ConFREE: Conflict-free Client Update Aggregation for Personalized Federated Learning

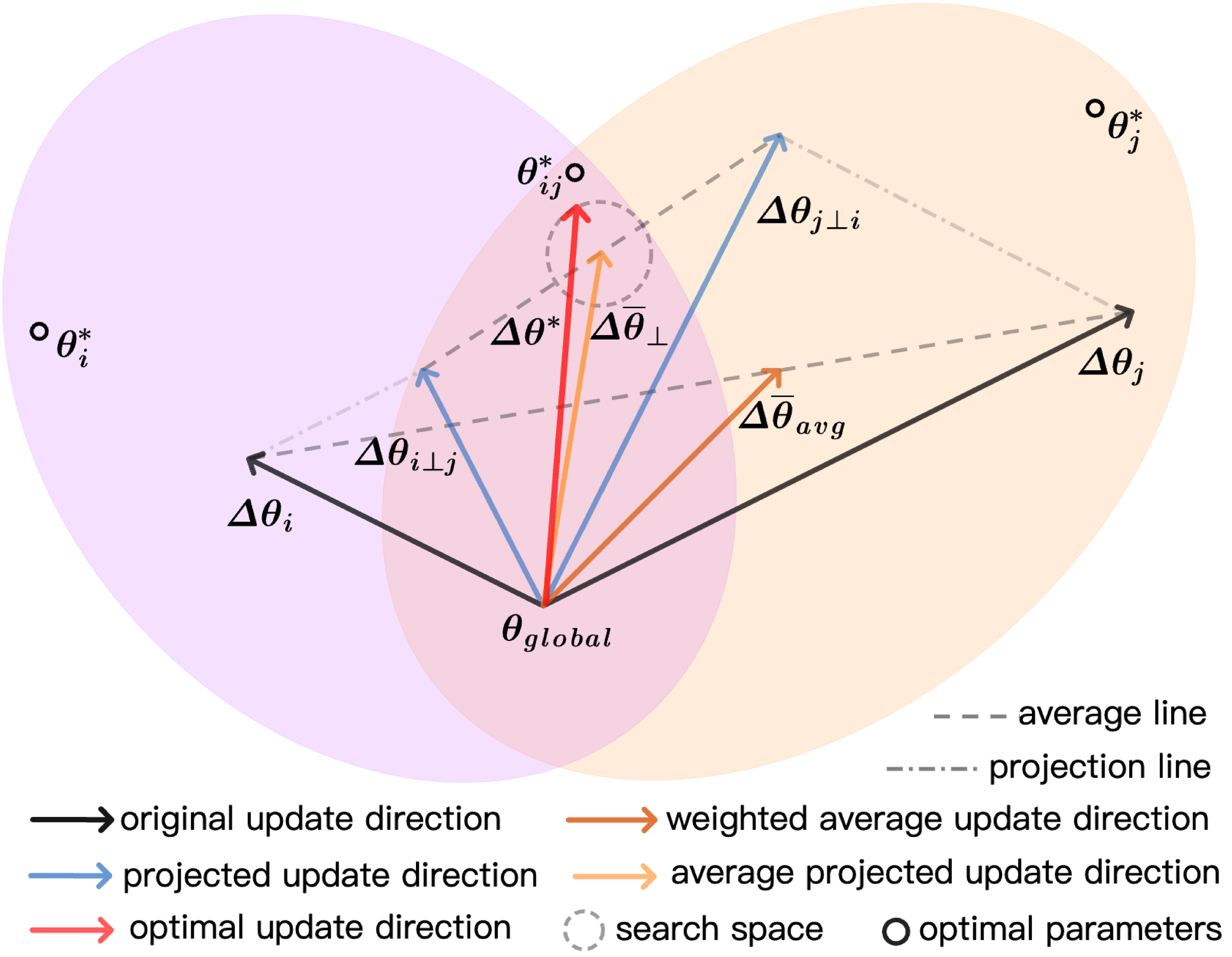
Hao Zheng[1,2], Zhigang Hu[1], Liu Yang[1*], Meiguang Zheng[1], Aikun Xu[1], and Boyu Wang[2*]

[1] Central South University [2] University of Western Ontario

Association for the Advancement of Artificial Intelligence
AAAI-25 / IAAI-25 / EAAI-25
FEBRUARY 25 – MARCH 4, 2025 | PHILADELPHIA, USA

Introduction and Motivation

- **Personalized Federated Learning:** aims to learn a tailored model for each client while still benefiting from the global model to enhance their models and address data scarcity.
- **Key Challenge:** Negative transfer (NF) is a critical challenge in PFL, and existing methods primarily focus on adapting the local data distribution on the client side, they only resist NF rather than fundamentally avoiding it.



- **Motivation:** when $\Delta\theta_i \cdot \Delta\theta_j < 0$, the update from client i negatively impacts client j because it leads to an increase in the loss of client j, which is referred to as client update conflict.

$$\Delta\mathcal{L}_j = \mathcal{L}_j(h_j, \theta_g + \Delta\theta_i) - \mathcal{L}_j(h_j, \theta_g) = -\frac{1}{\alpha}\Delta\theta_i \cdot \Delta\theta_j \tag{1}$$

$$\Delta\mathcal{L}_{\text{total}} = -\frac{1}{\alpha}\sum_{i=1}^{N}\sum_{j=1}^{N} p_i p_j \Delta\theta_i \cdot \Delta\theta_j \tag{2}$$

The classical FedAvg $\Delta\bar{\theta}_{\text{avg}}$ tends to be dominated by the $\Delta\theta_j$ from clients with more data, leading to insufficient optimization for client i.

Contributions

- ConFREE is the first method to address client update conflicts in pFL. By optimizing global model aggregation, it provides each client with more effective and comprehensive global information.
- ConFREE is model-agnostic, which requires a single modification to the server aggregation program that can be easily applied to enhance various client-side personalization techniques.

Methodology

- **Guidance Vector:** minimizes negative conflicts with parameter updates from other clients.

$$\Delta\bar{\theta}_\perp = \frac{1}{N}\sum_{i=1}^{N}\left(\Delta\theta_i - \sum_{j\neq i, \cos\theta_{ij}<0}\frac{\Delta\theta_i \cdot \Delta\theta_j}{\|\Delta\theta_j\|^2}\Delta\theta_j\right) \tag{3}$$

- **Optimization Objective:** maximize the worst local improvement among all clients.

$$\min \Delta\mathcal{L}_{\text{total}} \approx \max_{\mathbf{d}\in\boldsymbol{\Theta}} \frac{1}{N}\sum_{j=1}^{N}\min_{j\in[N]} \mathbf{d}\cdot\Delta\theta_j \quad \text{s.t.} \quad \|\mathbf{d} - \Delta\bar{\theta}_\perp\| \le c\|\Delta\bar{\theta}_\perp\|$$

- **Dual Problem:**

$$\min_{w\in\mathcal{W},\lambda\ge 0}\max_{\mathbf{d}\in\boldsymbol{\Theta}} \quad \mathbf{d}\cdot\Delta\theta_w - \frac{\lambda}{2}\left(\|\mathbf{d} - \Delta\bar{\theta}_\perp\|^2 - c^2\|\Delta\bar{\theta}_\perp\|^2\right) \tag{4}$$

where $d^* = \Delta\bar{\theta}_\perp + \frac{c\|\Delta\bar{\theta}_\perp\|}{\|\Delta\theta_{w^*}\|}\Delta\theta_{w^*}$ and $\lambda = \frac{\|\Delta\theta_{w^*}\|}{c\|\Delta\bar{\theta}_\perp\|}$.

- **Final Objective Function:**

$$w^* = \arg\min_{w\in\mathcal{W}} \quad \mathbf{d}\cdot\Delta\theta_w + c\|\Delta\bar{\theta}_\perp\|\|\Delta\theta_w\| \tag{5}$$

ConFREE

Input: N clients, local datasets $\{\mathcal{D}_1, \ldots, \mathcal{D}_N\}$, learning rates $\{\eta, \alpha\}$, total communication rounds T.
Output: Conflict-free global aggregation parameter θ_g^{t+1}.

Initialize global parameter θ_g.
for each global iteration $t = 1, 2, \ldots, T$ **do**
  for client $i = 1, 2, \ldots, N$ **in parallel do**
    Update local head h_i^{t+1} via $h_i^{t+1} \leftarrow h_i^t - \eta\nabla_{h_i^t}\mathcal{L}_i\left(h_i^t, \theta_g^t; D_i\right)$
    Update shared parameter θ_i^t via $\theta_i^t \leftarrow \theta_g^t - \alpha\nabla_{\theta_g^t}\mathcal{L}_i\left(h_i^{t+1}, \theta_g^t; D_i\right)$
    Compute shared parameter update $\Delta\theta_i^t = \theta_i^t - \theta_g^t$.
  end for
  for each client update $\Delta\theta_i^t$ **do**
    Adjust $\Delta\theta_i^t$ using projection for all $\Delta\theta_j^t$ (where $j \neq i$) with negative cosine similarity.
  end for
  Compute the conflict-free guidance vector by averaging the adjusted updates using
  $\Delta\bar{\theta}_\perp = \frac{1}{N}\sum_{i=1}^{N}\left(\Delta\theta_i - \sum_{j\neq i,\cos\theta_{ij}<0}\frac{\Delta\theta_i\cdot\Delta\theta_j}{\|\Delta\theta_j\|^2}\Delta\theta_j\right)$.
  Solve optimization problem to find w^* using
  $w^* = \arg\min_{w\in\mathcal{W}} \quad \mathbf{d}\cdot\Delta\theta_w + c\|\Delta\bar{\theta}_\perp\|\|\Delta\theta_w\|$.
  Compute λ and d^* based on the solution.
  Server Aggregation: $\theta_g^{t+1} \leftarrow \theta_g^t + d^*$.
end for
return θ_g^{t+1}.

Empirical Results

Methods	Pathological heterogeneous		Practical heterogeneous($\beta = 0.1$)				#S.	#T.
	Flowers102	CIFAR100	CIFAR10	CIFAR100	TINY	Flowers102		
Per-FedAvg	67.04	61.77	87.76	41.49	25.62	55.33	151.72	66.86
+ ConFREE	67.64 (↑ 0.60)	62.45 (↑ 0.68)	88.23 (↑ 0.47)	41.76 (↑ 0.27)	27.42 (↑ 1.80)	55.96 (↑ 0.63)	148.19	67.38
LG-FedAvg	63.89	63.32	89.23	47.63	33.75	51.48	148.39	46.71
+ ConFREE	64.09 (↑ 0.20)	63.48 (↑ 0.16)	89.52 (↑ 0.29)	47.88 (↑ 0.25)	33.96 (↑ 0.21)	52.72(↑ 1.24)	144.86	46.94
APFL	68.87	63.53	89.45	54.18	41.16	63.65	292.85	56.27
+ ConFREE	70.79 (↑ 1.92)	63.71 (↑ 0.18)	89.64 (↑ 0.19)	55.44 (↑ 1.26)	41.97 (↑ 0.81)	65.11 (↑ 1.46)	289.32	56.66
FedROD	69.21	63.79	89.93	51.02	39.74	61.02	159.55	54.30
+ ConFREE	69.56 (↑ 0.35)	64.02 (↑ 0.23)	90.80 (↑ 0.87)	52.57 (↑ 1.55)	40.98 (↑ 1.24)	61.52(↑ 0.50)	156.02	54.81
FedRep	68.62	67.41	90.40	51.33	41.72	57.52	151.52	56.90
+ ConFREE	68.92 (↑ 0.30)	67.92 (↑ 0.51)	90.61 (↑ 0.21)	51.41 (↑ 0.08)	42.31 (↑ 0.59)	57.81(↑ 0.29)	147.99	56.97
FedCP	71.38	70.47	91.30	55.33	44.87	60.73	539.17	99.49
+ ConFREE	71.73 (↑ 0.35)	70.98 (↑ 0.51)	91.72 (↑ 0.42)	55.68 (↑ 0.35)	45.62 (↑ 0.75)	61.51(↑ 0.78)	535.64	99.59
FedALA	68.82	67.33	90.67	57.24	44.56	62.92	683.23	55.91
+ ConFREE	70.64 (↑ 1.82)	67.52 (↑ 0.19)	91.32 (↑ 0.65)	57.70 (↑ 0.46)	45.01 (↑ 0.45)	63.94 (↑ 1.02)	679.72	56.39
FedPAC	74.48	72.60	90.86	61.94	46.63	68.56	160.01	110.67
+ ConFREE	74.68 (↑ 0.20)	73.01 (↑ 0.41)	91.42 (↑ 0.56)	62.85 (↑ 0.91)	46.83 (↑ 0.20)	69.10 (↑ 0.54)	156.48	147.20

Table 1: Comparison of eight SOTA methods combined with ConFREE. #S: communication overhead (MB), #T: computation overhead (s).

Methods	CIFAR100				Flowers102			
	C = 20		C = 60		C = 20		C = 60	
	$P = 30\%$	$P = 60\%$	$P = 30\%$	$P = 60\%$	$P = 30\%$	$P = 60\%$	$P = 30\%$	$P = 60\%$
APFL	52.85	53.64	43.56	44.02	62.92	63.94	45.88	50.31
+ ConFREE	53.49 (↑ 0.64)	54.11 (↑ 0.47)	44.90 (↑ 1.34)	44.46 (↑ 0.44)	64.14 (↑ 1.22)	64.48 (↑ 0.54)	49.30 (↑ 3.42)	52.53 (↑ 2.22)
FedROD	53.91	53.06	52.49	50.59	61.41	61.46	56.87	55.28
+ ConFREE	54.18 (↑ 0.27)	53.29 (↑ 0.23)	53.23 (↑ 0.74)	52.29 (↑ 1.70)	62.41 (↑ 1.00)	62.69 (↑ 1.23)	57.74 (↑ 0.87)	56.34 (↑ 1.06)
FedRep	50.95	51.03	44.36	44.04	58.44	57.42	49.40	48.72
+ ConFREE	51.11 (↑ 0.16)	51.39 (↑ 0.36)	44.66 (↑ 0.30)	44.37 (↑ 0.33)	58.93 (↑ 0.49)	58.10 (↑ 0.68)	49.93 (↑ 0.53)	49.16 (↑ 0.44)
FedCP	56.25	56.10	50.20	48.95	62.19	61.31	53.30	51.81
+ ConFREE	56.38 (↑ 0.13)	56.20 (↑ 0.10)	50.29 (↑ 0.09)	49.02 (↑ 0.07)	63.03 (↑ 0.84)	61.80 (↑ 0.49)	53.69 (↑ 0.39)	52.29 (↑ 0.48)
FedALA	56.62	56.97	58.01	58.15	62.19	62.48	56.54	58.71
+ ConFREE	56.98 (↑ 0.36)	57.45 (↑ 0.48)	58.40 (↑ 0.39)	58.60 (↑ 0.45)	63.50 (↑ 1.31)	63.70 (↑ 1.22)	59.33 (↑ 2.79)	60.88 (↑ 2.17)
FedPAC	61.45	61.61	61.93	62.84	67.49	68.08	58.95	59.04
+ ConFREE	62.09 (↑ 0.64)	62.23 (↑ 0.62)	62.63 (↑ 0.70)	63.34 (↑ 0.50)	67.83 (↑ 0.34)	68.86 (↑ 0.78)	60.20 (↑ 1.25)	60.78 (↑ 1.74)

Table 2: Comparison on different numbers of clients C and participation rates P on the CIFAR100 and Flowers102.

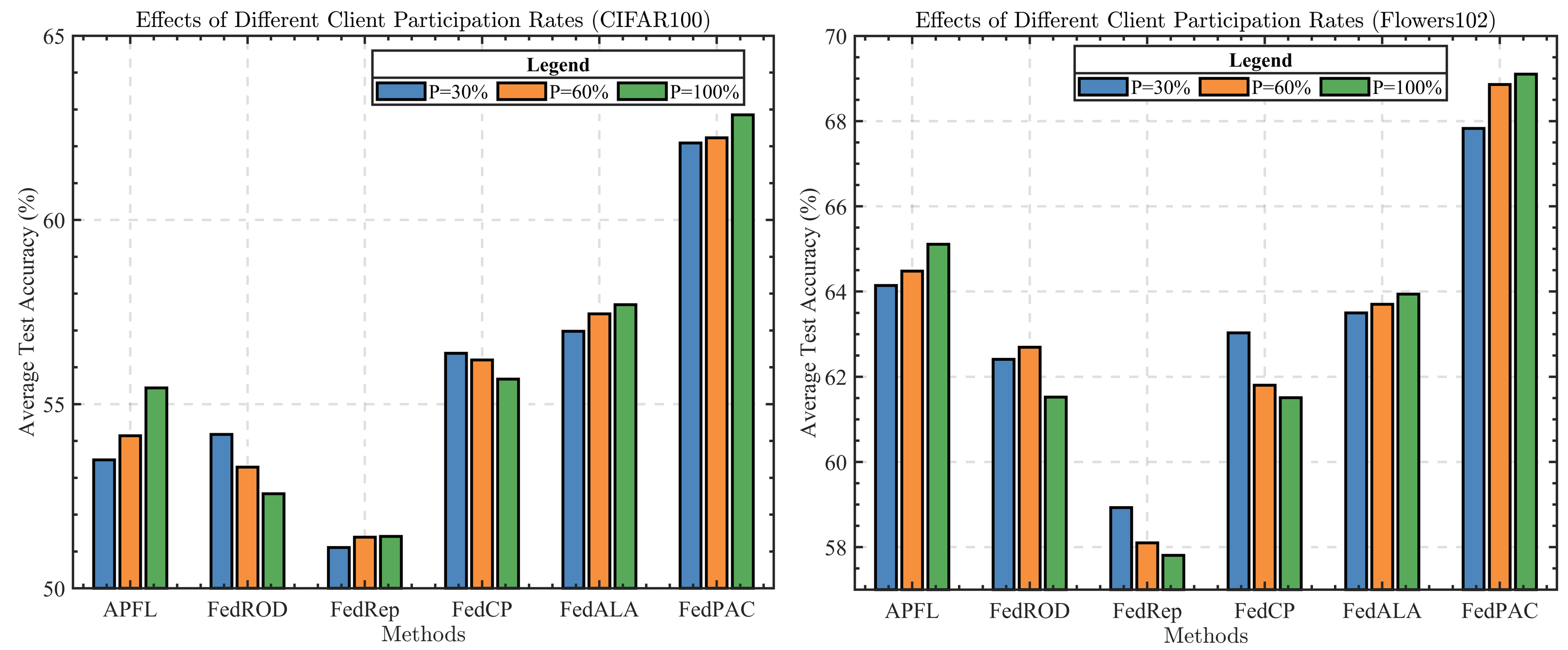


Figure 1: The average test accuracy of different client participation rates for six SOTA models combined with ConFREE on CIFAR100 and Flowers102.

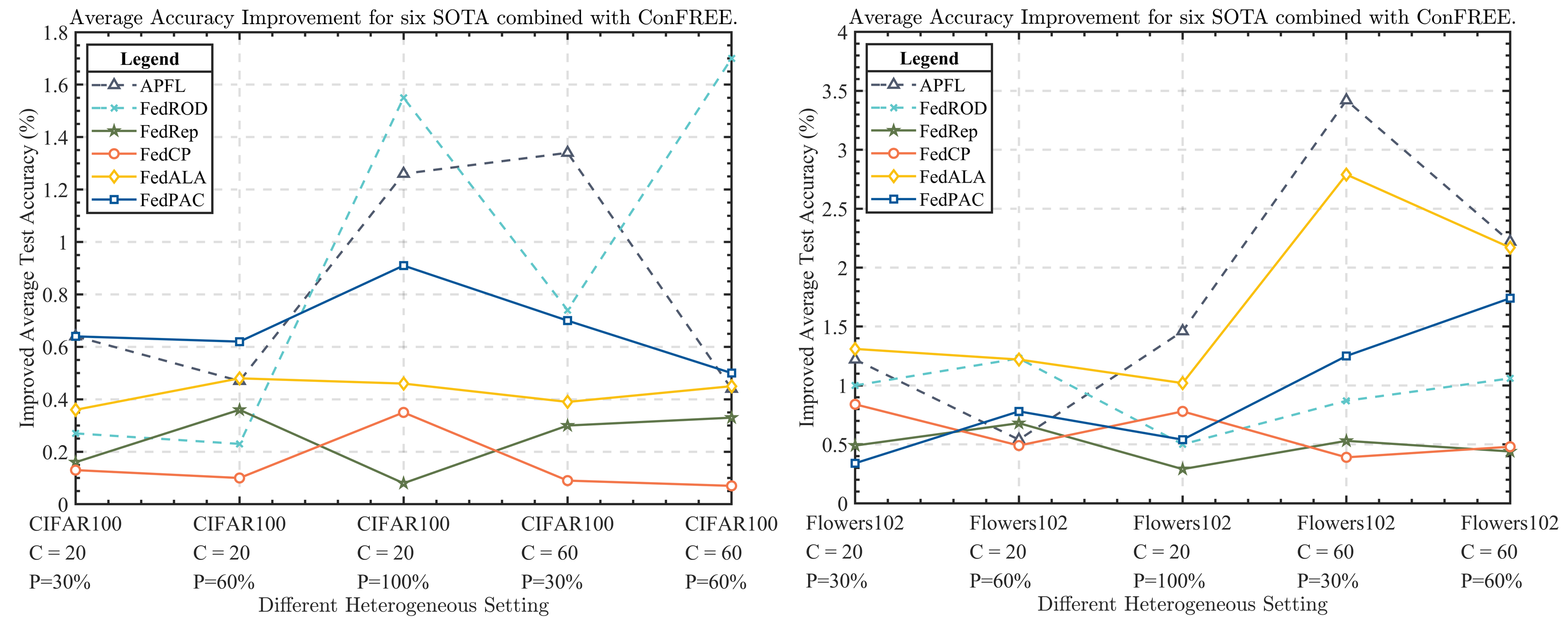


Figure 2: Comparison of six SOTA methods combined with ConFREE under different heterogenous setting on CIFAR100 and Flowers102.