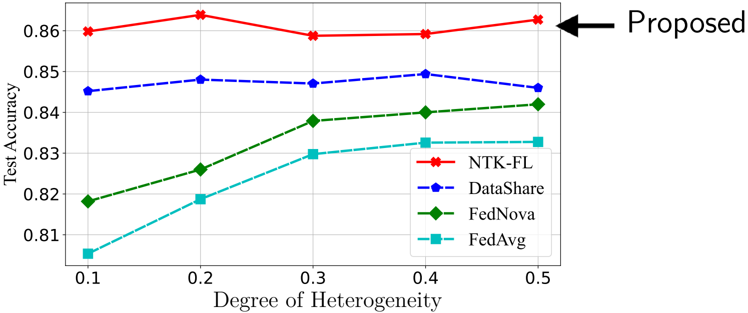
- Learning curves of different methods
  - NTK-FL approaches centralized learning

  [Centralized: selected clients share raw data]



Experiments: Non-IID Fashion-MNIST

- Test accuracy with various degrees of heterogeneity
  - NTK-FL is robust in different non-IID scenarios



Conclusion

- NTK-FL transmits more expressive Jacobian matrix
  - enable multi-step server update
  - reduce the negative influence of data heterogeneity
  - adaptively choose the number of update steps
- Please refer to our paper for potential challenges
  - additional communication cost
  - privacy concerns

Neural Tangent Kernel Empowered Federated Learning

Kai Yue[1] Richeng Jin[1] Ryan Pilgrim[2]
Chau-Wai Wong[1] Dror Baron[1] Huaiyu Dai[1]

[1]NC State University, USA [2]Independent Scholar

ICML 2022