SCIENCE CHINA
Information Sciences

• LETTER •

November 2024, Vol. 67, Iss. 11, 219302:1–219302:2
https://doi.org/10.1007/s11432-024-4160-3

Empowering over-the-air personalized federated learning via RIS

Wei SHI[1,2], Jiacheng YAO[1,2], Jindan XU[3*], Wei XU[1,2*], Lexi XU[4] & Chunming ZHAO[1,2]

[1] *National Mobile Communications Research Laboratory, Southeast University, Nanjing 210096, China;*
[2] *Purple Mountain Laboratories, Nanjing 211111, China;*
[3] *School of Electrical and Electronics Engineering, Nanyang Technological University, Singapore 639798, Singapore;*
[4] *Research Institute, China United Network Communications Corporation, Beijing 100048, China*

Received 24 June 2024/Revised 26 August 2024/Accepted 14 September 2024/Published online 10 October 2024

Federated learning (FL) is a promising distributed learning approach due to its privacy-enhancing characteristic [1–3]. To enhance communication efficiency of FL, over-the-air computation (AirComp) has emerged as a key technique by exploiting the waveform superposition property of multiple access channels [4, 5]. Although AirComp-enabled FL (AirFL) offers significant performance gains, it does not address the data heterogeneity in most real-life FL scenarios with non-independent and identically distributed local datasets. Such data heterogeneity hinders the generalization of a single global consensus model. To this end, preliminary works have been made to develop a personalized AirFL framework via clustering algorithms, where different models are trained for different clusters under the orchestration of the parameter server (PS) [6, 7]. However, this personalized framework requires large-scale antennas to combat interference, leading to a significant escalation in hardware cost.

As a cost-effective physical-layer technology, reconfigurable intelligent surface (RIS) has been extensively studied to support various communication applications due to its capability for smart channel reconstruction [8]. In this paper, we introduce low-cost RIS to achieve statistical interference elimination across different clusters and facilitate simultaneous multi-cluster computation over-the-air, thereby enhancing the efficiency of personalized AirFL.

Learning and communication model. We consider a personalized AirFL system consisting of K distributed devices, which are partitioned into M ($M < K$) disjoint clusters $\mathcal{K}_1, \ldots, \mathcal{K}_M$. A specific clustering method can be found in [6], which is not the focus of this article. Our goal is to find the optimal personalized model parameters $\mathbf{w}_m \in \mathbb{R}^D$ for each cluster $m \in [M]$ to minimize the loss function $\mathcal{L}_m(\mathbf{w}_m) = \frac{1}{|\mathcal{K}_m|}\sum_{k\in\mathcal{K}_m} F_k(\mathbf{w}_m, \mathcal{D}_k)$, where $F_k(\cdot, \mathcal{D}_k)$ is the loss function of device k with local dataset $\mathcal{D}_k$.

Distributed stochastic gradient descent (SGD) is adopted to optimize $\mathbf{w}_m$ in an iterative manner. First, at each training round t, the PS broadcasts the latest personalized models $\{\mathbf{w}_{m,t}\}_{m\in[M]}$ to each device. Then, based on the clustering mechanism, each device $k \in \mathcal{K}_m$ computes its local gradient $\mathbf{g}_{m,t,k} \in \mathbb{R}^D$ based on $\mathbf{w}_m$ and its local dataset $\mathcal{D}_k$, and reports it to the PS. Finally, after receiving all the local gradients, the PS calculates the global gradient of cluster m as

$$\mathbf{g}_{m,t} = \frac{1}{|\mathcal{K}_m|}\sum_{k\in\mathcal{K}_m} \mathbf{g}_{m,t,k}, \tag{1}$$

and updates the personalized model for cluster m through $\mathbf{w}_{m,t+1} = \mathbf{w}_{m,t} - \eta_{m,t}\mathbf{g}_{m,t}$, where $\eta_{m,t}$ is a chosen learning rate at the t-th training round. The above steps iterate until a convergence condition is met.

Note that the operation in (1) requires the PS to sum the local gradients of devices in each cluster separately. By applying AirComp, all devices simultaneously upload the analog signals of local gradients to the PS, achieving summation over-the-air. However, the analog nature of AirFL makes the PS cannot distinguish between the gradients of different clusters. In the following, we introduce an RIS-enabled personalized AirFL framework to address this challenge. Each cluster is assisted by an RIS with N reflecting elements to help realize the personalized model aggregation. To support simultaneous multi-cluster gradient estimation, at least M receiving antennas are required. Without loss of generality, we consider a PS equipped with M receiving antennas. Then, the received signal at the PS in the t-th round, $\mathbf{Y}_t = [\mathbf{y}_{1,t}, \mathbf{y}_{2,t}, \cdots, \mathbf{y}_{M,t}]^H \in \mathbb{C}^{M\times D}$, is given by

$$\mathbf{Y}_t = \sum_{k=1}^{K}\sqrt{p_k}\left(\sum_{i=1}^{M}\beta_{i,k}\mathbf{H}_{p,i}^H\boldsymbol{\Theta}_i\mathbf{h}_{i,k}\right)\bar{\mathbf{g}}_{m,t,k}^H + \mathbf{Z}_t, \tag{2}$$

where $\bar{\mathbf{g}}_{m,t,k} \triangleq \frac{1}{\sigma_{m,t,k}}(\mathbf{g}_{m,t,k} - u_{m,t,k}\mathbf{1})$ represents the normalized gradient, $u_{m,t,k}$ and $\sigma_{m,t,k}$ denote the mean and standard deviation of all entries in $\mathbf{g}_{m,t,k}$, p_k is the transmit power of device k, $\beta_{i,k}$ is the cascaded large-scale fading coefficient from device k to the PS through the i-th RIS, $\mathbf{H}_{p,i} = [\mathbf{h}_{p,i,1}, \mathbf{h}_{p,i,2}, \cdots, \mathbf{h}_{p,i,M}] \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_N \otimes \mathbf{I}_M)$ and $\mathbf{h}_{i,k} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_N)$ denote the small-scale fading channel from the i-th RIS to the PS and device k to the i-th RIS, respectively, $\mathbf{Z}_t = [\mathbf{z}_{1,t}, \mathbf{z}_{2,t}, \cdots, \mathbf{z}_{M,t}]^H$ is additive white Gaussian noise whose entries follow $\mathcal{CN}(0, \sigma^2)$, $\boldsymbol{\Theta}_i \triangleq \mathrm{diag}\left\{e^{j\theta_{i,1}}, \ldots, e^{j\theta_{i,n}}, \ldots, e^{j\theta_{i,N}}\right\}$ is the reflection matrix of the i-th RIS, and $\theta_{i,n} \in [0, 2\pi)$ is the phase shift introduced by the n-th RIS reflecting element. Then, based

* Corresponding author (email: jindan.xu@ntu.edu.sg, wxu@seu.edu.cn)

on the signal $\mathbf{y}_{m,t}$ at the m-th receiving antenna, the PS computes an estimated global gradient of cluster m as $\hat{\mathbf{g}}_{m,t} = \frac{\Re\{\mathbf{y}_{m,t}\}}{\lambda_m} + \sum_{k\in\mathcal{K}_m}\frac{u_{m,t,k}}{|\mathcal{K}_m|}\mathbf{1}$, where $\lambda_m > 0$ is a denoising factor introduced by the PS. It is rewritten as

$$\hat{\mathbf{g}}_{m,t} = \sum_{k\in\mathcal{K}_m}\ell_{m,k}\bar{\mathbf{g}}_{m,t,k} + \sum_{k\in\mathcal{K}_m}\frac{u_{m,t,k}}{|\mathcal{K}_m|}\mathbf{1} + \sum_{\substack{1\leqslant m'\leqslant M\\ m'\neq m}}\sum_{k'\in\mathcal{K}_{m'}}\ell_{m,k'}\bar{\mathbf{g}}_{m',t,k'} + \bar{\mathbf{z}}_{m,t}, \tag{3}$$

where $\ell_{m,k} = \frac{\sqrt{p_k}}{\lambda_m}\sum_{i=1}^{M}\beta_{i,k}\Re\left\{\mathbf{h}_{p,i,m}^H\mathbf{\Theta}_i\mathbf{h}_{i,k}\right\}$, $\forall m,k$, and $\bar{\mathbf{z}}_{m,t} \triangleq \frac{\Re\{\mathbf{z}_{m,t}\}}{\lambda_m}$ is the equivalent noise. Note that the estimated gradient is interfered by signals from other clusters and these interference cannot be eliminated since $M < K$. To this end, we propose an RIS phase shift configuration scheme that fortunately eliminates the interference from a statistical perspective in the following theorem.

Theorem 1. Statistical interference elimination, i.e., $\mathbb{E}[\ell_{m,k}] > 0$, $\forall k\in\mathcal{K}_m$ and $\mathbb{E}[\ell_{m,k'}] = 0$, $\forall k'\notin\mathcal{K}_m$, can be achieved by setting

$$\theta_{m,n} = -\angle h_{p,m,m,n}^* + \angle\sum_{k\in\mathcal{K}_m}h_{m,k,n}^*, \tag{4}$$

for $m\in[M]$ and $n\in[N]$, where $h_{p,m,m,n}$ and $h_{m,k,n}$ are the n-th elements of channel vectors $\mathbf{h}_{p,m,m}$ and $\mathbf{h}_{m,k}$, respectively.

Proof. See Appendix A.

According to **Theorem 1**, we conclude that favorable propagation can be achieved through phase matching using low-cost RIS reflecting elements, thereby eliminating the need for expensive large-scale receiving antennas. After the statistical interference elimination, we focus on joint design of power control and denoising factors from the following two perspectives to enhance the FL convergence.

1) Unbiased design: From the perspective of first-order moment, ensuring unbiased gradient estimation is of pivotal significance for guaranteeing FL convergence [4]. Hence, we consider the following unbiasedness-oriented method.

Proposition 1. By setting $p_k = \sigma_{m,t,k}^2\beta_{m,k}^{-2}\zeta_m^2$, $\forall k\in\mathcal{K}_m$, and $\lambda_m = \frac{\pi N\sqrt{|\mathcal{K}_m|}\zeta_m}{4}$, the gradient estimation in (3) is unbiased, where $\zeta_m = \min_{k\in\mathcal{K}_m}\frac{\sqrt{P_k}\beta_{m,k}}{\sigma_{m,t,k}\sqrt{D}}$ and P_k is the maximum transmit power.

Proof. See Appendix B.

2) Minimum mean squared error (MMSE) design: Apart from unbiasedness of the first-order moment, the second-order moment, known as MSE, also plays a decisive role in FL convergence [9]. For any given power control of p_k, we derive the optimal denoising factors in closed form in the following proposition.

Proposition 2. The optimal denoising factor of cluster m for minimizing MSE is equal to

$$\lambda_m^* = |\mathcal{K}_m|\frac{\sum_{i=1}^{M}\sum_{k\in\mathcal{K}_i}p_k\bar{h}_{m,k}^2\sigma_{i,t,k}^2 + \frac{\sigma^2}{2}}{\sum_{k\in\mathcal{K}_m}\sqrt{p_k}\bar{h}_{m,k}\sigma_{m,t,k}^3}, \tag{5}$$

where $\bar{h}_{m,k} \triangleq \sum_{i=1}^{M}\beta_{i,k}\Re\left\{\mathbf{h}_{p,i,m}^H\mathbf{\Theta}_i\mathbf{h}_{i,k}\right\}$.

Proof. See Appendix C.

Substituting the optimal λ_m^*, we formulate a power control optimization problem for minimizing the sum MSE, which can be solved via typical optimization methods and please refer to Appendix C for details.

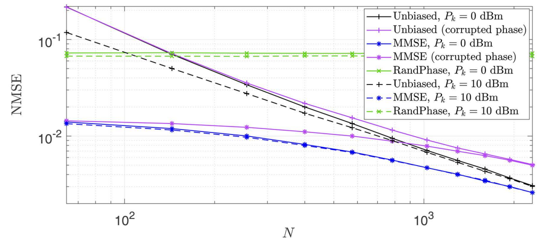

Figure 1 (Color online) NMSE versus N with $K = 100$ and $M = 5$.

The detailed implementation of the proposed RIS-based personalized AirFL approach is summarized in Appendix D.

Numerical Results. Assuming that the distances between the PS and each RIS are 200 m, and all the devices in each cluster $m\in[M]$ are uniformly distributed within a disk of radius 300 m centered at the m-th RIS. The path loss exponent for all the links is 2.2. Figure. 1 depicts the normalized MSE (NMSE) as a function of N for different values of P_k. It is observed that as P_k increases, the improvement in NMSE performance is marginal. This is due to the fact that an increase in transmit power amplifies not only the useful signals but interference. Furthermore, the NMSE of our proposed designs decreases linearly with large N on a log-log scale. This phenomenon becomes more obvious as P_k increases, owing to the diminishing impact of noise error. In addition, the NMSE curves with corrupted RIS phase shifts (i.e., 1-bit phase noise) are presented to validate the robustness of our proposed designs. The baseline utilizing random RIS phase shifts fails to obtain any effective performance enhancements as N and P_k increases, which demonstrates the importance of RIS phase shift configuration in **Theorem 1**.

Acknowledgements This work was supported by National Natural Science Foundation of China (Grant No. 62271137) and Fundamental Research Funds for the Central Universities (Grant Nos. 2242022k60002, 2242023K5003).

Supporting information Appendixes A–D. The supporting information is available online at info.scichina.com and link.springer.com. The supporting materials are published as submitted, without typesetting or editing. The responsibility for scientific accuracy and content remains entirely with the authors.

References

1 Xu W, Huang Y M, Wang W, et al. Toward ubiquitous and intelligent 6G networks: from architecture to technology. Sci China Inf Sci, 2023, 66: 130300
2 Xu W, Yang Z H, Ng D W K, et al. Edge learning for B5G networks with distributed signal processing: semantic communication, edge computing, and wireless sensing. IEEE J Sel Top Signal Process, 2023, 17: 9–39
3 Zhu G X, Lyu Z H, Jiao X, et al. Pushing AI to wireless network edge: an overview on integrated sensing, communication, and computation towards 6G. Sci China Inf Sci, 2023, 66: 130301
4 Yao J C, Xu W, Yang Z H, et al. Wireless federated learning over resource-constrained networks: digital versus analog transmissions. IEEE Trans Wireless Commun, 2024. doi: 10.1109/TWC.2024.3407822
5 Yao J C, Yang Z H, Xu W, et al. Imperfect CSI: a key factor of uncertainty to over-the-air federated learning. IEEE Wireless Commun Lett, 2023, 12: 2273–2277
6 Ghosh A, Chung J, Yin D, et al. An efficient framework for clustered federated learning. IEEE Trans Inform Theor, 2022, 68: 8076–8091
7 Sami H U, Güler B. Over-the-air personalized federated learning. In: Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Singapore, 2022. 1–5
8 Shi W, Xu W, You X H, et al. Intelligent reflection enabling technologies for integrated and green Internet-of-Everything beyond 5G: communication, sensing, and security. IEEE Wireless Commun, 2023, 30: 147–154
9 Zhang N F, Tao M X. Gradient statistics aware power control for over-the-air federated learning. IEEE Trans Wireless Commun, 2021, 20: 5115–5128