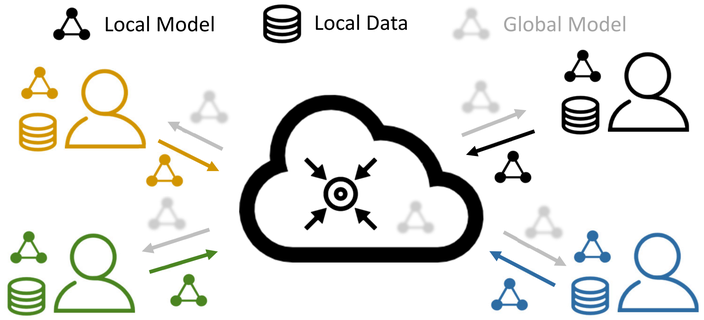
ELIMINATING DOMAIN BIAS FOR FEDERATED LEARNING IN REPRESENTATION SPACE

Jianqing Zhang[1], Yang Hua[2], Jian Cao[1], Hao Wang[3], Tao Song[1], Zhengui Xue[1], Ruhui Ma[1], Haibing Guan[1]

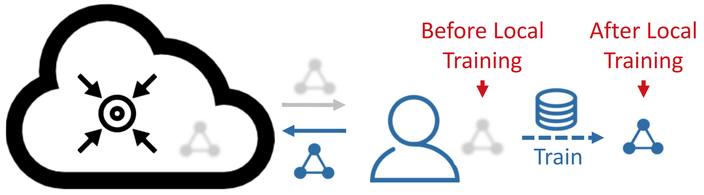
[1]Shanghai Jiao Tong University [2]Queen's University Belfast [3]Louisiana State University

Observation

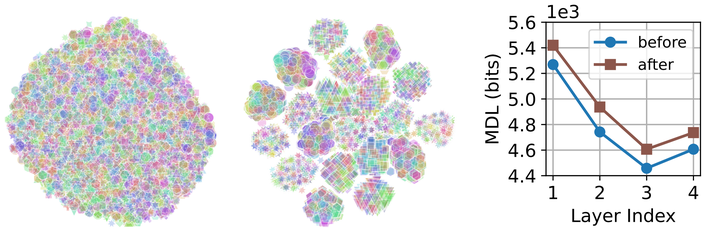
Background: Client-specific private data brings the **statistical heterogeneity** issue (as shown by the colorful icons below) in Federated Learning (FL). The poor global model hinders knowledge transfer.



With heterogeneous data, clients' local training turns the received global model into client-specific local models.

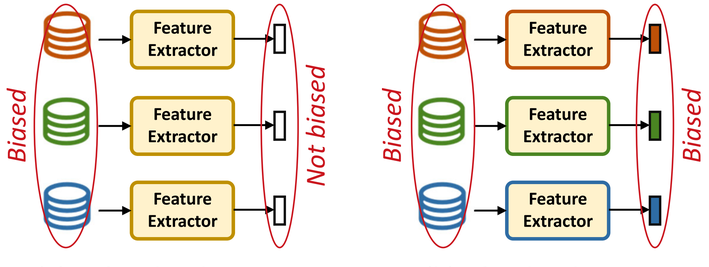


Representation bias and degeneration phenomenon: After local training, the feature representations are ***biased*** to client-specific domains. At the same time, representations' quality is also **degenerated** per-layer.



(a) Before. (b) After. (c) MDL change.

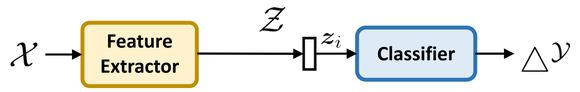
Representation bias and **representation degeneration** also exist in personalized FL (pFL).



(a) Before local training (b) After local training

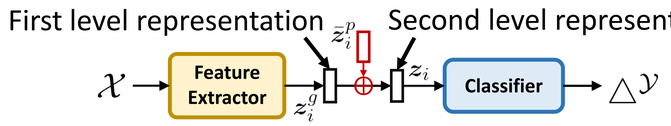
Our DBE improves FL methods by -22.35% in MDL and +32.30 in accuracy (%). FedAvg+DBE outperforms pFL methods by +11.36 in accuracy.

Domain Bias Eliminator (DBE)



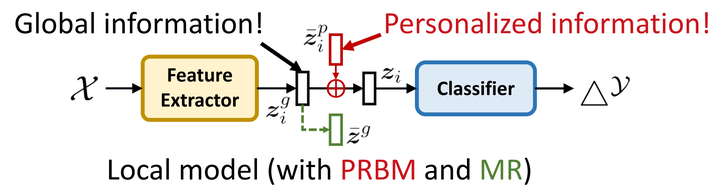
Local model (original)

Local loss (original): $\mathcal{L}_{\mathcal{D}_i}(\theta) := \mathbb{E}_{(x_i,y_i)\sim\mathcal{D}_i}[\ell(h(f(x_i;\theta^f);\theta^h),y_i)]$.



Local model (with PRBM)

PRBM stores personalized (***biased***) representation information (z_i^p) for each client, and makes the remaining information (z_i^g) to be global.
Local loss (with PRBM): $\mathcal{L}_{\mathcal{D}_i}(\theta_i) := \mathbb{E}_{(x_i,y_i)\sim\mathcal{D}_i}[\ell(h(f(x_i;\theta^f)+z_i^p;\theta^h),y_i)]$.
View the PRBM as a personalized translation transformation: $\texttt{PRBM}: \mathcal{Z}\mapsto\mathcal{Z}$.
Then, local loss (with PRBM): $\mathcal{L}_{\mathcal{D}_i}(\theta_i) := \mathbb{E}_{(x_i,y_i)\sim\mathcal{D}_i}[\ell(h(\texttt{PRBM}(f(x_i;\theta^f);z_i^p);\theta^h),y_i)]$.

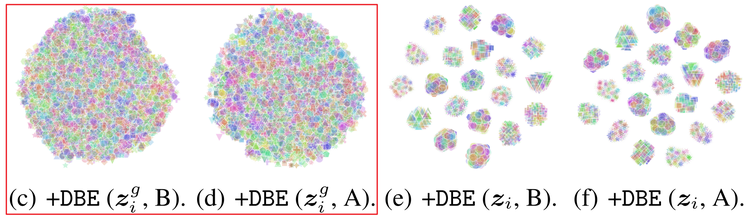


Local model (with PRBM and MR)

MR explicitly guides the feature extractor to generate z_i^g with global information.
Final local loss (with PRBM and MR):
$\mathcal{L}_{\mathcal{D}_i}(\theta_i) := \mathbb{E}_{(x_i,y_i)\sim\mathcal{D}_i}[\ell(h(\texttt{PRBM}(f(x_i;\theta^f);z_i^p);\theta^h),y_i)] + \kappa\cdot\texttt{MR}(z_i^g,\bar{z}^g)$.
$\bar{z}^g$ (**client-invariant mean**) is a global consensus obtained before FL.
DBE improves bi-directional knowledge transfer (Corollary 1 and Corollary 2).

Eliminating Representation Bias

t-SNE visualization for representations on Tiny-ImageNet (200 labels). "B" and "A" denote "before local training" and "after local training", respectively.



(c) +DBE (z_i^g, B). (d) +DBE (z_i^g, A). (e) +DBE (z_i, B). (f) +DBE (z_i, A).

The first level representation z_i^g in FedAvg+DBE is ***no longer biased*** to the local data domain of each client after local training. With the personalized translation transformation PRBM, z_i can fit the local domain of each client either before or after local training.

Improving Traditional FL with DBE

TINY and TINY* represent using 4-layer CNN and ResNet-18 on Tiny-ImageNet, respectively. DBE **promotes traditional FL methods by at most -22.35% in MDL (bits) and +32.30 in accuracy (%).**

Metrics	MDL				Accuracy			
Datasets	Cifar100	TINY	TINY*	AG News	Cifar100	TINY	TINY*	AG News
SCAFFOLD	1499	3661	3394	1931	33.08	23.26	24.90	88.13
FedProx	1523	3701	3570	2092	31.99	19.37	19.27	87.21
MOON	1516	3696	3536	1836	32.37	19.68	19.02	84.14
FedGen	1506	3675	3551	1414	30.96	19.39	18.53	89.86
SCAFFOLD+DBE	**1434**	**3549**	**3370**	**1743**	**63.61**	**45.55**	**45.09**	**96.73**
FedProx+DBE	**1439**	**3587**	**3490**	**1689**	**63.22**	**42.28**	**41.45**	**96.62**
MOON+DBE	**1432**	**3580**	**3461**	**1683**	**63.26**	**43.43**	**41.10**	**96.68**
FedGen+DBE	**1426**	**3563**	**3488**	**1098**	**63.26**	**42.54**	**41.87**	**97.16**

Outperforming pFL with FedAvg+DBE

Cifar100† represents the experiment with 100 clients and joining ratio $\rho = 0.5$ on Cifar100. **FedAvg+DBE outperforms the best SOTA pFL methods by up to +11.36 on Cifar100†** including the fine-tuning-based methods.

	Accuracy in Practical setting						Overhead	
	FMNIST	Cifar100	Cifar100†	TINY	TINY*	AG News	Total time	Time/iteration
Per-FedAvg	95.10	44.28	38.28	25.07	21.81	87.08	121 min	3.56 min
pFedMe	97.25	47.34	31.13	26.93	33.44	87.08	1157 min	10.24 min
Ditto	97.47	52.87	39.01	32.15	35.92	91.89	318 min	11.78 min
FedPer	97.44	49.63	41.21	33.84	38.45	91.85	83 min	1.92 min
FedRep	97.56	52.39	41.51	37.27	39.95	92.25	471 min	4.09 min
FedRoD	97.52	50.94	48.56	36.43	37.99	92.16	87 min	1.74 min
FedBABU	97.46	55.02	52.07	36.82	34.50	95.86	811 min	1.58 min
APFL	97.25	46.74	39.47	34.86	35.81	89.37	156 min	2.74 min
FedFomo	97.21	45.39	37.59	26.33	26.84	91.20	193 min	2.72 min
APPLE	97.06	53.22	—	35.04	39.93	84.10	132 min	2.93 min
FedAvg	85.85	31.89	28.81	19.46	19.45	87.12	365 min	1.59 min
FedAvg+DBE	**97.69**	**64.39**	**63.43**	**43.32**	**42.98**	**96.87**	171 min	1.60 min

Improving pFL with MR

MR can promote the local-to-global knowledge transfer between server and client. Therefore, pFL methods can benefit more from a better global model achieving higher accuracy on Tiny-ImageNet. However, their MR-equipped variants perform worse than FedAvg+DBE since **the representation bias still exists without using** PRBM.

+MR	Ditto	FedPer	FedRep	FedRoD	FedBABU	APFL	FedFomo
Accuracy	42.82	41.78	41.28	42.74	38.17	39.22	29.51
Improvement	10.67	7.94	4.01	6.31	1.35	4.36	3.18

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

- **Full paper:** https://arxiv.org/abs/2311.14975
- **Code:** https://github.com/TsingZ0/DBE
- **Related Project:** https://github.com/TsingZ0/PFL-Non-IID