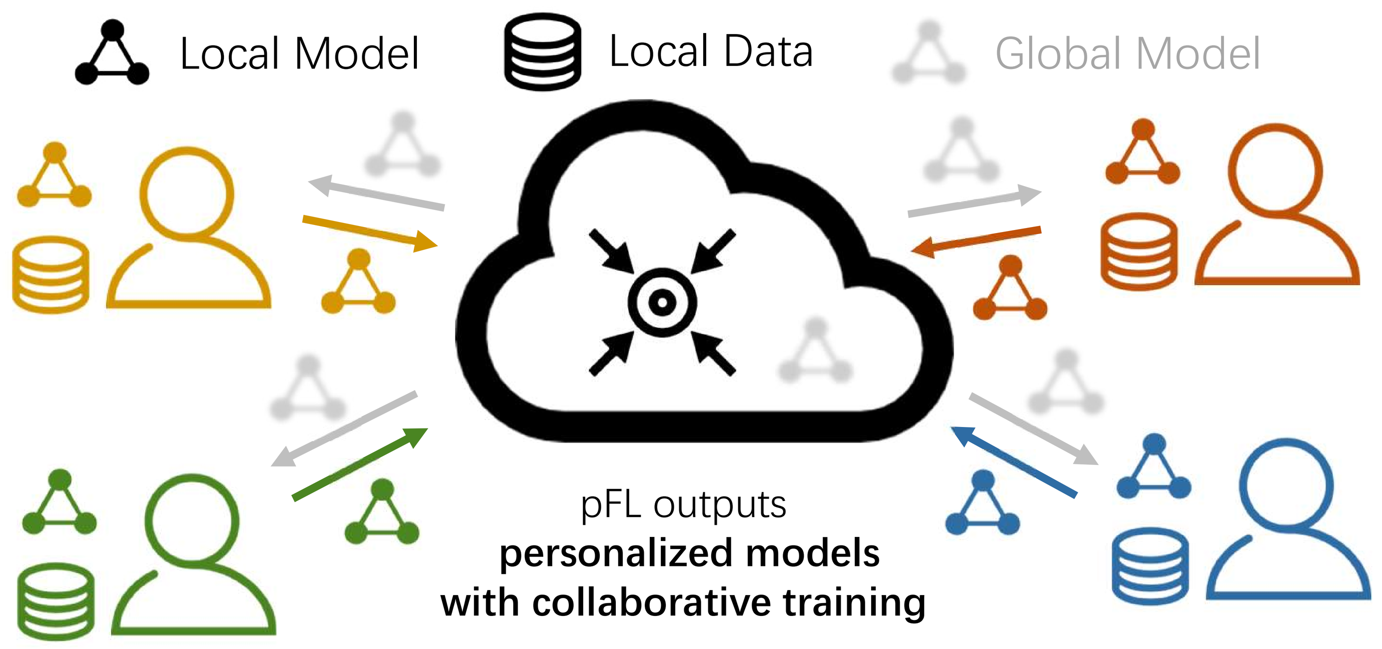
GPFL: Simultaneously Learning Global and Personalized Feature Information for Personalized Federated Learning

Jianqing Zhang[1], Yang Hua[2], Hao Wang[3], Tao Song[1], Zhengui Xue[1], Ruhui Ma[1], Jian Cao[1], Haibing Guan[1]
[1]Shanghai Jiao Tong University [2]Queen's University Belfast [3]Louisiana State University

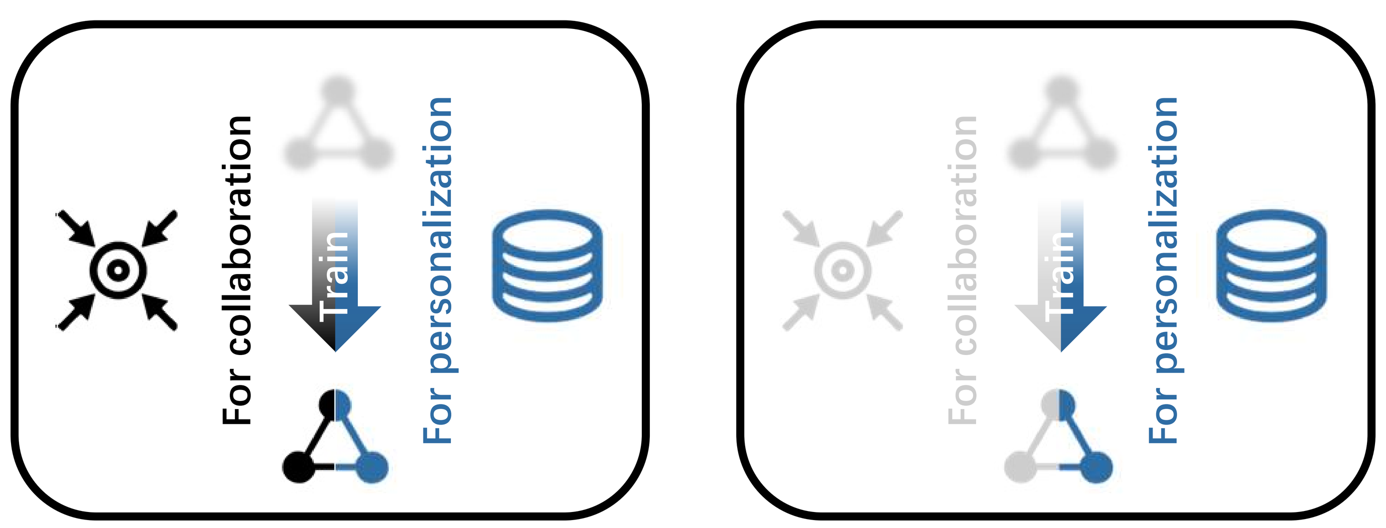
ICCV23 PARIS

Introduction

Background: Federated Learning (FL) has attracted increasing attention for its **collaborative learning** and **privacy-preserving** abilities. Then, personalized FL (pFL) comes along to tackle the statistical heterogeneity issue (as shown by the colorful icons below), which makes the standard FL hard to learn reasonable models. Gray color represents "poor".

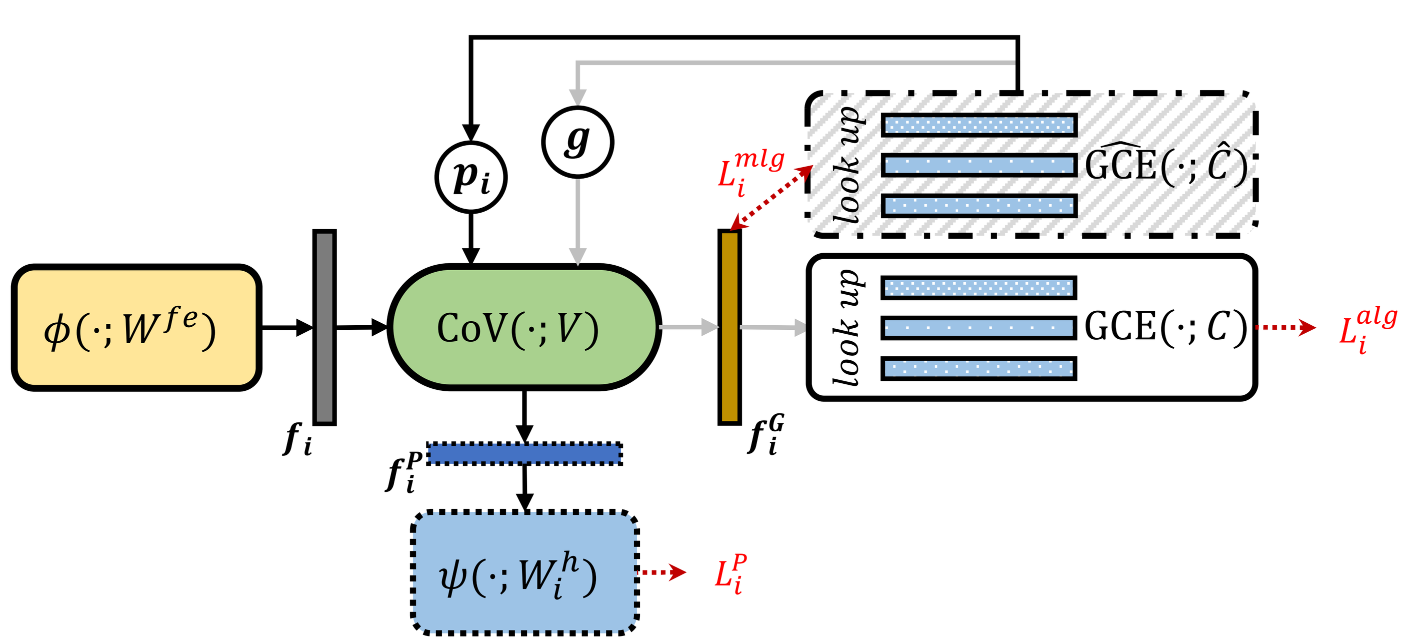


Motivation: An ideal pFL is one kind of FL with two goals: (1) *aggregating information for collaborative learning* and (2) *training reasonable personalized models*. However, local training mainly absorbs personalized information for clients' own tasks. Thus, we propose to **introduce more global information during local training**.



Two goals in idea pFL   Only training on local tasks

Overview: Illustration of client modules and data flow between them. Client i shares W^{fe}, V, C, and $\hat{C}$ while keeping W^h_i locally. By introducing trainable *Global Category Embedding layer* (GCE), we **simultaneously learn global and personalized feature information** during local training.



GPFL outperforms baselines by up to 8.99% on CV, NLP, and IoT tasks.

Core Components

Feature transformation. Using conditional computation techniques, we transform the extracted features f_i to f_i^G and f_i^P through the affine mapping

$$f_i^G = \sigma[(\gamma+1)\odot f_i + \beta], \quad f_i^P = \sigma[(\gamma_i+1)\odot f_i + \beta_i], \tag{1}$$

where γ, β, γ_i, and β_i are generated by *Conditional Valve* (CoV):

$$\{\gamma,\beta\} = \mathrm{CoV}(f_i, g; V), \quad \{\gamma_i,\beta_i\} = \mathrm{CoV}(f_i, p_i; V), \tag{2}$$

where g and p_i are the global and personalized conditional input, respectively.

Angle-level global guidance. To spread out the category embeddings during training for collaboration, we devise the angle-level guidance loss

$$\mathcal{L}_i^{alg} = -\log \frac{\exp(\mathrm{sim}(f_i^G, \mathrm{GCE}(y_i; C)))}{\sum_{u\in[U]} \exp(\mathrm{sim}(f_i^G, \mathrm{GCE}(u; C)))}. \tag{3}$$

Magnitude-level global guidance. To preserve global feature information, we devise the magnitude-level guidance loss

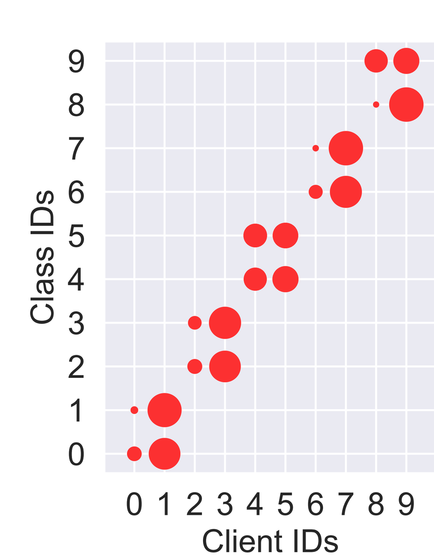
$$\mathcal{L}_i^{mlg} = ||f_i^G, \widehat{\mathrm{GCE}}(y_i;\hat{C})||_2. \tag{4}$$

Personalized tasks. Each client also learns personalized feature information f_i^P through personalized tasks

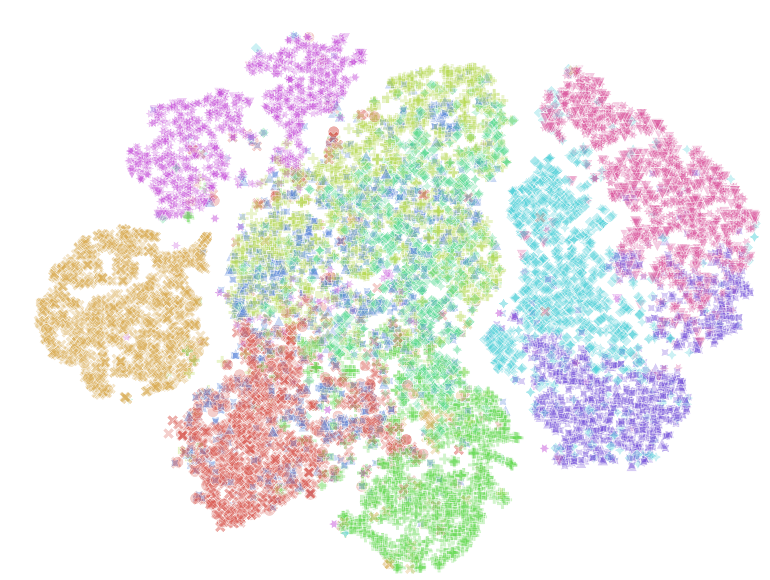
$$\mathcal{L}_i^P = \ell(\psi(f_i^P; W_i^h), y_i). \tag{5}$$

Learned Features

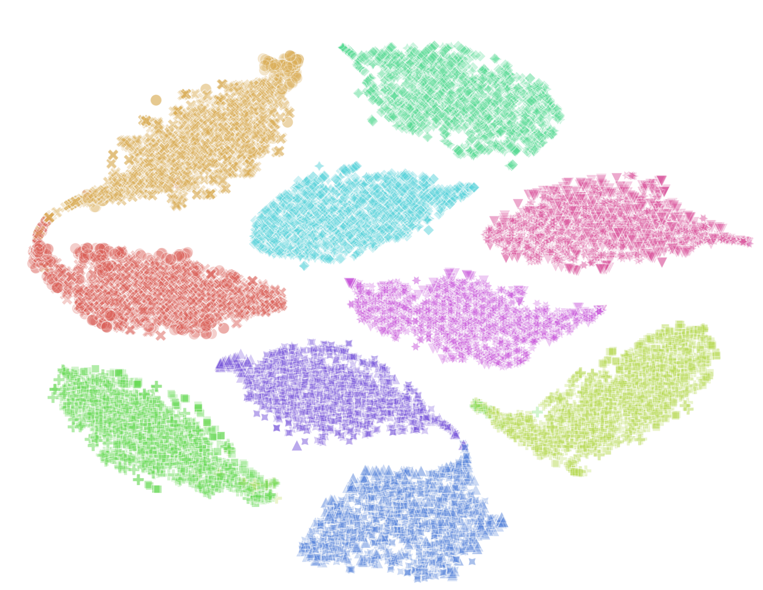
(a): Data distribution on each client. (b), (c), and (d): t-SNE visualizations of feature vectors on FMNIST with 10 clients.



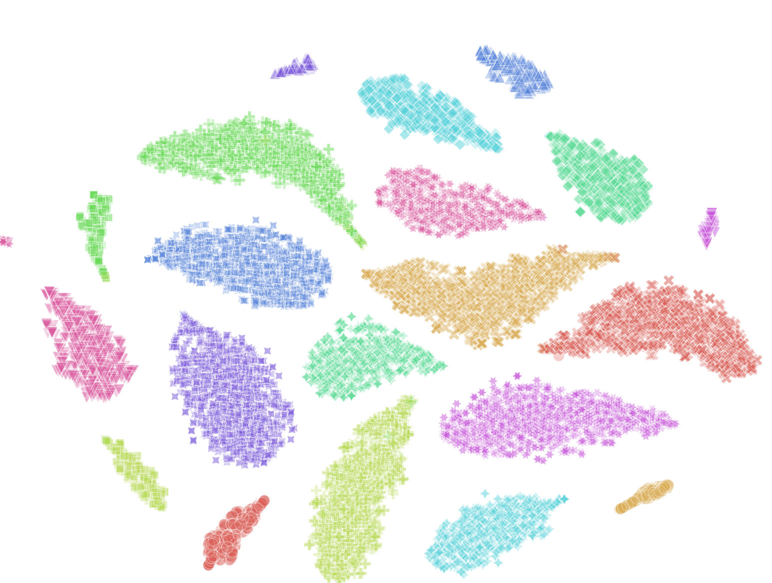
(a) Data distribution



(b) FedPer



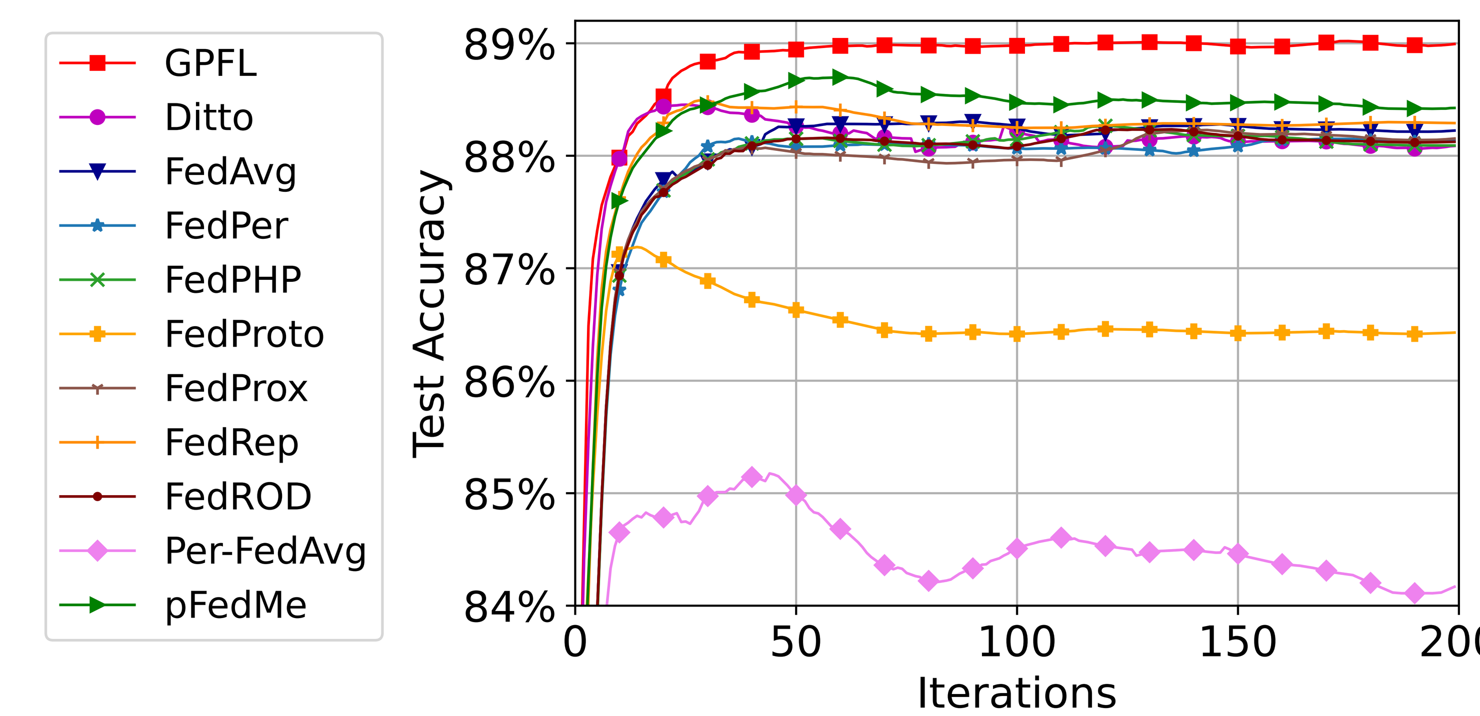
(c) FedProto



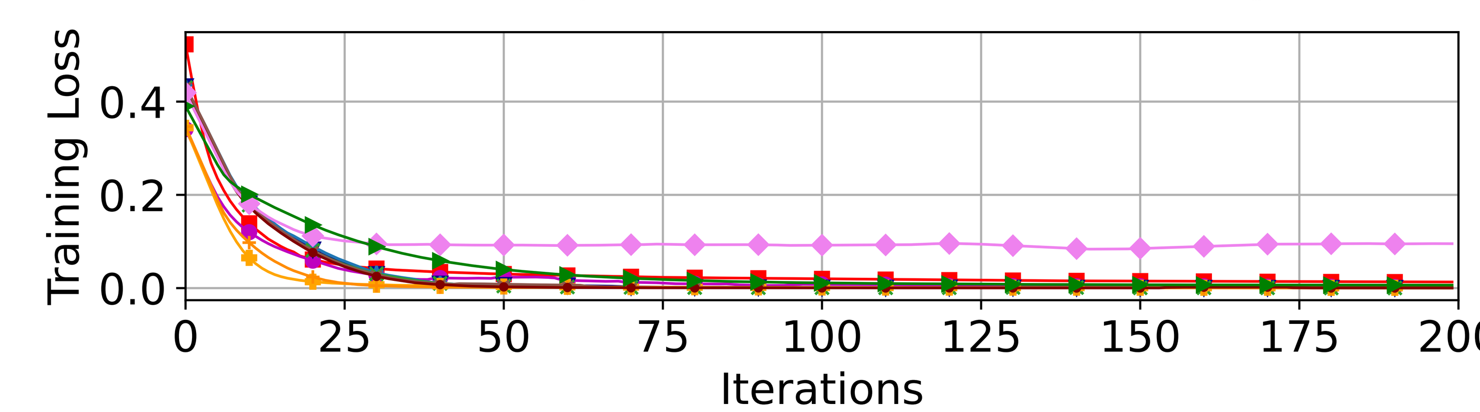
(d) GPFL

Mitigating Overfitting

The global embeddings mitigate the overfitting of client models.



(a) Test accuracy curves in the *feature shift* setting.



(b) Training loss curves in the *feature shift* setting.

Fairness

Standard deviation (%, ↓) [the coefficient of variation ($\times 10^{-2}$, ↓)] of client accuracy in 4 settings. paLS: pathological *label skew*, prLS: practical *label skew*.

Settings	**paLS**	**prLS**	***Feature Shift***	***Real World***
Clients	$N=20$	$N=100$	$N=4$	$N=30$
Datasets	TINY	Cifar100	Amazon Review	HAR
FedAvg	3.57 [25.14]	7.06 [22.10]	1.62 [1.84]	17.10 [19.61]
FedProx	3.51 [25.34]	7.08 [22.15]	1.60 [1.81]	17.35 [19.64]
Per-FedAvg	3.27 [11.65]	8.13 [22.54]	2.82 [3.29]	14.15 [18.35]
pFedMe	3.36 [12.12]	8.19 [17.63]	1.99 [2.25]	12.65 [13.81]
Ditto	3.84 [9.62]	9.89 [18.70]	2.12 [2.40]	13.20 [14.42]
FedPer	3.39 [8.51]	8.91 [22.07]	2.18 [2.47]	19.49 [25.79]
FedRep	3.53 [8.64]	8.99 [20.15]	2.15 [2.43]	21.16 [26.30]
FedRoD	3.46 [9.12]	8.87 [19.01]	2.24 [2.54]	16.93 [18.83]
FedPHP	3.81 [10.28]	9.45 [19.01]	1.96 [2.22]	13.81 [15.70]
FedProto	4.13 [11.23]	9.98 [21.18]	1.82 [2.08]	11.77 [13.89]
GPFL	**3.21 [7.20]**	**8.05 [14.10]**	**1.62 [1.80]**	**8.42 [8.98]**

More details and results (*e.g.*, privacy) can be found on:

- **Full paper:** https://arxiv.org/abs/2308.10279
- **Code:** https://github.com/TsingZ0/GPFL
- **Related Project:** https://github.com/TsingZ0/PFL-Non-IID