Federated Learning with Partial Model Personalization

Krishna Pillutla, Kshitiz Malik, Abdelrahman Mohamed, Michael Rabbat, Maziar Sanjabi, Lin Xiao

Personalized federated learning

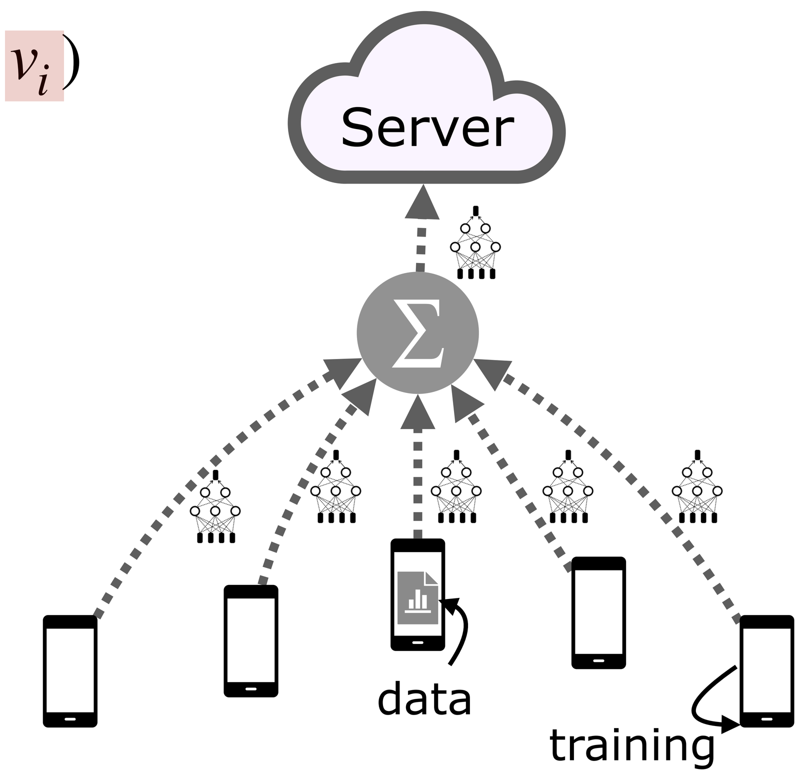
Model on client $i = (u, v_i)$

Objective:

$$\min_{u, v_1, \cdots, v_n} \frac{1}{n} \sum_{i=1}^{n} F_i(u, v_i)$$

u: shared parameters

v_i: personal parameters



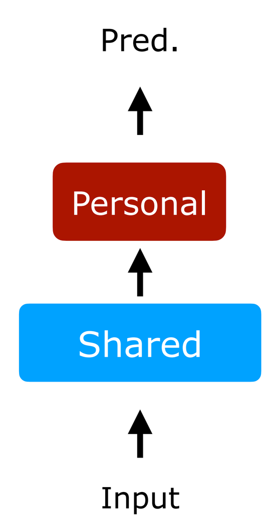
Our Contributions

1. Theory: Analysis of 2 popular optimization algos

2. Extensive experiments: text, vision, and speech settings

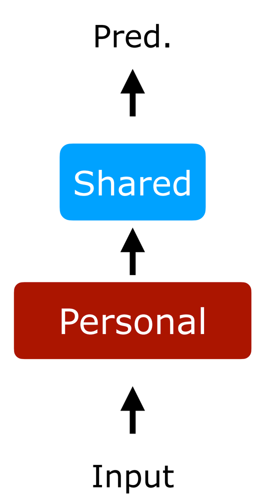
Personalization Architectures

Personalized output layer



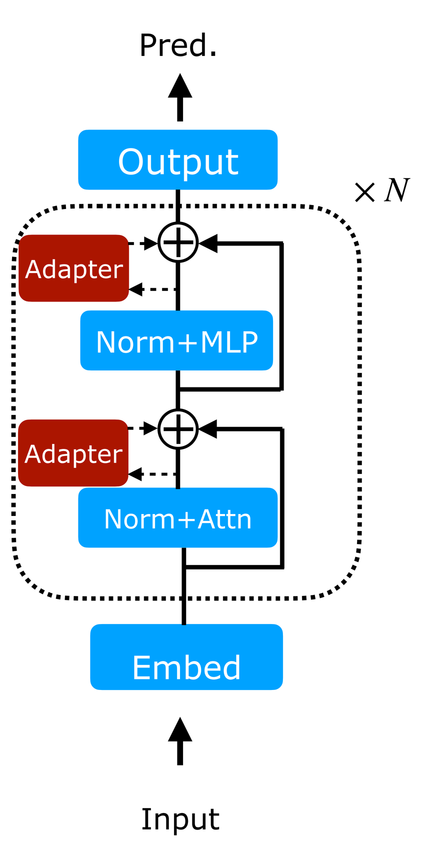
Arivazhagan et al. (2019)
Collins et al. ICML (2021)

Personalized input layer

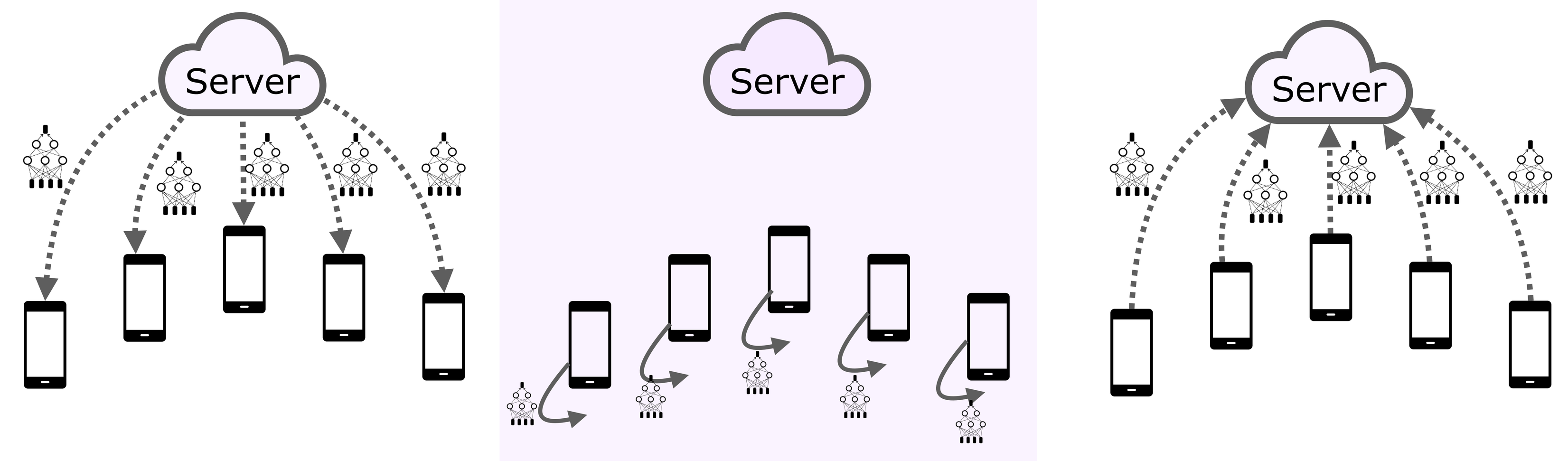


Liang et al. (2019)

Personalized adapters



Optimization Algorithms



Alternating update

$$v_i^+ = v_i - \gamma \nabla_v F_i(u, v_i)$$

$$u_i^+ = u - \gamma \nabla_u F_i(u, v_i^+)$$

Collins et al. ICML (2021)
Singhal et al. NeurIPS (2021)

Simultaneous update

$$v_i^+ = v_i - \gamma \nabla_v F_i(u, v_i)$$

$$u_i^+ = u - \gamma \nabla_u F_i(u, v_i)$$

Liang et al. (2019)
Arivazhagan et al. (2019)

1. Theory

Theorem For smooth, nonconvex functions and client sampling, we have the rates:

Alternating update: $\dfrac{\sigma_1^2}{\sqrt{t}}$

Simultaneous update: $\dfrac{\sigma_2^2}{\sqrt{t}}$

where $\sigma_1^2 < \sigma_2^2$ under typical scenarios

Key technical challenge: Dependent random variables due to random sampling of clients

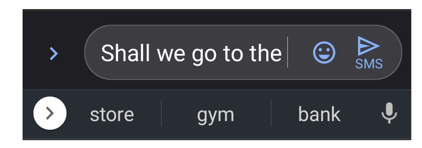
Methodology: virtual full participation

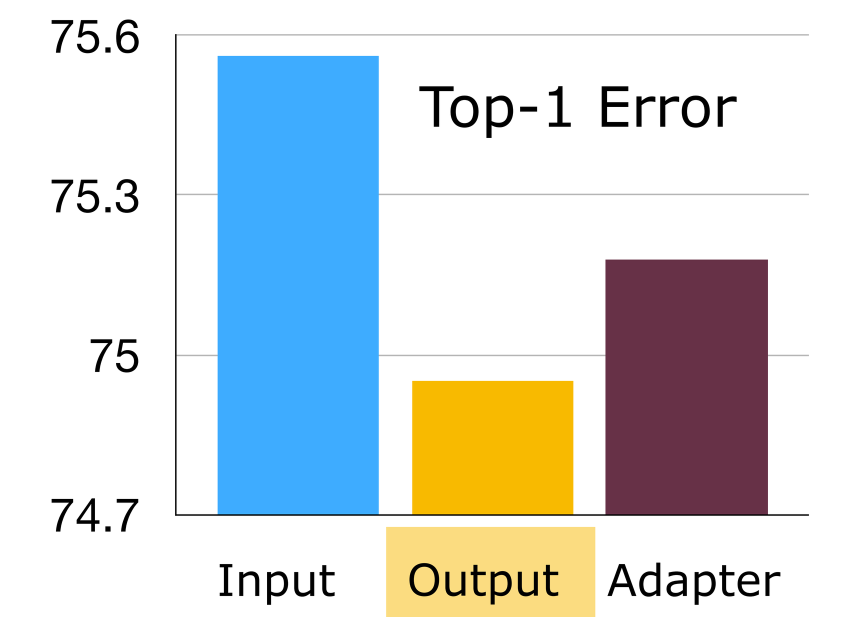
Experimentally, small but consistent trend of alternating > simultaneous

2. Experiments

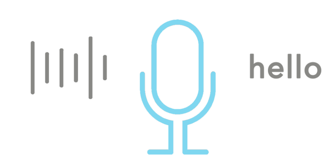
Best personalization architecture depends on nature of data heterogeneity

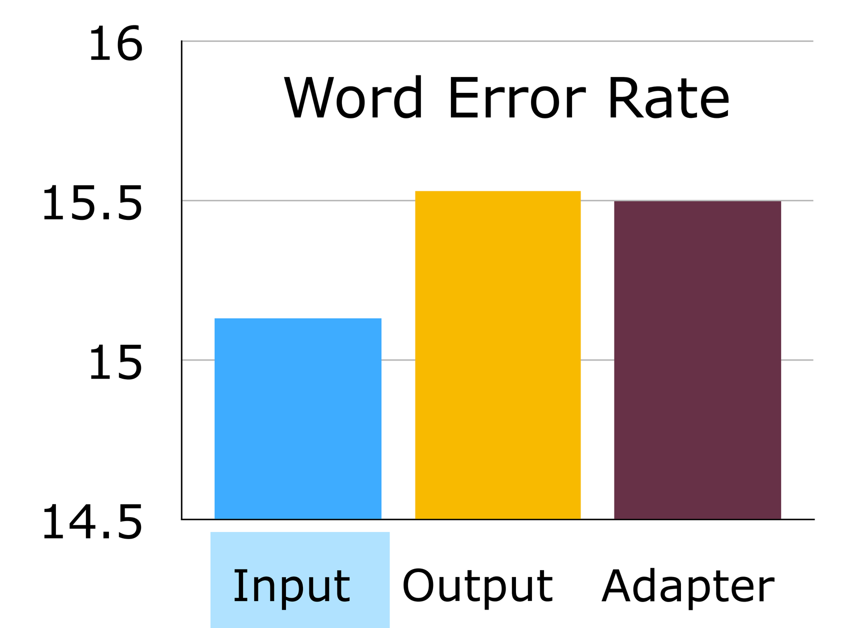
Next word prediction



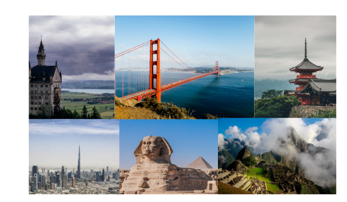


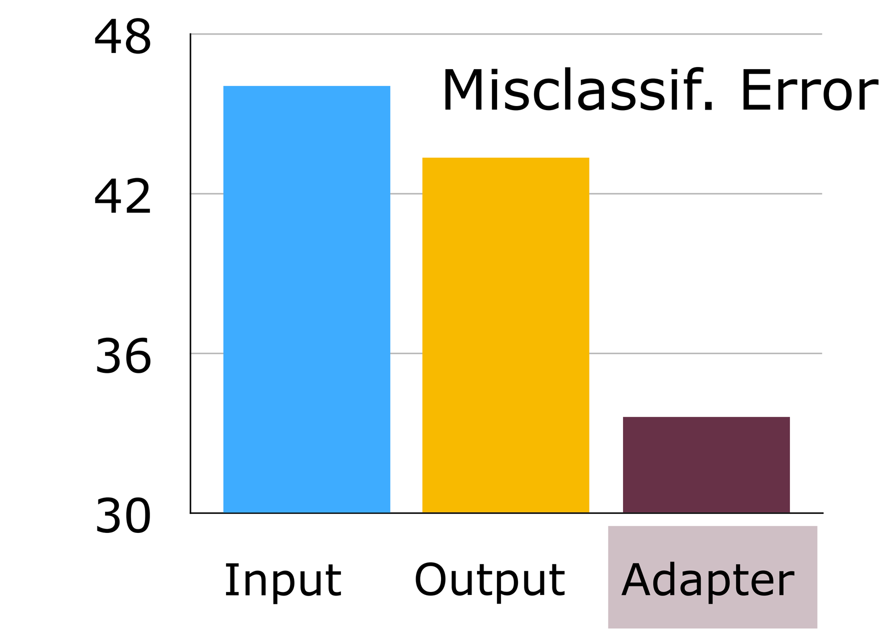
Speech recognition





Landmark recognition





Choose your personalization architecture wisely!

krishnap25
KrishnaPillutla
krishnap25.github.io

Code

SCAN ME