

智能反射面赋能的联邦边缘学习及其在车联网中的应用

王平^{1,2}, 杨志伟¹, 李贺举¹

(1. 同济大学电子与信息工程学院, 上海 201804; 2. 上海自主智能无人系统科学中心, 上海 201210)

摘要: 针对无线链路和数据分布的异构性导致在 FEEL 训练中很难实现无线通信和模型精度的最佳权衡的问题, 提出了一种智能反射面 (RIS) 赋能的空中联邦边缘学习系统, 其利用智能反射面的信道可重构性自适应地配置信号传播环境, 并利用空中计算实现联邦边缘学习模型的快速聚合。具体来说, 首先刻画无线信道和数据异构影响下的联邦优化算法收敛行为, 并以此构造统一的无线资源优化问题, 通过联合设计收发端波束成形方案和 RIS 相移来优化学习性能。仿真结果验证了所提方案的有效性, 并证明 RIS 可以在数据异构前提下提高空中联邦边缘学习系统准确性。最后, 探讨了其在车联网中应用的可能性。

关键词: 联邦边缘学习; 数据异构; 智能反射面; 空中计算; 收敛分析; 车联网

中图分类号: TN92

文献标志码: A

DOI: 10.11959/j.issn.1000-436x.2023192

Federated edge learning with reconfigurable intelligent surface and its application in Internet of vehicles

WANG Ping^{1,2}, YANG Zhiwei¹, LI Heju¹

1. College of Electrical and Information Engineering, Tongji University, Shanghai 201804, China

2. Shanghai Research Institute for Intelligent Autonomous Systems, Shanghai 201210, China

Abstract: Aiming at the problem that it is difficult to achieve an optimal trade-off between wireless communication and model accuracy in FEEL training caused by the heterogeneity of wireless links and data distribution, a reconfigurable intelligent surface (RIS) enabled federated edge learning system was proposed, which exploited the channel reconfigurability of RIS to adaptively manipulate the signal propagation environment, and utilized the over-the-air computation (Air-comp) to achieve fast model aggregation. Specifically, the convergence behavior of the FEEL algorithm under the influence of wireless channels and data heterogeneity was rigorously derived, and accordingly, an unified wireless resources optimization problem was constructed with the goal of minimizing the learning loss by jointly designing the transceiver design and the RIS phase shift. Simulation results demonstrate that the proposed scheme achieves substantial performance improvement compared against several baselines, and prove that RIS can play an important role in improving the accuracy of Aircomp enabled FEEL systems under data heterogeneity. Finally, the probability of applying it into Internet of vehicles is discussed.

Keywords: federated edge learning, data heterogeneity, reconfigurable intelligent surface, over-the-air computation, convergence analysis, Internet of vehicles

0 引言

作为物联网中一个有潜力的研究分支, 车联网

可有效改善道路安全问题和驾驶者的驾乘环境, 有望成为智能交通系统中重要的数据传输与控制平台。同时, 随着下一代无线网络的蓬勃发展, 以机

收稿日期: 2023-06-19; 修回日期: 2023-09-10

通信作者: 杨志伟, yx-yangzw@dfmc.com.cn

基金项目: 上海市科学技术委员会基金资助项目 (No.21511102400)

Foundation Item: The Project of Shanghai Municipal Science and Technology Commission (No.21511102400)

器学习为核心的应用方案在分析车联网边缘数据方面发挥着重要的作用, 可进行自主化的推理、建模以及决策^[1]。然而, 由于无线资源限制、时延约束及数据隐私安全方面的挑战, 强迫车联网移动终端将其收集的本地隐私数据上传到中心节点(如云服务器进行集中模型训练)是不切实际的。为此, 联邦边缘学习(FEEL, federated edge learning)通过允许分布式终端设备在数据本地化的前提下共享本地模型协同训练, 避免敏感数据通过无线信道传输, 从而有效地保证了车联网终端隐私和数据安全^[2]。此外, 中心节点从云服务器下沉到网络边缘, 使移动终端只需与基站(BS, base station)端通信, 由此显著降低通信成本, 克服由网络拥塞导致的过度传播时延问题^[3]。然而, 由于采样对象或采样终端规格存在差异, 移动终端数据往往是非独立同分布(non-IID, not identically and independently distributed)的, 即各终端的本地数据往往不服从同一分布, 表现出较大的差异性^[4]。当数据是独立同分布(IID, identically and independently distributed), 即局部分布相同时, 局部损失函数在大量样本上收敛于相同的期望损失, 而 non-IID 数据则会影响全局模型的预测准确率, 阻碍全局模型收敛, 从而使联邦学习任务不能取得预期效果。

同时, 由于移动终端通过无线信道连接到基站, 基站端接收到的模型参数不可避免地受到信道衰落和加性噪声的影响^[5]。由于终端间的无线环境不同, 其与基站端的通信链路也存在较大的异构性。因此, 在模型更新过程中设计有效的上行通信系统是 FEEL 训练的关键问题。针对上行通信设计, 目前已有研究主要利用正交信道来实现本地模型的并发上传^[3,6], 以尽可能地避免终端干扰, 然而, 当边缘终端数量较多时, 此种通信方式通信效率低下, 且所需带宽资源较多。针对此问题, 有工作设计了基于空中计算(Aircomp, over-the-air computation)的 FEEL 无线聚合方案, 以提升频谱效率。然而, 绝大多数工作侧重于 IID 设定下的空中计算分析与设计, 其并未统筹考虑通信及 non-IID 数据分布对空中联邦学习性能的影响。同时, 由于无线环境的不可控性, 如何设计高效的空中聚合方案, 并统筹规划无线资源分配方案, 从而在有限资源约束下最大化联邦性能收益, 仍然是一个亟待解决的问题。

近年来, 智能反射面(RIS, reconfigurable intelligent surface)作为一种极具前景的通信新范例,

引起了无线通信领域的普遍关注, 其可为 5G 和 6G 系统提供智能且可重构的无线信道传输环境。具体来说, 智能反射面是一个二维平面, 其由大量无源反射元件构成, 并且每个元件都能对入射信号独立产生一个可控的振幅和(或)相位, 这从根本上解决了无线信道衰落损伤和干扰问题。基于此, 文献[7-8]将 RIS 引入基于空中计算的 FEEL 框架中, 并以理论和实验证明 RIS 在缓解空中聚合误差方面的有效性, 然而, 其工作依赖于 IID 数据分布的假设。

为此, 本文提出了一种 non-IID 数据分布下智能反射面赋能的空中联邦边缘学习可靠通信方案。具体来说, 本文利用智能反射面的信道可重构性来提高接收信号功率, 并利用空中计算实现联邦边缘学习模型的快速聚合。本文的主要贡献如下。

1) 本文设计了智能反射面赋能的空中计算模型聚合方案。通过分析由 non-IID 数据分布和无线信道衰落联合造成的模型聚合误差, 本文推导出 non-IID 环境下联邦边缘学习的训练损失上界。

2) 通过分析收敛损失上界, 本文进一步构造关于收发端波束成形方案和 RIS 相移矩阵的“通信-学习”优化问题, 并引入一种低复杂度的凸近似策略, 将非凸无线资源优化问题解耦, 直到算法收敛。

1 相关工作

随着 5G 技术的发展, 移动边缘计算(MEC, mobile edge computing)^[9-10]正在成为未来网络的一项关键技术, 以在边缘网络下实现超低功耗和超低延迟应用。联邦边缘学习通过结合联邦学习和 MEC, 可有效增强边缘数据隐私和安全保障, 以提供完备的边缘智能服务^[11-12]。

然而, 由于无线信道衰落和资源的有限性, 通信效率成为在有低时延、隐私和安全保障需求的无线网络中大规模部署联邦学习的关键性能指标^[3,6,13-14]。为此, 文献[6]聚焦正交多址接入技术, 分析了无线数据包传输误差对模型参数更新过程的影响, 并通过优化用户选择和上行链路资源块分配方案, 最小化训练损失。然而, 基于正交多址接入的聚合方式拥有极低的频谱利用效率, 其所需通信资源随着网络规模的增加而线性增长。针对这一问题, 文献[13-14]研究了基于 Aircomp 的非正交多址接入聚合方案, 其利用多址信道的波形叠加特性, 将终端在宽带信道上同时传输的梯度更新进行模拟空中聚合, 以此降低联邦学习通信开销, 提升频谱资

源利用效率。

然而,就现有的工作来说,基于 Aircomp 的联邦学习聚合方案依然面临着一些亟待系统性研究的挑战,如弱信道补偿问题。具体来说,实现空中更新聚合需要从不同移动终端接收到的模型参数具有相同的振幅,因此,空中计算网络需要通过功率控制补偿不同终端的信道衰落,以满足空中计算所需的信道一致性,因此其性能会受限于边缘终端的弱信道,这在一定程度上损失了空中计算网络的通信性能增益。针对此问题,有必要刻画空中计算网络通信误差对联邦学习性能的影响,以此实现最优的“通信-学习”折中。文献[15-16]分别使用数据比例和所选终端数量来近似学习性能,并通过调整通信错误容忍度约束来实现“通信-学习”的统一规划。然而,上述工作可能不能完全描述联邦学习的训练性能,从而导致系统性能的损失。为解决这一问题,文献[17]开发了统一的联邦学习分析框架,来定量地描述模型聚合误差对空中计算 FEEL 收敛性的影响,然而,其工作侧重于 IID 数据分布下的学习设计,这一定程度上与联邦学习的 non-IID 特性相违背。为此,文献[18]通过引入 non-IID 数据分布引发的训练误差,设计了一种基于 Aircomp 的无线聚合方案,然而,其未考虑到 FEEL “通信-学习”的联合设计问题。

综上,如何在 non-IID 数据分布设定下,联合量化信道衰落及数据异构对联邦学习性能的影响,以此设计合理的无线资源分配方案,从而在有限的资源约束下最大化联邦性能收益,仍然是现阶段联邦学习亟待解决的重要问题。

2 问题描述

2.1 联邦学习模型

如图 1 所示,本文考虑由一个 J 根天线的基站和 K 个单天线边缘终端组成的联邦边缘学习系统,其中部署了智能反射面辅助通信。边缘终端 $k \in \mathcal{U} \triangleq \{1, 2, \dots, K\}$ 存有本地数据集 \mathcal{D}_k , 其由已标记的 D_k 对数据样本特征和关联标签 $\mathbf{z} = \{\mathbf{x}, \mathbf{y}\}$ 组成。假设终端在基站的协调下合作训练一个全局模型参数向量 $\mathbf{w} \in \mathbb{R}^D$ 。定义损失函数 $f(\mathbf{w}, \mathbf{z})$, 其量化模型 \mathbf{w} 在训练样本 $\mathbf{z} = \{\mathbf{x}, \mathbf{y}\}$ 上的预测偏差。那么,从形式上来说,边缘终端 k 的局部学习目标是

最小化本地数据集 \mathcal{D}_k 上的经验损失函数,即

$$F_k(\mathbf{w}) \triangleq \frac{1}{D_k} \sum_{\mathbf{z} \in \mathcal{D}_k} f(\mathbf{w}, \mathbf{z}) \quad (1)$$

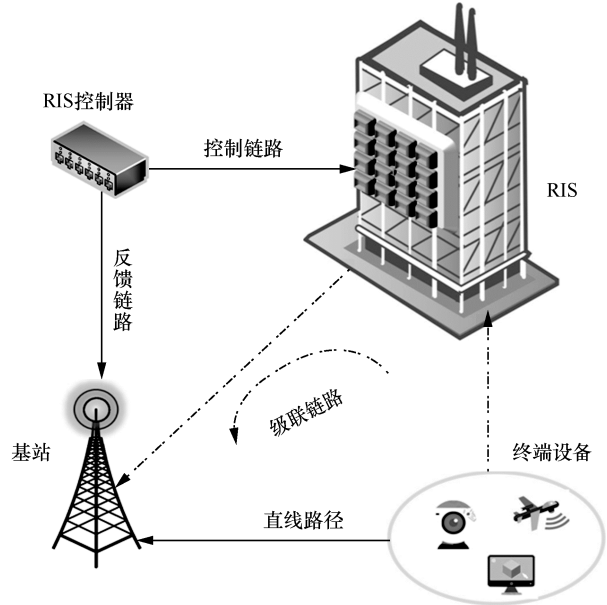


图 1 系统模型

由此,联邦学习的全局优化目标可表示为

$$F(\mathbf{w}) \triangleq \sum_{k=1}^K \frac{D_k}{D_t} F_k(\mathbf{w}) = \frac{1}{D_t} \sum_{k=1}^K \sum_{\mathbf{z} \in \mathcal{D}_k} f(\mathbf{w}, \mathbf{z}) \quad (2)$$

其中, $D_t = \sum_{k=1}^K D_k$ 。此时,最佳模型参数 \mathbf{w}^* 应使得全局损失函数最小化,即表示为

$$\mathbf{w}^* \triangleq \arg \min_{\mathbf{w}} F(\mathbf{w}) \quad (3)$$

为了方便后续的收敛分析,参考文献[7,18],本文假设 $D_k = D_{k'}, \forall k \neq k'$, 即边缘终端数据集大小相同,则全局优化目标重写为 $F(\mathbf{w}) \triangleq \left(\frac{1}{K}\right) \sum_{k=1}^K F_k(\mathbf{w})$ 。注意,本文设计思路可容易扩展到 $D_k \neq D_{k'}, \forall k \neq k'$ 场景。

为了衡量 non-IID 程度,本文定义 2 个新的中间变量,即 $F^* = F(\mathbf{w}^*)$ 和 $F_k^* = \arg \min_{\mathbf{w}} F_k(\mathbf{w})$, 其分别表示 F 和 F_k 的最小值。那么, $\Gamma \triangleq F^* - \left(\frac{1}{K}\right) \sum_{k=1}^K F_k^*$ 则可衡量本地模型与全局模型的偏离程度^[19]。

具体来说,联邦边缘学习框架可分为如下步骤。

1) 基站端将当前的全局模型(由最新的权重向量 \mathbf{w}_t 描述)广播给边缘终端。

2) 每个边缘终端将 w_t 赋值给本地模型 w_t^k , 即 $w_t^k = w_t$, 并使用其本地数据集在接收到的全局模型基础上来更新本地模型。为了降低 non-IID 分布对学习性能的影响, 首先利用 FedProx 算法^[19]限制局部模型相对于全局模型的偏离程度, 此时, 终端 k 的局部学习目标可改写为

$$F_k(w, \mathcal{D}_k) \triangleq \frac{1}{|\mathcal{D}_k|} \sum_{z \in \mathcal{D}_k} f(w, z) + \frac{\mu}{2} \|w - w_t^k\|^2 \quad (4)$$

假定终端使用小批量随机梯度下降 (SGD) 本地训练 E 次迭代生成 w_{t+E}^k , 其中单次更新过程总结为

$$w_{t+1}^k = w_t^k - \eta_t \nabla F(w_t^k, \xi_{t+1}^k) \quad (5)$$

其中, ξ_{t+1}^k 是从终端 k 的本地数据集中独立随机采样的小批数据集。

3) 边缘终端将最新的本地模型 $w_{t+E}^k - w_t^k$ 上传到基站端, 并由其聚合生成新一轮次的全局模型, 即

$$w_{t+E} = w_t + \frac{1}{K} \sum_{k=1}^K (w_{t+E}^k - w_t^k) \quad (6)$$

2.2 RIS 赋能的空中计算模型

假设一个块衰落信道模型, 其中信道系数在整个训练过程中保持不变。值得说明的是, 本文的设计可容易扩展到非块衰落信道时的情况, 仅需要在每次迭代开始之前适当地进行信道估计和参数求解。RIS 由 L 个被动反射元件组成, 记为 $\mathcal{L} \triangleq \{1, 2, \dots, L\}$, 其与控制器关联, 并且, 控制器通过调节 RIS 元件的反射系数来控制信号反射。在 RIS 控制器和 BS 之间部署了单独的控制链路, 以获取反射系数设计所需的信息^[18]。定义终端 k -BS、终端 k -RIS 和 RIS-BS 的信道系数分别为 $h_k^d \in \mathbb{C}^J$ 、 $h_k^c \in \mathbb{C}^L$ 和 $G \in \mathbb{C}^{J \times L}$ 。并且, 对于每个信道, 假设信道状态信息 (CSI) 在 BS 和 RIS 上都能完美估计。假设 FEEL 模型聚合过程中, RIS 相移保持不变。 $\Theta = \text{diag}\{\theta\} \in \mathbb{C}^{L \times L}$ 表示 RIS 相移矩阵, 其中, $\theta = [\theta_1, \theta_2, \dots, \theta_L]^T$, 且 $|\theta_l| = 1$ 。

如上所述, 终端 k -BS 路径有效信道系数可表示为

$$h_k(\theta) \triangleq G\Theta h_k^c + h_k^d = G \text{diag}\{h_k^c\} \theta + h_k^d \quad (7)$$

本文拟采用空中计算技术实现 FEEL 模型的快速聚合。具体来说, 边缘终端使用相同的无线电资源并发地传输其本地模型更新。接着, BS 利用信道的信号叠加特性, 从接收到的信号中恢复聚合模

型。与传统的正交多路访问协议 (终端使用正交信道传输) 相比, 空中计算的带宽需求或通信时延不随终端数量的增加而增加, 极大地缓解了 FEEL 中的通信压力。

在第 t 轮模型聚合中, 将梯度传输时间划分为 D 时隙。定义时隙 $d \in \{1, \dots, D\}$ 中终端 k 的发射信号为 $\{x_k^t[d]\}$, 则 BS 端处对应的接收信号 $\{y_t[d]\}$, 其为直线信道与终端-RIS-BS 级联信道信号的叠加, 即

$$y_t[d] = \sum_{k \in \mathcal{K}} (G\Theta h_k^c + h_k^d) x_k^t[d] + n[d] \quad (8)$$

其中, $n[d] \in \mathbb{C}^J$ 表示加性白高斯噪声 (AWGN) 向量, 其分量服从分布 $\mathcal{CN}(0, \sigma_n^2)$ 。

2.3 基于空中计算的模型聚合

定义最新本地模型 $w_t^k - w_{t-E}^k$ 的第 d 个元素为 $\varpi_k^t[d]$ 。接着, 为了便于功率控制, 将 $\varpi_k^t[d]$ 映射到发射信号 $x_k^t[d]$ 。具体来说, 首先计算终端 k 的局部模型统计量 (即均值和方差), 即

$$\bar{\varpi}_k^t = \frac{1}{D} \sum_{d=1}^D \varpi_k^t[d], \quad (v_k^t)^2 = \frac{1}{D} \sum_{d=1}^D (\varpi_k^t[d] - \bar{\varpi}_k^t)^2 \quad (9)$$

假设边缘终端可通过传统的 OMA 技术无损地上传 $\{\bar{\varpi}_k^t, (v_k^t)^2\}$ 到 BS。此时, 终端 k 传送序列 $x_k^t[d]$ 表示为

$$x_k^t[d] = p_k^t s_k^t[d], \quad s_k^t[d] \triangleq \frac{\varpi_k^t[d] - \bar{\varpi}_k^t}{v_k^t}, \quad \forall d \quad (10)$$

其中, $p_k^t \in \mathbb{C}$ 表示传输均衡因子。可以看到, $\varpi_k^t[d]$ 被映射为零均值单位方差的符号 $s_k^t[d]$ 。因此, $\mathbb{E}[|x_k^t[d]|^2] = |p_k^t|^2$, 其中 $|p_k^t|^2$ 指示传输功率, 其满足单独的发射功率限制, 即 $|p_k^t|^2 \leq P_0, \forall k$ 。

将式(10)代入式(8), BS 端在时隙 d 上的接收信号为

$$y[d] = \sum_{k \in \mathcal{K}} h_k \frac{p_k^t}{v_k^t} (\varpi_k^t[d] - \bar{\varpi}_k^t) + n[d] \quad (11)$$

利用线性估计器, BS 可获得 $r^t[d] = \frac{1}{K} \sum_{k=1}^K \varpi_k^t[d]$ 的估计版本 $\hat{r}^t[d]$, 即

$$\hat{r}^t[d] = \frac{1}{K\sqrt{\beta^t}} f^H y^t[d] + \bar{\varpi}^t =$$

$$\frac{\sum_{k \in \mathcal{L}'} \frac{\mathbf{f}^H \mathbf{h}_k' p_k' (\varpi_k' [d] - \bar{\omega}_k')}{v_k'} + \mathbf{f}^H \mathbf{n} [d]}{K \sqrt{\beta^t}} + \bar{\omega}^t \quad (12)$$

其中, $\bar{\omega}^t \triangleq \frac{1}{K} \sum_{k=1}^K \bar{\omega}_k^t$; $\mathbf{f} \in \mathbb{C}^L$ 表示归一化的接收机波束成形矢量, 其满足 $|\mathbf{f}|^2 = 1$; $\beta^t > 0$ 表示归一化标量; 附加项 $\bar{\omega}^t$ 是由于在式(10)的归一化中减去了 $\bar{\omega}_k^t$ 。需要说明的是, 由于本文假设的是一个块衰落信道模型, 且参照后续 \mathbf{f} 的设计准则, \mathbf{f} 在整个训练过程中保持不变, 即 $\mathbf{f}^t = \mathbf{f}, \forall t$ 。当然, 本文的设计思路可容易地扩展到 \mathbf{f}^t 在不同轮次变化的场景。

由于衰落和通信噪声的存在, 此估计过程不可避免地会出现估计误差。当 BS 收集到所有的 $\hat{r}^t[d]$ 后, 接收向量 $\hat{\mathbf{r}}^t$ 可表示为 $\hat{\mathbf{r}}^t = [\hat{r}^t[1], \hat{r}^t[2], \dots, \hat{r}^t[D]]^T$, 由此,

$$\mathbb{E} \left[\|\mathbf{r}^t - \hat{\mathbf{r}}^t\|_2^2 \right] = \frac{\sum_{d=1}^D \mathbb{E} \left[\sum_{k \in \mathcal{L}'} \left(1 - \frac{\mathbf{f}^H \mathbf{h}_k p_k'}{\sqrt{\beta^t v_k'}} \right) (\varpi_k' [d] - \bar{\omega}_k') + \frac{|\mathbf{f}^H \mathbf{n} [d]|^2}{\beta^t} \right]}{K^2} \stackrel{(a)}{\geq} \frac{1}{\beta^t K^2} \sum_{d=1}^D \mathbb{E} \left[|\mathbf{f}^H \mathbf{n} [d]|^2 \right] = \frac{D \sigma_n^2}{\beta^t K^2} \quad (13)$$

其中, $\mathbf{r}^t = [r^t[1], r^t[2], \dots, r^t[D]]^T$; 当 $\frac{\mathbf{f}^H \mathbf{h}_k p_k'}{\sqrt{\eta^t v_k'}} = 1$ 时, 不等式(a)的 $\mathbb{E} \left[\|\mathbf{r}^t - \hat{\mathbf{r}}^t\|_2^2 \right]$ 取得最小值, 则

$$p_k^t = \frac{\sqrt{\beta^t v_k'} (\mathbf{f}^H \mathbf{h}_k)^H}{|\mathbf{f}^H \mathbf{h}_k|^2} \quad (14)$$

可得 $\hat{r}^t[d] = \frac{1}{K} \sum_{k \in \mathcal{L}'} \varpi_k' [d] + \frac{1}{K \sqrt{\beta^t}} \mathbf{f}^H \mathbf{n} [d]$ 。由此

可知, $\hat{r}^t[d]$ 是 $r^t[d] = \frac{1}{K} \sum_{k=1}^K \varpi_k' [d]$ 的无偏估计。由于

发射功率限制, 可得 $|p_k^t|^2 = \frac{\beta^t (v_k')^2}{|\mathbf{f}^H \mathbf{h}_k|^2} \leq P_0, \forall k$, 即

$$\beta^t \leq P_0 \frac{|\mathbf{f}^H \mathbf{h}_k|^2}{(v_k')^2}, \forall k \Rightarrow \beta^t \leq \min_{k \in \mathcal{L}'} \frac{P_0 |\mathbf{f}^H \mathbf{h}_k|^2}{(v_k')^2} \quad (15)$$

由于 $\mathbb{E} \left[\|\mathbf{r}^t - \hat{\mathbf{r}}^t\|_2^2 \right]$ 与 η^t 成反比, 可得

$$\beta^t = \min_{k \in \mathcal{L}'} \frac{P_0 |\mathbf{f}^H \mathbf{h}_k|^2}{(v_k')^2} \quad (16)$$

此时

$$\mathbb{E} \left[\|\mathbf{r}^t - \hat{\mathbf{r}}^t\|_2^2 \right] = \frac{D \sigma_n^2}{\beta^t K^2} = \frac{D \sigma_n^2}{P_0 K^2} \max_{k \in \mathcal{L}'} \frac{(v_k')^2}{|\mathbf{f}^H \mathbf{h}_k|^2} \quad (17)$$

由式(6)可得

$$\mathbf{w}_t = \mathbf{w}_{t-E} + \hat{\mathbf{r}}^t = \frac{1}{K} \sum_{k=1}^K \mathbf{w}_k^t + \frac{1}{K \sqrt{\beta^t}} \tilde{\mathbf{w}} \quad (18)$$

其中, $\tilde{\mathbf{w}} \in \mathbb{C}^D$, 其分量表示为 $\tilde{w}[d] = \mathbf{f}^H \mathbf{n} [d]$ 。

3 收敛分析及问题建模

接下来, 着重阐述 RIS 赋能的联邦边缘学习系统收敛理论, 并具体分析无线信道及 RIS 对空中联邦学习全局模型收敛上限的影响。

3.1 基础假设

首先, 对于收敛性分析的假设和定义进行说明。

假设 1 本地目标函数 F_k 是 L 平滑的, 其满足

$$F_k(\mathbf{v}) \leq F_k(\mathbf{w}) + (\mathbf{v} - \mathbf{w})^T \nabla F_k(\mathbf{w}) + \frac{L}{2} \|\mathbf{v} - \mathbf{w}\|^2.$$

假设 2 本地目标函数 F_k 是 μ -强凸的, 其满足

$$F_k(\mathbf{v}) \geq F_k(\mathbf{w}) + (\mathbf{v} - \mathbf{w})^T \nabla F_k(\mathbf{w}) + \frac{\mu}{2} \|\mathbf{v} - \mathbf{w}\|^2.$$

假设 3 对于从用户 k 本地数据集上均匀采样的小批量数据集 ξ_t^k , 所得到的随机梯度是无偏的, 即 $\mathbb{E} \left[\nabla F_k(\mathbf{w}_t^k, \xi_t^k) \right] = \nabla F_k(\mathbf{w}_t^k)$, 且随机梯度偏差是有界的, 即 $\mathbb{E} \left[\left\| \nabla F_k(\mathbf{w}_t^k, \xi_t^k) - \nabla F_k(\mathbf{w}_t^k) \right\|^2 \right] \leq \sigma_k^2$ 。

假设 4 随机梯度的期望 $\mathbb{E} \left[\left\| \nabla F_k(\mathbf{w}_t^k, \xi_t^k) \right\|^2 \right]$ 是有界的, 其满足 $\mathbb{E} \left[\left\| \nabla F_k(\mathbf{w}_t^k, \xi_t^k) \right\|^2 \right] \leq H^2$ 。

上述假设在联邦学习收敛性分析中被广泛地采用, 参考文献[16,20-21]。

3.2 收敛分析

在上述假设的基础上, 本节给出 RIS 赋能的空中联邦边缘学习系统收敛性定理。

定理 1 令 $\gamma = \max \left\{ \frac{8L}{\mu}, E \right\}$, $\kappa = \frac{L}{\mu}$, $\eta_t = \frac{2}{\mu(t+\gamma)}$, 则在上述假设和 non-IID 数据分布下, RIS 赋能的空中联邦边缘学习系统收敛性满足

$$\mathbb{E}\left[\|F(\bar{\mathbf{w}}_T) - F^*\|^2\right] \leq \frac{2\kappa}{T + \gamma} \left[\frac{U}{\mu} + \left(2L + \frac{\mu E}{4}\right) \|\bar{\mathbf{w}}_0 - \mathbf{w}^*\|^2 \right] \quad (19)$$

其中

$$U = \sum_{k=1}^K \frac{\sigma_k^2}{K^2} + 6L\Gamma + 8(E-1)^2 H^2 + \max_{k \in \mathcal{L}} \frac{4(E-1)^2 H^2 \sigma_n^2}{P_0 K^2} \frac{1}{|\mathbf{f}^H \mathbf{h}_k|^2} \quad (20)$$

证明过程见附录 1。

上述收敛上界依赖于联邦边缘学习系统通信环境的配置，即智能反射面相位设计和收发端波束成形设计。直观地说，由通信引入的误差会累积起来，并反映在新的全局模型中，从而影响 FEEL 的收敛性。若上述通信配置设计得当，收敛速度可以在很大程度上得到保证。由定理 1 还可以看出，收敛上界与超参数有一定的单调关系。例如，边界随着 E 的增大而增大，这与文献[22]的结果一致。并且 non-IID 数据分布的影响体现在 Γ 引入的误差项上。当 Γ 减小时，收敛上界单调减小；当 Γ 趋于零时，可得到 IID 数据分布的收敛上界。

3.3 问题建模

参照定理 1，空中计算通信系统会在收敛上界中引入一个误差项，其揭示了智能反射面相位偏移设计和收发端波束成形对收敛速度和学习性能的影响。 $\max_{k \in \mathcal{L}} \frac{1}{|\mathbf{f}^H (\mathbf{G} \text{diag}\{\mathbf{h}_k^c\} \boldsymbol{\theta} + \mathbf{h}_k^d)|^2}$ 越大，学习性能越差。由此，可以建立一个关于收发端波束成形方案和 RIS 配置矩阵的“通信-学习”优化问题，表示为

$$\begin{aligned} (\mathcal{P}) \quad & \min_{\mathbf{f}, \boldsymbol{\theta}} \max_{k \in \mathcal{L}} u_k(\mathbf{f}, \boldsymbol{\theta}) = -|\mathbf{f}^H \mathbf{h}_k|^2 \\ & \text{s.t. } |\theta_l|^2 = 1, \quad \forall l = 1, \dots, L \\ & \|\mathbf{f}\|^2 = 1 \end{aligned} \quad (21)$$

4 联合优化

由于变量的完全耦合以及单位模非凸约束，问题 \mathcal{P} 无法直接获取最优解。为此，本文采用连续凸逼近 SCA 策略^[18,23]，通过求解一系列凸近似问题来迭代更新 \mathbf{f} 和 $\boldsymbol{\theta}$ 。具体来说，定义 $\mathbf{H}_k = \mathbf{G} \text{diag}\{\mathbf{h}_k^c\}$ ，则在迭代 i 时，引入凸代理函数 $\tilde{u}_k^{(i)}(\mathbf{f}, \boldsymbol{\theta})$ ，其可以

看作 $u_k(\mathbf{f}, \boldsymbol{\theta})$ 的近似，表示为

$$\begin{aligned} \tilde{u}_k^{(i)}(\mathbf{f}, \boldsymbol{\theta}) = & u_k(\mathbf{f}^{(i)}, \boldsymbol{\theta}^{(i)}) + \text{Re}\left\{(\mathbf{f} - \mathbf{f}^{(i)})^H \nabla_{\mathbf{f}} u(\mathbf{f}^{(i)}, \boldsymbol{\theta}^{(i)})\right\} + \\ & \text{Re}\left\{(\boldsymbol{\theta} - \boldsymbol{\theta}^{(i)})^H \nabla_{\boldsymbol{\theta}} u(\mathbf{f}^{(i)}, \boldsymbol{\theta}^{(i)})\right\} + \\ & \tau \|\mathbf{f} - \mathbf{f}^{(i)}\|^2 + \tau \|\boldsymbol{\theta} - \boldsymbol{\theta}^{(i)}\|^2 = \\ & \tau \|\mathbf{f} - \mathbf{f}^{(i)}\|^2 + \tau \|\boldsymbol{\theta} - \boldsymbol{\theta}^{(i)}\|^2 + \left|(\mathbf{f}^{(i)})^H \mathbf{h}_k(\boldsymbol{\theta}^{(i)})\right|^2 - \\ & 2\text{Re}\left\{\mathbf{f}^H \mathbf{h}_k(\boldsymbol{\theta}^{(i)}) \mathbf{h}_k(\boldsymbol{\theta}^{(i)})^H \mathbf{f}^{(i)}\right\} - \\ & 2\text{Re}\left\{\boldsymbol{\theta}^H \mathbf{H}_k^H \mathbf{f}^{(i)} (\mathbf{f}^{(i)})^H \mathbf{h}_k(\boldsymbol{\theta}^{(i)})\right\} + \\ & 2\text{Re}\left\{(\mathbf{f}^{(i)})^H \mathbf{h}_k(\boldsymbol{\theta}^{(i)}) (\boldsymbol{\theta}^{(i)})^H \mathbf{H}_k^H \mathbf{f}^{(i)}\right\} \end{aligned} \quad (22)$$

其中， $\nabla_{\mathbf{f}} u(\mathbf{f}^{(i)}, \boldsymbol{\theta}^{(i)})$ 和 $\nabla_{\boldsymbol{\theta}} u(\mathbf{f}^{(i)}, \boldsymbol{\theta}^{(i)})$ 分别表示 $\mathbf{f} = \mathbf{f}^{(i)}$ 和 $\boldsymbol{\theta} = \boldsymbol{\theta}^{(i)}$ 时 $u(\mathbf{f}, \boldsymbol{\theta})$ 的共轭梯度；正则化参数 τ 用于确保强凸性。由此， $\mathbf{f}^{(i+1)}$ 和 $\boldsymbol{\theta}^{(i+1)}$ 可通过求解以下优化问题得到

$$\{\mathbf{f}^{(i+1)}, \boldsymbol{\theta}^{(i+1)}\} = \arg \min_{\mathbf{f}, \boldsymbol{\theta}} \max_{k \in \mathcal{L}} \tilde{u}_k^{(i)}(\mathbf{f}, \boldsymbol{\theta}) \quad (23)$$

由于 $\|\mathbf{f}\|^2 = 1$ ，且 $\|\boldsymbol{\theta}\|^2 = L$ ，上述问题等价于

$$\begin{aligned} \{\mathbf{f}^{(i+1)}, \boldsymbol{\theta}^{(i+1)}\} = \arg \min_{\mathbf{f}, \boldsymbol{\theta}, \kappa} & \\ \text{s.t. } \kappa \geq c_k^{(i)} - 2\text{Re}\left\{\mathbf{f}^H \mathbf{a}_k^{(i)} + \boldsymbol{\theta}^H \mathbf{b}_k^{(i)}\right\}, & k \in \mathcal{L} \end{aligned} \quad (24)$$

其中

$$\begin{aligned} \mathbf{a}_k^{(i)} & \triangleq \left(\tau + \mathbf{h}_k(\boldsymbol{\theta}^{(i)}) \mathbf{h}_k(\boldsymbol{\theta}^{(i)})^H\right) \mathbf{f}^{(i)} \\ \mathbf{b}_k^{(i)} & \triangleq \tau \boldsymbol{\theta}^{(i)} + \mathbf{H}_k^H \mathbf{f}^{(i)} (\mathbf{f}^{(i)})^H \mathbf{h}_k(\boldsymbol{\theta}^{(i)}) \\ c_k^{(i)} & \triangleq \left|(\mathbf{f}^{(i)})^H \mathbf{h}_k(\boldsymbol{\theta}^{(i)})\right|^2 + 2\tau(L+1) + \\ & 2\text{Re}\left\{(\mathbf{f}^{(i)})^H \mathbf{h}_k(\boldsymbol{\theta}^{(i)}) (\boldsymbol{\theta}^{(i)})^H \mathbf{H}_k^H \mathbf{f}^{(i)}\right\} \end{aligned} \quad (25)$$

上述问题仍然是非凸的。为此，本文采用拉格朗日对偶算法将其近似。定义拉格朗日对偶向量 $\boldsymbol{\xi} = [\xi_1, \xi_2, \dots, \xi_K]^T \geq \mathbf{0}$ ，则上述对偶问题表示为

$$\begin{aligned} \mathcal{J}^{(i)}(\mathbf{f}, \boldsymbol{\theta}, \kappa, \boldsymbol{\xi}) = & \kappa + \sum_{k \in \mathcal{L}} \xi_k \left(c_k^{(i)} - \kappa - 2\text{Re}\left\{\mathbf{f}^H \mathbf{a}_k^{(i)} + \boldsymbol{\theta}^H \mathbf{b}_k^{(i)}\right\}\right) = \\ & \sum_{k \in \mathcal{L}} \xi_k c_k^{(i)} + \kappa \left(1 - \sum_{k \in \mathcal{L}} \xi_k\right) - 2\text{Re}\left\{\mathbf{f}^H \sum_{k \in \mathcal{L}} \xi_k \mathbf{a}_k^{(i)} + \boldsymbol{\theta}^H \sum_{k \in \mathcal{L}} \xi_k \mathbf{b}_k^{(i)}\right\} \end{aligned} \quad (26)$$

由此, $\mathbf{f}^{(i+1)}$ 和 $\boldsymbol{\theta}^{(i+1)}$ 可通过求解以下的优化问题得到

$$\{\mathbf{f}^{(i+1)}, \boldsymbol{\theta}^{(i+1)}\} = \arg \max_{\boldsymbol{\xi}} \min_{\mathbf{f}, \boldsymbol{\theta}, \kappa} \mathcal{J}^{(i)}(\mathbf{f}, \boldsymbol{\theta}, \kappa, \boldsymbol{\xi}) \quad (27)$$

由 $\mathcal{J}^{(i)}(\mathbf{f}, \boldsymbol{\theta}, \kappa, \boldsymbol{\xi})$ 的定义可知, 当 $1 - \sum_{k \in \mathcal{L}} \xi_k \neq 0$ 时, $\min_{\mathbf{f}, \boldsymbol{\theta}, \kappa} \mathcal{J}^{(i)}(\mathbf{f}, \boldsymbol{\theta}, \kappa, \boldsymbol{\xi}) = -\infty$; 反之, 当 $1 - \sum_{k \in \mathcal{L}} \xi_k = 0$ 时, 可得

$$\begin{aligned} \mathcal{J}^{(i)}(\mathbf{f}, \boldsymbol{\theta}, \kappa, \boldsymbol{\xi}) = & \sum_{k \in \mathcal{L}} \xi_k c_k^{(i)} - 2 \operatorname{Re} \left\{ \mathbf{f}^H \sum_{k \in \mathcal{L}} \xi_k \mathbf{a}_k^{(i)} + \boldsymbol{\theta}^H \sum_{k \in \mathcal{L}} \xi_k \mathbf{b}_k^{(i)} \right\} \stackrel{(a)}{\geq} \\ & \sum_{k \in \mathcal{L}} \xi_k c_k^{(i)} - 2 \left\| \sum_{k \in \mathcal{L}} \xi_k \mathbf{a}_k^{(i)} \right\|_2 - 2 \left\| \sum_{k \in \mathcal{L}} \xi_k \mathbf{b}_k^{(i)} \right\|_2 \end{aligned} \quad (28)$$

其中, 不等式(a)源于 Cauchy-Schwarz 不等式。当且仅当 $\mathbf{f}^{(i+1)}(\boldsymbol{\xi}) = \frac{\sum_{k \in \mathcal{L}} \xi_k \mathbf{a}_k^{(i)}}{\left\| \sum_{k \in \mathcal{L}} \xi_k \mathbf{a}_k^{(i)} \right\|_2}$, $\boldsymbol{\theta}^{(i+1)}(\boldsymbol{\xi}) = e^{j \arg \left\{ \sum_{k \in \mathcal{L}} \xi_k \mathbf{b}_k^{(i)} \right\}}$ 时

等式成立。 $\boldsymbol{\xi}$ 可通过求解以下优化问题得到

$$\begin{aligned} \min_{\boldsymbol{\xi}} & - \sum_{k \in \mathcal{L}} \xi_k c_k^{(i)} + 2 \left\| \sum_{k \in \mathcal{L}} \xi_k \mathbf{a}_k^{(i)} \right\|_2 + \left\| \sum_{k \in \mathcal{L}} \xi_k \mathbf{b}_k^{(i)} \right\|_2 \\ \text{s.t. } & \boldsymbol{\xi} \geq 0, \\ & \sum_{k \in \mathcal{L}} \xi_k = 1 \end{aligned} \quad (29)$$

当式(29)用凸优化求解器求解时, 若采用内点法, 其最坏情况复杂度上界为 $O(K^3)$ 。假设最大迭代次数为 I_{\max} , 则 SCA 策略复杂度上界为 $O(I_{\max} K^3)$ 。

5 实验设计与分析

接下来, 通过仿真实验评估所提方法的性能, 并与几种基准方案作比较。

5.1 实验设定

如图 2 所示, 本文考虑三维笛卡儿坐标系下的仿真设定。其中, BS 和 RIS 分别放置在 $(-50, 0, 10)$ 和 $(0, 0, 10)$ 处。边缘终端随机分布在 2 个半径为 10 m 的圆形区域, 即区域 I $\propto \{(x, y, 0) : -20 \leq x \leq 0, -10 \leq y \leq 10\}$ 以及区域 II $\propto \{(x, y, 0) : 100 \leq x \leq 120, -10 \leq y \leq 10\}$ 。简单起见, 假设终端中有一半随机分布在区域 I, 另一半随机分布在区域 II。

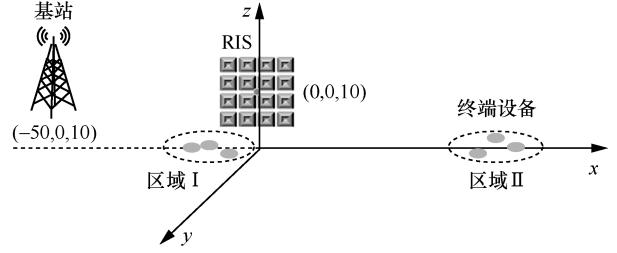


图 2 仿真设定

与文献[18]类似, 本文相关信道系数定义为大尺度衰落和小尺度衰落之积。具体来说, 小尺度衰落服从标准高斯分布, 并且终端到基站端的路径损耗模型表示为

$$P_{\text{DB}} = G_{\text{BS}} G_{\text{D}} \left(\frac{3 \times 10^8}{4\pi f_c d_{\text{DB}}} \right)^{\text{PL}} \quad (30)$$

其中, G_{BS} 和 G_{D} 分别表示基站端和终端的天线增益; f_c 表示载波频率; PL 表示路径损耗因子; d_{DB} 表示终端到基站端的距离。终端-RIS-BS 的路径损耗模型建模为

$$P_{\text{DRB}} = G_{\text{BS}} G_{\text{D}} G_{\text{RIS}} \frac{L^2 d_x d_y \left(\frac{3 \times 10^8}{f_c} \right)^2}{64\pi^3 d_{\text{RB}}^2 d_{\text{DR}}^2} \quad (31)$$

其中, G_{RIS} 表示 RIS 端的天线增益; d_x 和 d_y 分别表示 RIS 单元的水平 and 垂直大小; d_{RB} 及 d_{DR} 分别表示 RIS 到基站端的距离以及终端到 RIS 端的距离。如无特别说明, 仿真参数如表 1 所示。

表 1 仿真参数

参数	值
K	40
J	5
L	50
P_0 / dB	-10
σ_n^2 / dB	-100
μ	1
E	1
τ	1
PL	3.76
f_c / MHz	915
$G_{\text{BS}} / \text{dBi}$	5
$G_{\text{D}} / \text{dBi}$	0
$G_{\text{DR}} / \text{dBi}$	5
d_x	$\frac{3 \times 10^7}{f_c}$
d_y	$\frac{3 \times 10^7}{f_c}$

在 FEMNIST 数据集^[24]上执行图像分类任务，以验证所提方法性能。具体来说，首先构建一个 4 层的卷积神经网络 (CNN)，其中包括 2 个 5×5 卷积层以及 2 个全连接层。为了展现所提方案 (OPT) 的性能，本文设置了如下的基准方案进行了性能比较。

1) DGP: 无 RIS 辅助通信，即 $\theta = 0$ 。通过微分几何规划^[25]，设计接收机波束成形矢量 \mathbf{f} ，以最小化通信均方误差。

2) RDGP: 采用随机 RIS 相移，并通过微分几何规划，设计接收机波束成形矢量 \mathbf{f} ，以最小化通信均方误差。

3) SDR: 针对问题 \mathcal{P} ，RIS 相移 θ 以及接收机波束成形矢量 \mathbf{f} 通过半正定松弛算法^[26]迭代优化。

此外，将无噪声传输方案 (S-FedAvg) 作为性能上界，用于比较。

5.2 性能分析

图 3 展示了 non-IID 设定下的联邦学习测试精度比较。从图 3 可以看到，本文方案取得了比其他基准方案更优异的收敛性能。并且，值得一提的是，相较于无 RIS 辅助的 DGP 设计，除了本文方案，其余 3 种方案都实现了显著的性能提升，这证明了在 non-IID 设定下，RIS 同样可以提高空中 FEEL 系统准确性。

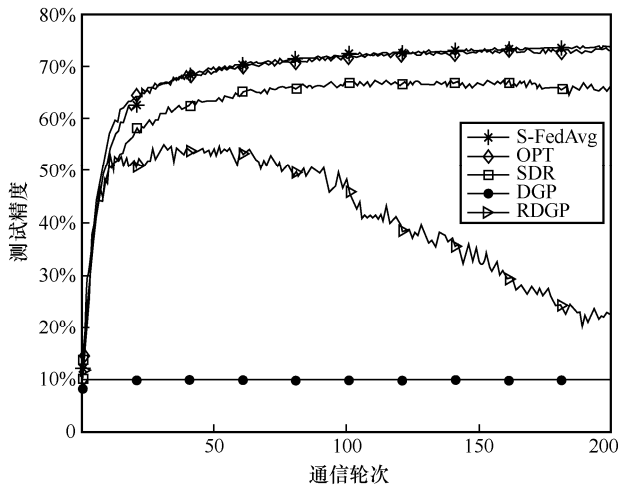


图 3 non-IID 设定下的联邦学习测试精度比较

图 4 展示了 IID 设定下 (即 $\Gamma = 0$) 的联邦学习测试精度比较。从图 4 可以看出，相较于 non-IID 设定，SDR、OPT 以及 S-FedAvg 皆实现了一定程度的性能提升，原因在于，当 $\Gamma = 0$ 时，定理 1 所示的收敛上界中消除了关于 non-IID 数据分布的误差项，提升了学习性能。

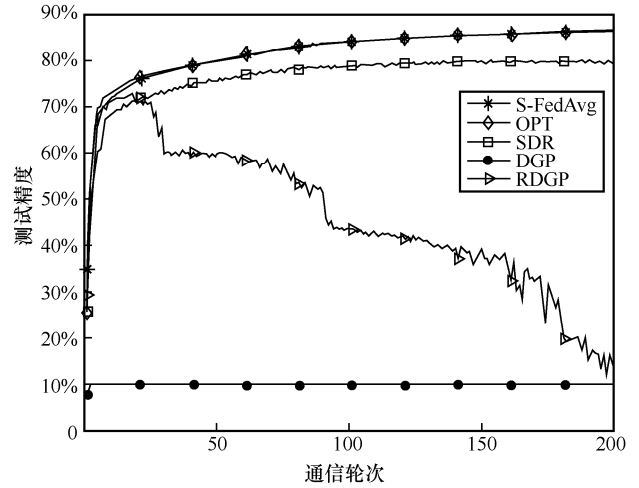


图 4 IID 设定下的联邦学习测试精度比较

图 5 展示了 non-IID 设定下不同反射单元数量对测试精度的影响。显然，SDR 和 OPT 都可以在较多的反射单元数量下实现相当大的性能提升，这是因为反射单元数量的增加会提高调配智能反射面适配信道的自由度，即通过设计智能反射面的相移产生更窄波束形成，以增加通信链路的可靠性，降低基站端的聚合误差。然而，所提方案与其他方案之间的性能差距仍然存在。

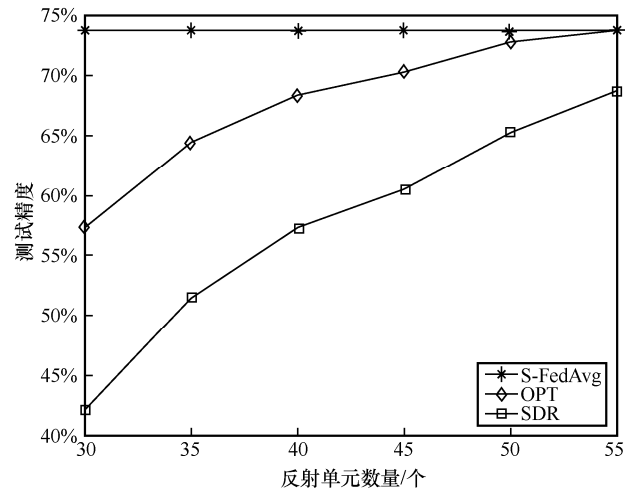


图 5 non-IID 设定下不同反射单元数量对测试精度的影响

如图 6 所示，进一步假设每个 RIS 相移只能取有限个离散值，即

$$\theta_l \in \mathcal{F} = \left\{ \exp\left(\frac{j2\pi m}{2^b}\right) \right\}_{m=0}^{2^b-1} \quad (32)$$

其中，比特数 $b \in \mathcal{B} \triangleq \{1, 2, 3, \infty\}$ 。当 $b = \infty$ 时， \mathcal{F} 为连续集，即 $|\theta_l| = 1$ 。可以看到，随着 b 的增加，

系统性能逐渐提升,这是因为低位相移引入了额外匹配误差,从而降低了训练性能。

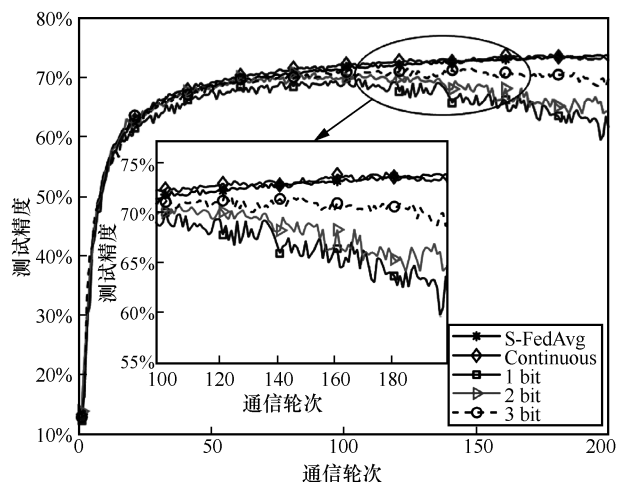


图6 non-IID 设定下不同离散比特对测试精度的影响

接下来,考虑不完美 CSI 对测试精度的影响,如图7所示。其中,Continuous 表示 $b=\infty$ 。具体来说,假设从边缘终端到 BS 的直线路径和通过 RIS 的级联链路是不完美的。直线信道由 $\hat{h}_k^d = h_k^d + \Delta h_k^d$ 给出,其中 h_k^d 是在 BS 处已知的估计信道, Δh_k^d 是信道误差。级联链路表示为 $\hat{C}_k = G \text{diag}\{h_k^c\} = C_k + \Delta C_k$, 其中, C_k 和 ΔC_k 分别表示对应的估计信道和估计误差。指示 CSI 不确定性的误差向量 Δh_k^d 和 $\text{vec}\{\Delta C_k\}$ 遵循圆对称复高斯分布,即 $\Delta h_k^d \sim \mathcal{CN}(\mathbf{0}, \varepsilon_k^2 \mathbf{I})$, $\text{vec}\{\Delta C_k\} \sim \mathcal{CN}(\mathbf{0}, \varepsilon_k^2 \mathbf{I})$ 。从图7中可以看到,所提方案对不完美 CSI 具有一定的鲁棒性。

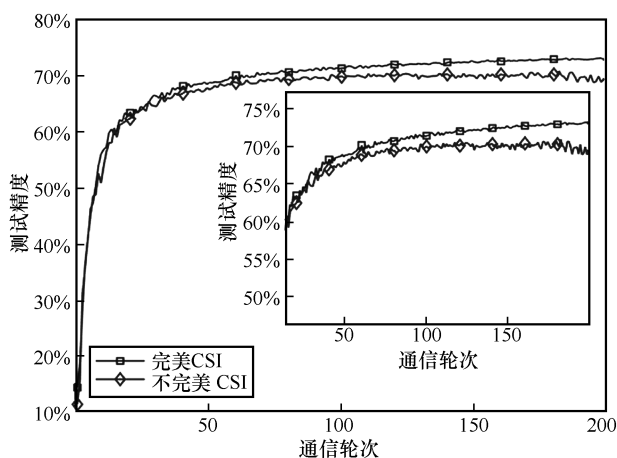


图7 non-IID 设定下不完美 CSI 对测试精度的影响

6 在车联网中的应用探讨

在特大城市如北京、上海,车辆或其他移动终

端容易受高层建筑遮挡而造成信号传播盲区,将 RIS 安装在视野更开阔的建筑物表面,可以中继移动终端的信号给基站,利用智能反射面的信道可重构性来提高接收信号功率,并利用空中计算实现模型快速聚合增强无线资源的收益,如图8所示。

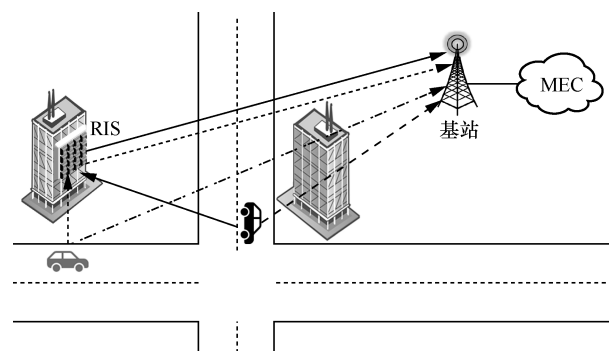


图8 智能反射面赋能的联邦边缘学习在车联网中的应用

7 结束语

针对 non-IID 设定下的联邦边缘学习框架,本文提出一种智能反射面赋能的空中计算聚合模型。通过量化由 non-IID 数据分布和无线信道衰落共同引起的模型聚合误差,首先推导出 non-IID 环境下空中联邦边缘学习的训练损失上界,并以此为基础,建立通信资源约束下以收发端波束成形和 RIS 相移为变量的损失上界最小化问题;同时,引入一种低复杂度的凸逼近策略,高效求解最优通算参数,以减弱甚至消除通信误差对联邦学习收敛性能的影响。最后,通过仿真实验验证了所提方案的有效性。

未来,将考虑更多通用的场景,例如非凸损失函数下的“通信-学习”联合建模和折中优化,以及引入模型量化技术进一步降低传输开销等,从而为深层次理解联邦边缘学习融合无线通信新范式提供新的理论支撑。另外,还可探讨将 RIS 用于 V2V (vehicle to vehicle) 场景,有助于消除移动终端之间信号干扰,增强每个移动终端的接收信号质量。

附录 1 RIS 赋能的空中联邦边缘学习系统收敛性证明

首先,定义时间轴刻度为 SGD 更新轮次,如式(33)所示

$$t = \underbrace{1, \dots, E}_{\text{round 1}}, \underbrace{E+1, \dots, 2E}_{\text{round 2}}, \dots, \underbrace{(T-1)E+1, \dots, TE}_{\text{round T}} \quad (33)$$

设 w_t^k 为第 t 次迭代时第 k 个终端上的模型权重, w_t 为

第 t 次迭代时的全局模型。在标准 FedAvg 中, 终端在全局聚合之前执行 E 次局部迭代。因此, \mathbf{w}_t 只在特定的区间上, 即 $\mathcal{I}_E \triangleq \{nE | n=1, 2, \dots, T\}$ 才具有实际意义。定义 $\mathbf{v}_{t+1}^k = \mathbf{w}_t^k - \eta_t \nabla F(\mathbf{w}_t^k, \xi_{t+1}^k)$ 。可以看到, 当且仅当 $t+1 \notin \mathcal{I}_E$ 时, $\mathbf{w}_{t+1}^k = \mathbf{v}_{t+1}^k$; 反之, $t+1 \in \mathcal{I}_E$, 边缘终端通过空中计算网络将其权重 \mathbf{v}_{t+1}^k 传输到 BS 端并解码。因此, 定义

$$\begin{aligned} \mathbf{u}_{t+1}^k &\triangleq \begin{cases} \mathbf{v}_{t+1}^k, & t+1 \notin \mathcal{I}_E \\ \frac{1}{K} \sum_{k=1}^K \mathbf{v}_{t+1}^k, & t+1 \in \mathcal{I}_E \end{cases} \\ \bar{\mathbf{w}}_{t+1} &\triangleq \begin{cases} \mathbf{v}_{t+1}^k, & t+1 \notin \mathcal{I}_E \\ \frac{1}{K} \sum_{k=1}^K \mathbf{v}_{t+1}^k + \frac{\tilde{\mathbf{w}}}{K\sqrt{\beta^{t+1}}}, & t+1 \in \mathcal{I}_E \end{cases} \end{aligned} \quad (34)$$

定义虚拟变量

$$\begin{aligned} \bar{\mathbf{v}}_t &= \frac{1}{K} \sum_{k=1}^K \mathbf{v}_t^k, \bar{\mathbf{u}}_t = \frac{1}{K} \sum_{k=1}^K \mathbf{u}_t^k, \bar{\mathbf{w}}_t = \frac{1}{K} \sum_{k=1}^K \mathbf{w}_t^k \\ \bar{\mathbf{g}}_t &= \frac{1}{K} \sum_{k=1}^K \nabla F(\mathbf{w}_t^k), \mathbf{g}_t = \frac{1}{K} \sum_{k=1}^K \nabla F(\mathbf{w}_t^k, \xi_t^k) \end{aligned} \quad (35)$$

因此, $\bar{\mathbf{v}}_{t+1} = \bar{\mathbf{w}}_t - \eta_t \mathbf{g}_t$ 且 $\mathbb{E}[\mathbf{g}_t] = \bar{\mathbf{g}}_t$ 。当 $t+1 \in \mathcal{I}_E$ 时, 有

$$\bar{\mathbf{w}}_{t+1} = \bar{\mathbf{u}}_{t+1} + \frac{\tilde{\mathbf{w}}}{K\sqrt{\beta^{t+1}}} \quad (36)$$

接下来, 引入在定理 1 的证明过程中一些必要的引理, 其具体证明过程参考文献[20]。其中, 引理 1 是根据损失函数的光滑性和凸性推导出来的; 引理 2 源于 SGD 方差有界性, 而引理 3 由损失函数的有界梯度导出。

引理 1 若假设 1 和假设 2 成立, 令 $\eta_t \leq \frac{1}{4L}$, 则满足

$$\begin{aligned} \mathbb{E}[\|\bar{\mathbf{v}}_{t+1} - \mathbf{w}^*\|^2] &\leq (1-\eta_t\mu)\mathbb{E}[\|\bar{\mathbf{w}}_t - \mathbf{w}^*\|^2] + \eta_t^2\mathbb{E}[\|\mathbf{g}_t - \bar{\mathbf{g}}_t\|^2] + \\ &6L\eta_t^2\Gamma + 2\mathbb{E}\left[\frac{1}{K}\sum_{k=1}^K\|\bar{\mathbf{w}}_t - \mathbf{w}_t^k\|^2\right] \end{aligned} \quad (37)$$

引理 2 若假设 3 成立, 满足

$$\mathbb{E}[\|\mathbf{g}_t - \bar{\mathbf{g}}_t\|^2] \leq \sum_{k=1}^K \frac{\sigma_k^2}{K^2} \quad (38)$$

引理 3 若假设 4 成立, η_t 非增且对于任意 t 均满足 $\eta_t \leq 2\eta_{t+E}$, 则有

$$\mathbb{E}\left[\frac{1}{K}\sum_{k=1}^K\|\bar{\mathbf{w}}_t - \mathbf{w}_t^k\|^2\right] \leq 4\eta_t^2(E-1)^2H^2 \quad (39)$$

当 $t+1 \notin \mathcal{I}_E$, $\bar{\mathbf{w}}_{t+1} = \bar{\mathbf{v}}_{t+1}$ 时, 利用引理 1、引理 2 和引理 3, 可得

$$\begin{aligned} \mathbb{E}[\|\bar{\mathbf{w}}_{t+1} - \mathbf{w}^*\|^2] &= \mathbb{E}[\|\bar{\mathbf{v}}_{t+1} - \mathbf{w}^*\|^2] \leq \\ &(1-\eta_t\mu)\mathbb{E}[\|\bar{\mathbf{w}}_t - \mathbf{w}^*\|^2] + \eta_t^2\mathbb{E}[\|\mathbf{g}_t - \bar{\mathbf{g}}_t\|^2] + \\ &6L\eta_t^2\Gamma + 2\mathbb{E}\left[\frac{1}{K}\sum_{k=1}^K\|\bar{\mathbf{w}}_t - \mathbf{w}_t^k\|^2\right] \leq \\ &(1-\eta_t\mu)\mathbb{E}[\|\bar{\mathbf{w}}_t - \mathbf{w}^*\|^2] + \\ &\eta_t^2\mathbb{E}\left[\sum_{k=1}^K\frac{\sigma_k^2}{K^2} + 6L\Gamma + 8(E-1)^2H^2\right] \end{aligned} \quad (40)$$

当 $t+1 \in \mathcal{I}_E$ 时, 边缘终端通过空中计算网络将其模型传输到 BS 端并解码, 则有

$$\begin{aligned} \mathbb{E}[\|\bar{\mathbf{w}}_{t+1} - \mathbf{w}^*\|^2] &= \mathbb{E}[\|\bar{\mathbf{w}}_{t+1} - \bar{\mathbf{u}}_{t+1} + \bar{\mathbf{u}}_{t+1} - \mathbf{w}^*\|^2] \leq \\ &\mathbb{E}[\|\bar{\mathbf{w}}_{t+1} - \bar{\mathbf{u}}_{t+1}\|^2] + \mathbb{E}[\|\bar{\mathbf{u}}_{t+1} - \mathbf{w}^*\|^2] + \\ &2\mathbb{E}[\langle \bar{\mathbf{w}}_{t+1} - \bar{\mathbf{u}}_{t+1}, \bar{\mathbf{u}}_{t+1} - \mathbf{w}^* \rangle] \stackrel{(a)}{=} \\ &\mathbb{E}[\|\bar{\mathbf{w}}_{t+1} - \bar{\mathbf{u}}_{t+1}\|^2] + \mathbb{E}[\|\bar{\mathbf{v}}_{t+1} - \mathbf{w}^*\|^2] \end{aligned} \quad (41)$$

其中, 等式(a)源于 $\mathbb{E}[\bar{\mathbf{w}}_{t+1} - \bar{\mathbf{u}}_{t+1}] = 0$ 。由此可得

$$\mathbb{E}[\|\bar{\mathbf{w}}_{t+1} - \bar{\mathbf{u}}_{t+1}\|^2] = \mathbb{E}\left[\left\|\frac{\tilde{\mathbf{w}}}{K\sqrt{\beta^{t+1}}}\right\|_2^2\right] = \frac{D\sigma_n^2}{P_0K^2} \max_{k \in \mathcal{I}} \frac{\mathbb{E}[(v_k^{t+1})^2]}{|f^H h_k|^2} \quad (42)$$

接下来, 提出引理 4 进一步导出 $\mathbb{E}[(v_k^{t+1})^2]$, $\forall k$ 的上界。

引理 4 若假设 4 成立, η_t 非增且对于任意 t 均满足 $\eta_{t-E} \leq 2\eta_t$, 则对于 $\forall k$ 和 $\forall t$, 有

$$\mathbb{E}[(v_k^t)^2] \leq \frac{4}{D}\eta_t^2(E-1)^2H^2 \quad (43)$$

$$\mathbb{E}[(v_k^t)^2] = \frac{1}{D}\mathbb{E}\left[\sum_{d=1}^D(\omega_k^t[d] - \frac{1}{D}\sum_{d'=1}^D\omega_k^t[d'])^2\right] =$$

$$\frac{1}{D}\mathbb{E}\left[\left(\sum_{d=1}^D(\omega_k^t[d])^2 - \frac{1}{D}\left(\sum_{d=1}^D\omega_k^t[d]\right)^2\right)\right] \stackrel{(a)}{\leq}$$

$$\frac{1}{D}\mathbb{E}\left[\left(\sum_{d=1}^D(\omega_k^t[d])^2\right)\right] = \frac{1}{D}\mathbb{E}[\|\mathbf{w}_t^k - \mathbf{w}_{t-E}^k\|_2^2] \stackrel{(b)}{\leq}$$

$$\frac{1}{D}\eta_{t-E}^2(E-1)^2H^2 \stackrel{(c)}{\leq} \frac{1}{D}4\eta_t^2(E-1)^2H^2 \quad (44)$$

其中, 不等式(a)源于 $\frac{1}{D}\left(\sum_{d=1}^D\omega_k^t[d]\right)^2 \geq 0$; 不等式(b)源于假设 4; 不等式(c)源于 $\eta_{t+E} \leq 2\eta_t$ 。因此, $\mathbb{E}[\|\bar{\mathbf{w}}_{t+1} - \mathbf{w}^*\|^2]$ 可表示为

$$\mathbb{E}[\|\bar{\mathbf{w}}_{t+1} - \mathbf{w}^*\|^2] \leq (1-\eta_t\mu)\mathbb{E}[\|\bar{\mathbf{w}}_t - \mathbf{w}^*\|^2] + \eta_t^2U \quad (45)$$

其中

$$U = \sum_{k=1}^K \frac{\sigma_k^2}{K^2} + 6L\Gamma + 8(E-1)^2 H^2 + \max_{k \in \mathcal{K}} \frac{4(E-1)^2 H^2 \sigma_n^2}{\frac{P_0 K^2}{|\mathbf{f}^H \mathbf{h}_k|^2}} \quad (46)$$

令 $\Delta_t = \mathbb{E} \left[\|\bar{\mathbf{w}}_t - \mathbf{w}^*\|^2 \right]$, 式(46)可写为

$$\Delta_{t+1} \leq (1 - \eta_t \mu) \Delta_t + \eta_t^2 U \quad (47)$$

令 $\eta_t = \frac{\lambda}{t + \gamma}$, 其中, $\lambda \geq \frac{1}{\mu}$ 且 $\gamma \geq 0$, 则

$\eta_t \leq \min \left\{ \frac{1}{\mu}, \frac{1}{4L} \right\} = \frac{1}{4L}$ 且 $\eta_t \leq 2\eta_{t+E}$ 。接下来, 将用归纳

法证明 $\Delta_t \leq \frac{\nu}{t + \gamma}$, 其中

$$\nu = \max \left\{ \frac{\lambda^2 U}{\lambda \mu - 1}, (\gamma + 1) \Delta_0 \right\} \quad (48)$$

首先, 参照 ν 定义, 当 $t = 0$ 时, $\Delta_0 \leq \frac{\nu}{\gamma}$ 。假设结论对

于某些 $t > 0$ 成立, 则

$$\begin{aligned} \Delta_{t+1} &\leq (1 - \eta_t \mu) \Delta_t + \eta_t^2 U = \\ &\left(1 - \frac{\lambda \mu}{t + \gamma} \right) \frac{\nu}{t + \gamma} + \frac{\lambda^2 U}{(t + \gamma)^2} = \\ &\frac{t + \gamma - 1}{(t + \gamma)^2} \nu + \left[\frac{\lambda^2 U}{(t + \gamma)^2} - \frac{\lambda \mu - 1}{(t + \gamma)^2} \right] \leq \frac{\nu}{t + 1 + \gamma} \end{aligned} \quad (49)$$

根据强凸性假设, 有

$$\mathbb{E} \left[\|F(\bar{\mathbf{w}}_t) - F^*\|^2 \right] \leq \frac{L}{2} \Delta_t \leq \frac{L}{2} \frac{\nu}{t + \gamma} \quad (50)$$

若令 $\lambda = \frac{2}{\mu}$, $\gamma = \max \left\{ 8 \frac{L}{\mu} - 1, E \right\}$, $\kappa = \frac{L}{\mu}$, 则

$\eta_t = \frac{2}{\mu} \frac{1}{t + \gamma}$ 。则

$$\begin{aligned} \nu &= \max \left\{ \frac{\lambda^2 U}{\lambda \mu - 1}, (\gamma + 1) \Delta_0 \right\} \stackrel{(a)}{\leq} \\ &\frac{\lambda^2 U}{\lambda \mu - 1} + (\gamma + 1) \Delta_0 = 4 \frac{U}{\mu^2} + (\gamma + 1) \Delta_0 \leq \\ &4 \frac{U}{\mu^2} + \left(8 \frac{L}{\mu} + E \right) \|\bar{\mathbf{w}}_0 - \mathbf{w}^*\|^2 \end{aligned} \quad (51)$$

因此有

$$\begin{aligned} \mathbb{E} \left[\|F(\bar{\mathbf{w}}_t) - F^*\|^2 \right] &\leq \frac{L}{2(t + \gamma)} \left[4 \frac{U}{\mu^2} + \left(8 \frac{L}{\mu} + E \right) \|\bar{\mathbf{w}}_0 - \mathbf{w}^*\|^2 \right] = \\ &\frac{2\kappa}{t + \gamma} \left[\frac{U}{\mu} + \left(2L + \frac{\mu E}{4} \right) \|\bar{\mathbf{w}}_0 - \mathbf{w}^*\|^2 \right] \end{aligned} \quad (52)$$

令 $t = T$, 证毕。

参考文献:

- [1] CHEN M Z, GÜNDÜZ D, HUANG K B, et al. Distributed learning in wireless networks: recent progress and future challenges[J]. IEEE Journal on Selected Areas in Communications, 2021, 39(12): 3579-3605.
- [2] LYU X C, REN C S, NI W, et al. Optimal online data partitioning for geo-distributed machine learning in edge of wireless networks[J]. IEEE Journal on Selected Areas in Communications, 2019, 37(10): 2393-2406.
- [3] LI H J, WANG R, ZHANG W, et al. One bit aggregation for federated edge learning with reconfigurable intelligent surface: analysis and optimization[J]. IEEE Transactions on Wireless Communications, 2023, 22(2): 872-888.
- [4] 温依霖, 赵乃良, 曾艳, 等. 基于本地模型质量的客户端选择方法[J]. 计算机工程, 2023, 49(6): 131-143.
WEN Y L, ZHAO N L, ZENG Y, et al. Client selection method based on local model quality[J]. Computer Engineering, 2023, 49(6): 131-143.
- [5] NGUYEN V-D, SHARMA S K, VU T X, et al. Efficient federated learning algorithm for resource allocation in wireless IoT networks[J]. IEEE Internet of Things Journal, 2021, 8(5): 3394-3409.
- [6] CHEN M Z, YANG Z H, SAAD W, et al. A joint learning and communications framework for federated learning over wireless networks[J]. IEEE Transactions on Wireless Communications, 2021, 20(1): 269-283.
- [7] WANG Z B, QIU J H, ZHOU Y, et al. Federated learning via intelligent reflecting surface[J]. IEEE Transactions on Wireless Communications, 2022, 21(2): 808-822.
- [8] LIU H, YUAN X J, ZHANG Y J A. Reconfigurable intelligent surface enabled federated learning: a unified communication-learning design approach[J]. IEEE Transactions on Wireless Communications, 2021, 20(11): 7595-7609.
- [9] WANG S Q, TUOR T, SALONIDIS T, et al. When edge meets learning: adaptive control for resource-constrained distributed machine learning[C]//Proceedings of IEEE Conference on Computer Communications. Piscataway: IEEE Press, 2018: 63-71.
- [10] MAO Y Y, YOU C S, ZHANG J, et al. A survey on mobile edge computing: the communication perspective[J]. IEEE Communications Surveys & Tutorials, 2017, 19(4): 2322-2358.
- [11] BRIK B, KSENTINI A, BOUAZIZ M. Federated learning for UAVs-enabled wireless networks: use cases, challenges, and open problems[J]. IEEE Access, 2020, 8: 53841-53849.
- [12] NGUYEN D C, DING M, PATHIRANA P N, et al. Federated learning for Internet of things: a comprehensive survey[J]. IEEE Communications Surveys & Tutorials, 2021, 23(3): 1622-1658.
- [13] SUN Y X, ZHOU S, NIU Z S, et al. Dynamic scheduling for over-the-air federated edge learning with energy constraints[J].

- IEEE Journal on Selected Areas in Communications, 2022, 40(1): 227-242.
- [14] LI L T, YANG L W, GUO X, et al. Delay analysis of wireless federated learning based on saddle point approximation and large deviation theory[J]. IEEE Journal on Selected Areas in Communications, 2021, 39(12): 3772-3789.
- [15] ZHU G X, WANG Y, HUANG K B. Broadband analog aggregation for low-latency federated edge learning[J]. IEEE Transactions on Wireless Communications, 2020, 19(1): 491-506.
- [16] YANG K, JIANG T, SHI Y M, et al. Federated learning via over-the-air computation[J]. IEEE Transactions on Wireless Communications, 2020, 19(3): 2022-2035.
- [17] FAN X, WANG Y, HUO Y, et al. Joint optimization of communications and federated learning over the air[J]. IEEE Transactions on Wireless Communications, 2022, 21(6): 4434-4449.
- [18] SERY T, SHLEZINGER N, COHEN K, et al. Over-the-air federated learning from heterogeneous data[J]. IEEE Transactions on Signal Processing, 2021, 69: 3796-3811.
- [19] LI X, HUANG K, YANG W, et al. On the convergence of FedAvg on non-IID data[J]. arXiv Preprint, arXiv: 1907.02189, 2019.
- [20] LI T, SAHU A K, ZAHEER M, et al. Federated optimization in heterogeneous networks[J]. arXiv Preprint, arXiv: 1812.06127, 2018.
- [21] ZHENG S H, SHEN C, CHEN X. Design and analysis of uplink and downlink communications for federated learning[J]. IEEE Journal on Selected Areas in Communications, 2021, 39(7): 2150-2167.
- [22] ZHAO Y, LI M, LAI L, et al. Federated learning with non-IID data[J]. arXiv Preprint, arXiv: 1806.00582, 2018.
- [23] SCUTARI G, FACCHINEI F, SONG P R, et al. Decomposition by partial linearization: parallel optimization of multi-agent systems[J]. IEEE Transactions on Signal Processing, 2014, 62(3): 641-656.
- [24] XIAO H, RASUL K, VOLLGRAF R. Fashion-MNIST: a novel image dataset for benchmarking machine learning algorithms[J]. arXiv Preprint, arXiv: 1708.07747, 2017.
- [25] ZHU G X, CHEN L, HUANG K B. MIMO over-the-air computation: beamforming optimization on the Grassmann manifold[C]//Proceedings of 2018 IEEE Global Communications Conference (GLOBECOM). Piscataway: IEEE Press, 2019: 1-6.
- [26] LUO Z Q, MA W K, SO A M C, et al. Semidefinite relaxation of quadratic optimization problems[J]. IEEE Signal Processing Magazine, 2010, 27(3): 20-34.

[作者简介]



王平（1978-），男，湖北黄冈人，博士，同济大学副教授、博士生导师，主要研究方向为车联网。

杨志伟（1982-），男，河南许昌人，同济大学博士生，主要研究方向为车路云协同智能驾驶。

李贺举（1995-），男，安徽阜阳人，同济大学博士生，主要研究方向为无线通信和联邦学习。