FedPeWS: Personalized Warmup via Subnetworks for Enhanced Heterogeneous Federated Learning

Nurbek Tastan[1], Samuel Horváth[1], Martin Takáč[1], Karthik Nandakumar[1,2]

[1]Mohamed bin Zayed University of Artificial Intelligence (MBZUAI), [2]Michigan State University (MSU)

Motivation

- Federated Learning (FL) enables collaborative training while preserving privacy.
- Extreme heterogeneity in client data slows convergence and impacts performance.
- Existing FL methods struggle when client data distributions vary significantly.
- **Solution:** `FedPeWS` introduces a personalized warmup phase, improving convergence and model performance.

Problem Formulation and Contributions

Our goal is to minimize a sum-structured federated learning optimization objective

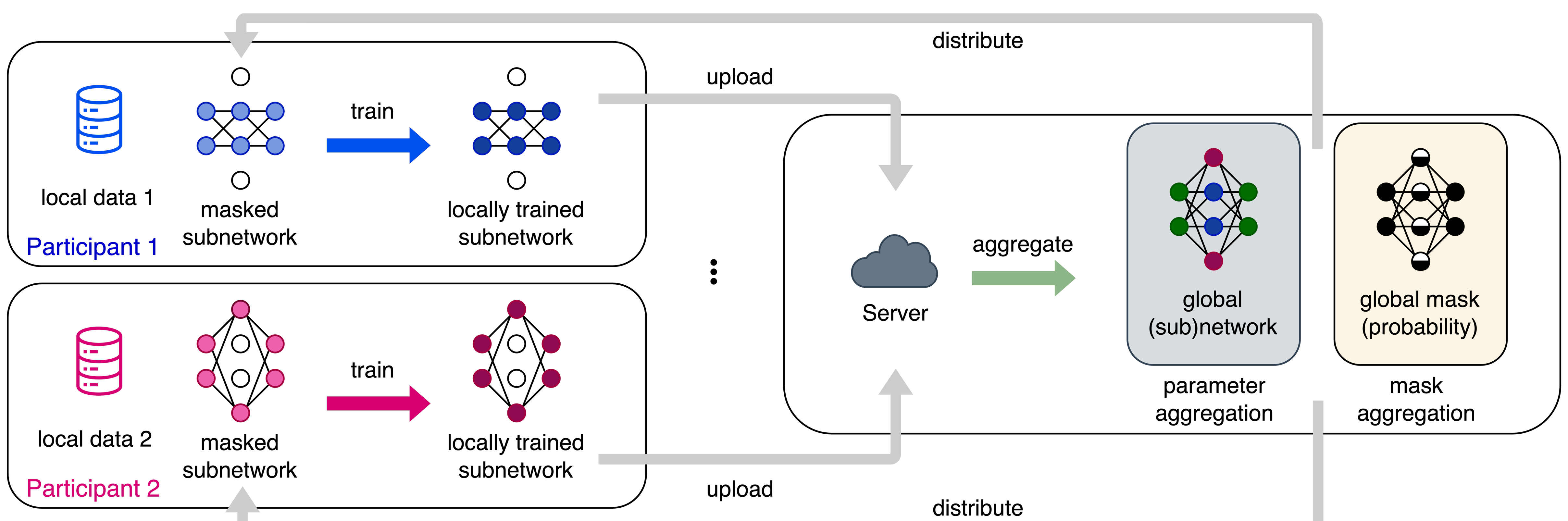
$$\boldsymbol{x}^\star \leftarrow \arg\min_{\boldsymbol{x}\in\mathbb{R}^d}\left[f(\boldsymbol{x}) := \frac{1}{N}\Sigma_{i=1}^N f_i(\boldsymbol{x})\right],$$

$$f_i(\boldsymbol{x}) := \mathbb{E}_{\xi\sim\mathcal{D}_i}[F_i(\boldsymbol{x},\xi)]$$

Key Contributions:

- We propose `FedPeWS`, a novel FL approach that mitigates heterogeneity-induced conflicts by introducing a neuron-level personalized warmup phase, improving generalization and convergence speed.
- Our algorithm identifies optimal subnetworks for each client by jointly learning personalized masks and parameter updates.
- For small-scale FL with known data distributions, we introduce `FedPeWS-Fixed`, which eliminates mask learning by assigning predefined subnetworks, reducing computational overhead.
- Extensive experiments on synthetic, CIFAR10-MNIST, medical datasets (PathMNIST, OCTMNIST, TissueMNIST), and CIFAR100 confirm FedPeWS consistently improves convergence and accuracy in both extreme non-IID and IID scenarios.

Proposed Methodology



Conceptual illustration of training personalized subnetworks in FL.

Phase 1. Warmup

- Clients train only a subset of the model (masked neurons)
- Minimizes conflicts before global aggregation

The update rule:

$$\boldsymbol{x}_i^t = \boldsymbol{x}_i^{t-1} - \eta_\ell \nabla f_i\left(\boldsymbol{x}_i^{t-1}\odot m_i^{t-1}, \xi_i^{t-1}\right), \forall i, \forall t.$$

Identification of subnetworks

1. Neuron-level score mask vectors $s_i \in \mathbb{R}^h$, $h \ll d$.
2. $\mathcal{G}: \mathbb{R}^h \to \{0,1\}^d$ - mask generation function that generates the binary parameter-level masks m_i from neuron-level score mask vectors, i.e., $m_i = \mathcal{G}(s_i)$. $\mathcal{G}$ consists of three steps:

- convert s_i into probabilities by applying a sigmoid function, i.e. $\theta_i = \sigma(s_i)$, where $\theta_i \in [0,1]^h$.
- sample binary neuron masks $\tilde{m}_i$ from a Bernoulli distribution with parameter θ_i, i.e., $\tilde{m}_i(\ell) \sim Bernoulli(\theta_i(\ell)), \forall \ell \in [h]$.
- map the neuron-level masks $\tilde{m}_i$ to parameter-level masks m_i.

3. **Procedure I. Mask training**

$$\mathcal{L}_s = f_i(\boldsymbol{x}_i^t \odot \mathcal{G}(s_i^t), \xi_i^t) - \lambda\|\sigma(s_i^t) - \theta_{g\setminus\{i\}}^t\|_2^2;$$
$$s_i^{t+1} \leftarrow s_i^t - \eta_s \nabla_s \mathcal{L}_s.$$

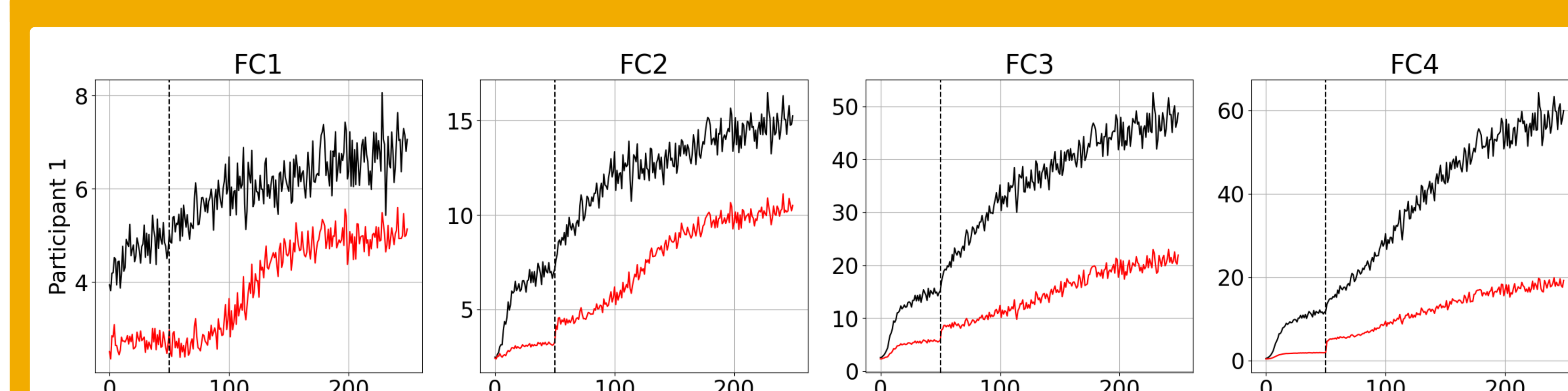
4. **Procedure II. Weight training**

$$\mathcal{L}_{\boldsymbol{x}} = f_i(x_i^t \odot \mathcal{G}(s_i^t), \xi_i^t); \quad \boldsymbol{x}_i^{t+1} \leftarrow \boldsymbol{x}_i^t - \eta_l \nabla_{\boldsymbol{x}} \mathcal{L}_{\boldsymbol{x}}.$$

Phase 2. Standard FL

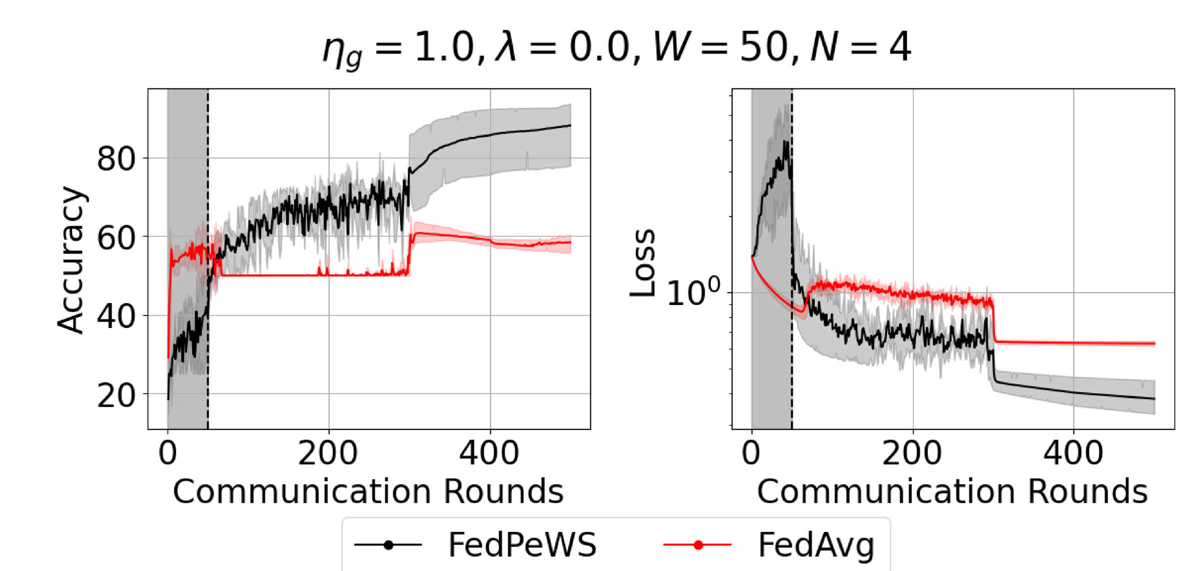
- Full model updates start after warmup, leading to better convergence

Important Result



Neuron activation study on synthetic dataset. The experiment uses the `FedPeWS-Fixed` method with $W = 50$ warmup rounds, indicated by the vertical dashed line.

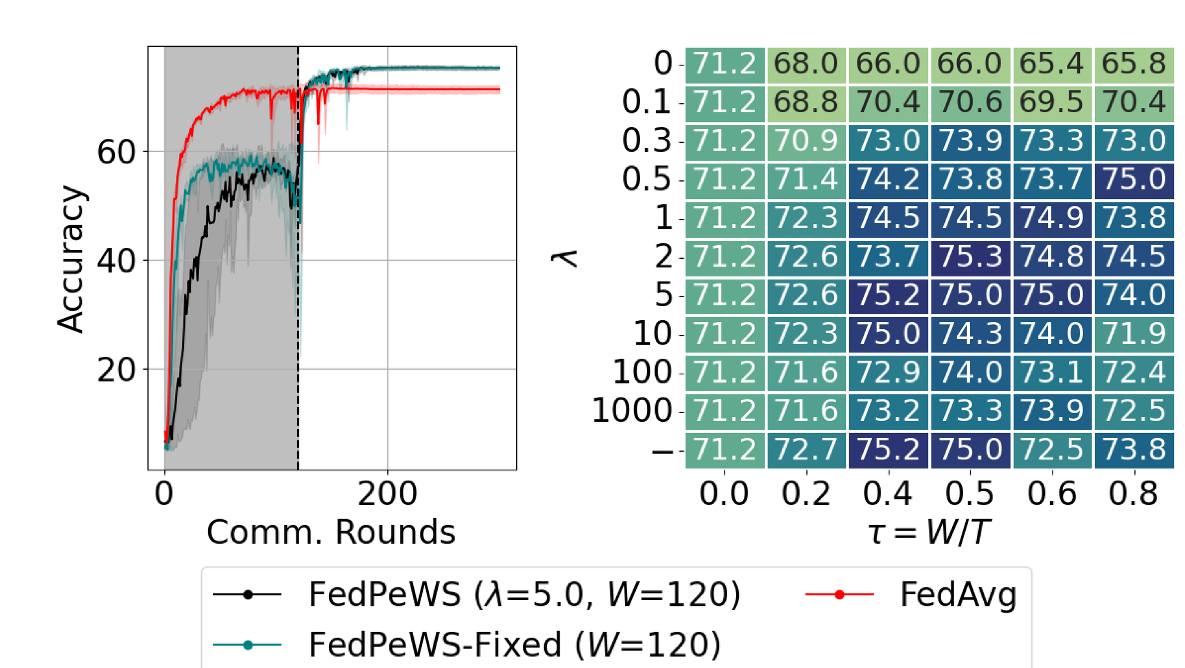
Experiments



Visualization of accuracy and losson synthetic dataset with $N = 4$ participants each having 1 distinct class.

Dataset	CIFAR-MNIST	{P-O-T}MNIST
Fedavg	71.78 ± 0.66	52.83 ± 1.26
FedProx	72.27 ± 0.88	51.28 ± 1.03
SCAFFOLD	71.83 ± 0.24	53.05 ± 0.60
FedNova	71.63 ± 0.98	53.05 ± 0.83
MOON	71.84 ± 1.09	52.10 ± 0.19
`FedAvg+PeWS`	$\mathbf{75.83 \pm 0.88}$	$\mathbf{55.12 \pm 0.56}$
`FedProx+PeWS`	75.04 ± 0.85	54.67 ± 0.43

Comparison to the state-of-the-art algorithms in federated learning that tackle heterogeneity.



Performance visualization and sensitivity analysis on CIFAR-MNIST dataset.

Mohamed bin Zayed University of Artificial Intelligence