



D2D 辅助的双阶段空中计算模型聚合方案

张冉强¹, 邓娜^{1*}, 卫海超², 邢成文³, 赵楠¹

1. 大连理工大学信息与通信工程学院, 大连 116024

2. 大连海事大学信息科学技术学院, 大连 116026

3. 北京理工大学信息与电子学院, 北京 100081

* 通信作者. E-mail: dengna@dlut.edu.cn

收稿日期: 2023-12-21; 修回日期: 2024-02-24; 接受日期: 2024-06-17; 网络出版日期: 2024-08-23

国家自然科学基金 (批准号: 62201115, 62325103, 62371086)、辽宁省自然科学基金联合基金 (批准号: 2023-MSBA-015)、航空科学基金 (批准号: 2022Z001063001) 和中央高校基本科研业务费专项资金 (批准号: DUT24MS015, 3132024246) 资助项目

摘要 在无线联邦学习中, 不同位置的设备信道质量差异性使得基于空中计算的联邦学习的模型聚合误差由信道质量最差的设备所主导. 为此提出了一种终端直通 (device-to-device, D2D) 辅助的空中计算联邦学习方案, 其中选择信道质量好的辅助设备辅助边缘设备将本地模型更新至服务器. 构建了最小化均方误差的优化问题, 提出了一种交替优化的算法, 以低复杂度优化所有设备和基站的运行参数. 为了评估该方案的性能, 本文通过理论分析验证了所提方案相比于传统的基于空中计算的模型聚合方案具有优势. 同时, 设计了两种不同的设备分布场景, 并基于神经网络和真实数据集构建了联邦学习实验. 结果表明, 所提算法能够很快收敛, 与传统的空中计算方案以及现有的基于调度和基于中继的空中计算方案相比, 该方案能够显著减小模型聚合误差, 并提高联邦学习的预测精度.

关键词 D2D, 联邦学习, 空中计算, 交替优化算法

1 引言

当前, 机器学习广泛应用推动了人工智能技术的广泛发展^[1], 例如图像识别和自然语言处理等^[2]. 典型的机器学习过程包含训练过程和推理过程两个部分, 全部由云计算提供支持, 即通过集中式的云数据中心实现大规模计算、存储以及对整个数据集的访问. 然而, 新兴的智能移动设备和高风险应用, 如无人机、智能汽车和增强现实等, 对低时延和隐私保护提出了关键性需求, 而基于云计算的机器学习方法已无法满足^[3]. 因此, 设备直接在本地数据上进行训练和推理的方式越来越有吸引力, 不需要向云或网络发送大量原始数据. 这种新兴技术被称为边缘机器学习^[4], 由移动边缘计算技术提供支持将云计算能力下推至网络边缘^[5,6]. 尽管移动边缘智能服务具有很多优势, 但其主要瓶颈在于计算、存储、能源和带宽等资源的有限性. 为解决这一问题, 现有研究开始通过硬件和软件的协同设计以及使

引用格式: 张冉强, 邓娜, 卫海超, 等. D2D 辅助的双阶段空中计算模型聚合方案. 中国科学: 信息科学, 2024, 54: 2487–2502, doi: 10.1360/SSI-2023-0402
Zhang R Q, Deng N, Wei H C, et al. D2D-assisted two-stage model aggregation scheme based on over-the-air computation (in Chinese). Sci Sin Inform, 2024, 54: 2487–2502, doi: 10.1360/SSI-2023-0402

用模型压缩技术, 来降低推理过程中的存储、时间和能耗消耗^[7,8]. 另外, 还提出了各种先进的分布式优化算法^[9~11], 利用分布在多个设备上的计算能力和数据, 实现资源的有效协同, 以加速训练过程.

联邦学习是一种分布式机器学习方法, 允许多个参与方在保护数据隐私的前提下, 共同建立一个全局模型^[12~14]. 通过在本地设备上进行模型训练, 参与方能够有效地合作, 从而充分利用各方的数据资源, 提高模型性能, 同时保护敏感数据的安全. 当在无线网络中实现联邦学习时, 基站与设备之间模型交互是通过无线链路实现的, 因此有限的无线资源成为这种大模型交互共享的瓶颈. 在联邦学习的模型聚合环节中, 需要将不同设备训练的参数进行加权平均. 考虑到这个操作与多址接入信道波形叠加特性类似, 空中计算是一种实现无线网络中联邦学习模型聚合的高效通信方法^[15~17]. 空中计算所需的带宽和通信延迟与设备数量无关, 显著增强了系统可扩展性, 提高了通信效率. 然而, 空中计算不可避免地受到噪声和无线信道衰落造成的失真影响. 为了使上传的本地模型在接收端对齐, 信道较好的设备必须降低发射功率, 导致模型聚合误差由信道最差的设备所主导^[18]. 因此, 如何设计基于空中计算的模型聚合方案以提高联邦学习性能值得深入研究.

为了减小空中计算联邦学习的模型聚合误差, 一种思路是放弃信道质量差的设备所训练出的模型. 文献 [19] 提出了一种基于信道条件的调度方案来排除信道条件最差的设备. 文献 [15] 考虑联合设备选择和波束成形设计, 在目标均方误差 (mean square error, MSE) 约束下最大化所选设备的数量. 文献 [20] 提出了一种基于截断的功率控制方法来丢弃具有深度衰落信道的设备. 文献 [21] 提出同时利用 CSI 和数据重要性联合中心接收机设计和用户调度的方案, 建立了最小化聚合信号失真的非凸优化问题, 并将该问题解耦为两个子问题进行求解. 文献 [22] 提出了一种动态设备调度机制, 通过功率控制策略选择符合条件的终端设备上传其局部模型参与全局模型聚合. 虽然这些方案在模型上传时丢弃信道质量差的设备可以有效控制空中计算的误差, 然而, 从训练中丢弃设备往往会减少参与模型训练的数据的数量, 则不可避免地降低联邦学习的模型收敛速度, 并损害最终聚合模型的推理性能, 尤其当用户具有非独立同分布的数据时, 该问题尤为严重^[23].

另一种思路是通过先进的通信技术来提升信道质量差的设备的通信质量. 文献 [24] 提出了一种两层中继辅助联邦学习框架, 其中局部模型通过空中计算上传到中继节点, 中继节点再通过空中计算转发给基站, 这种设计利用中继提高了边缘设备的通信质量. 但是所提方案假设每个设备只连接一个中继节点, 无法利用中继网络协同分集的优势. 文献 [18] 部署了多个半双工中继, 协助设备将本地模型更新到边缘服务器, 并提出了一种基于交替优化的算法, 优化接收端、中继和设备的收发参数以最小化模型失真, 这种设计通过中继和协同分集有效提升了边缘设备的通信质量. 然而引入中继设备会增加网络的复杂性, 需要额外的设备和安装维护成本, 特别是中继设备的数量和布置需要精心规划和管理. 文献 [25] 采用分层联邦学习架构, 利用边缘服务器组织终端设备形成簇, 并在全局模型聚合之前在每个簇中先进行簇模型聚合, 以适应大规模联邦学习训练. 文献 [26] 提出了一个中继辅助的大规模联邦学习框架, 并研究了在功耗和 MSE 约束下联邦学习系统中的设备调度问题. 作者制定了一个联合设备调度和功率分配问题, 通过将原始的非凸优化问题转化为多个稀疏优化问题来处理. 并提出设备调度算法成功解决了这些稀疏子问题以最大化参与调度的设备数量. 文献 [27] 考虑采用移动设备构建中继网络, 以提高联邦学习的效率. 其中, 移动设备帮助彼此将模型以多跳中继的方式更新上传到基站, 该方案充分利用中继方式提高了网络的整体通信质量. 但是该系统中每个设备的数据只传输给一个移动设备, 没有有效利用中继网络的协同分集的优势, 并且该方案没有以空中计算作为模型聚合方式, 导致了较大的通信时延和基站模型聚合时延. 以上方案存在训练数据减少、无法利用中继网络协同分集优势、中继网络复杂度高, 以及模型聚合时延大等问题.

针对高路损和深衰落问题, 本文采用终端直通 (device-to-device, D2D) 技术来改善设备的通信质

表 1 策略对比
Table 1 Strategy comparison

Related work	Enhancement techniques	Training devices	Feature
[15]	Joint beamforming and scheduling	Part	Incomplete training data
[18]	Relay-assisted aggregation	All	Extra relay station
[19]	Channel quality-based scheduling	Part	Incomplete training data
[20]	Truncated power control-based scheduling	Part	Incomplete training data
[21]	CSI and data importance-based scheduling	Part	Incomplete training data
[22]	Joint power allocation and scheduling	Part	Incomplete training data
[24]	Relay-assisted aggregation	All	Extra relay station
[25]	Hierarchical structure	Part	Extra intermediate sever
[26]	Relay-assisted aggregation and joint power allocation and scheduling	Part	Extra relay station, incomplete training data
[27]	Multi-hop relay-assisted aggregation	All	Extra relay station
Our paper	The proposed scheme	All	Complete data, no extra device

量, 其中 D2D 是蜂窝网络中邻近设备之间不经由基站转发而直接交互信息的一种通信技术. 3GPP 组织提出了一种名为 Sidelink 的 D2D 通信技术, 并纳入到 Release 12 及后续通信标准 [28]. 因此, 为了使小区内的所有设备都能参与到联邦学习训练中并保证空中计算的准确性, 本文提出 D2D 辅助的双阶段空中计算模型聚合方案, 辅助高路损或深衰落设备进行模型上传, 降低聚合模型的 MSE, 从而提升联邦学习性能. 本文方案与现有方案的对比如表 1 所示. 首先在第一阶段, 信道质量差的用户同时向多个信道质量好的用户以及基站发送模型数据, 在第二阶段, 信道质量好的设备聚合本设备的模型数据和收到的模型数据, 然后发送给基站, 基站聚合两个阶段收到的模型数据得到本轮联邦学习的聚合模型, 该方案充分利用了无线联邦学习系统的协作多样性, 提高了通信效率和联邦学习模型训练速度. 为了降低空中计算联邦学习的模型聚合误差, 分析了影响均方误差的关键因素, 形成了最小化 MSE 的优化问题, 并提出了一种交替优化算法用于优化设备发射系数、加权因子和基站的降噪接收系数以有效降低模型聚合误差. 通过理论分析, 本文确定了使所提方案的 MSE 不大于传统方案的充分条件, 从而在理论上证明了所提方案的性能优势. 同时, 设计了两种不同的设备分布场景以及构建了一个基于空中计算的联邦学习任务. 结果表明该算法具有较快的收敛速度, 与传统的空中计算方案以及现有的基于调度和基于中继的空中计算方案相比, 显著降低了模型聚合误差, 提高联邦学习预测精度.

2 系统模型与问题建模

2.1 传统的空中计算联邦学习

如图 1(a) 所示, 考虑一个传统的空中计算联邦学习系统, 其中部署了 1 个基站和 M 个设备. 设 $\mathcal{M} = \{1, \dots, M\}$ 表示设备集合, 通过使用本地数据协同训练共享的学习模型. 通常联邦学习的目标函数为

$$\min_{\mathbf{w} \in \mathbb{C}^\xi} \mathcal{F}(\mathbf{w}) = \sum_{m \in \mathcal{M}} \rho_m \mathcal{F}_m(\mathbf{w}), \quad (1)$$

其中 $\mathbf{w} \in \mathbb{C}^\xi$ 为模型参数向量, $\rho_m = \frac{|\mathcal{D}_m|}{\sum_{m \in \mathcal{M}} |\mathcal{D}_m|}$ 是设备 m 的权重, \mathcal{D}_m 表示设备 m 中的数据集合,

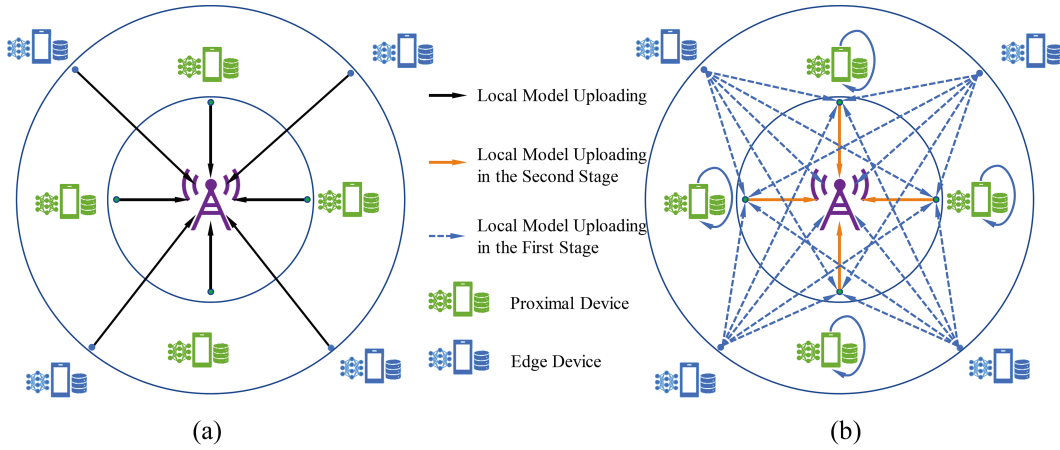


图 1 (网络版彩图) (a) 传统的和 (b) D2D 辅助的双阶段空中计算联邦学习系统

Figure 1 (Color online) Over-the-air computation based federated learning systems. (a) Traditional federated learning system; (b) D2D-assisted two-stage federated learning system

$|\mathcal{D}_m|$ 表示设备 m 所拥有的数据的数量. 设备 m 处的局部经验损失函数表示为

$$\mathcal{F}_m(\mathbf{w}) = \frac{1}{|\mathcal{D}_m|} \sum_{(\mathbf{x}_j, \mathbf{y}_j) \in \mathcal{D}_m} \mathcal{L}(\mathbf{w}; \mathbf{x}_j, \mathbf{y}_j), \quad (2)$$

其中 \mathbf{x}_j 和 \mathbf{y}_j 分别为特征向量和对应的标签, $\mathcal{L}(\mathbf{w}; \mathbf{x}_j, \mathbf{y}_j)$ 是单个样本的损失函数.

为了解决 (1) 中的最小化问题, 可以采用联邦平均算法^[29], 该算法通过对各个局部模型的参数求加权平均, 实现全局模型的更新. 空中计算技术利用多址信道的波形叠加特点, 可以高效地实现联邦平均算法. 具体来说, 参与联邦学习的多个设备同时发送各自的局部模型信号, 这些信号在空中直接叠加, 以加权平均的形式聚合到基站, 即完成了全局模型的更新. 在这个过程中, 每个设备的发射信号都有对应的实际权重系数参与最终的加权求和运算, 其中, 每个设备的实际权重系数由发射系数、信道系数和接收系数组成. 联邦学习需要多轮循环操作, 假设当前为第 r 轮, 则每轮循环的操作如下.

(1) 全局模型分发: 基站将当前全局模型 $\mathbf{w}^{[r-1]}$ 通过无线链路广播给所有设备 \mathcal{M} . 由于蜂窝系统下行资源丰富, 假设下行传输总是成功的, 即所有设备都能无差错地接收到全局模型参数进行局部训练.

(2) 局部模型更新: 在接收到全局模型 $\mathbf{w}^{[r-1]}$ 的基础上, 每个设备 $m \in \mathcal{M}$ 运行局部更新算法, 如随机梯度下降算法, 借助局部数据集生成更新后的局部模型 $\mathbf{w}_m^{[r]}$.

(3) 局部模型聚合: 参与联邦学习的设备同时将更新后的局部模型通过用于空中计算的时频资源发送给基站^[30]. 在不考虑信道衰落和噪声的理想通信情况下, 基站得到的聚合模型为

$$\mathbf{w}^{[r]} = \sum_{m \in \mathcal{M}} \rho_m \mathbf{w}_m^{[r]}. \quad (3)$$

然而, 由于无线通信的衰落和噪声, 设备上传 $\mathbf{w}_m^{[r]}$ 不可避免地会产生误差.

为体现实际无线环境的复杂性, 本文采用 Nakagami 模型, 通过形状参数 β 的调节能够描述不同信道衰落分布特性. 任意两个节点 i 和 j 之间的路径损耗为 $d_{i,j}^{-\alpha}$, 其中 $d_{i,j}$ 表示节点 i 与 j 之间的距离, α 表示路径损耗指数, 小尺度衰落系数 $l_{i,j}$ 的概率密度函数为 $f(x; \beta, \omega) = \frac{2\beta^\beta}{\Gamma(\beta)\omega^\beta} x^{2\beta-1} e^{-\frac{\beta}{\omega} x^2}$, 信道系数 h 表示为 $l_{i,j} d_{i,j}^{-\alpha/2}$. 假设基站和设备都使用单天线进行信号的收发, 并且各个设备的信号之间统计独立. 此外, 假设设备在给定的单个时隙内保持静止.

为了对抗路径损耗和小尺度衰落,每个设备都需要进行功率控制.因此,先用线性函数把 $\mathbf{w}_m^{[r]}$ 归一化为具有零均值和单位方差的 $\mathbf{s}_m^{[r]} \triangleq \Psi(\mathbf{w}_m^{[r]})$, 即 $\mathbb{E}(\mathbf{s}_m^{[r]}(\mathbf{s}_m^{[r]})^H) = \mathbf{I}$, $\mathbb{E}(\mathbf{s}_m^{[r]}(\mathbf{s}_{m'}^{[r]})^H) = \mathbf{0}$, $\forall m, m' \in \mathcal{M}$, $m \neq m'$. 然后,调整发射和接收系数以接近理想的聚合模型,在每一轮中,假设模型向量 $\mathbf{s}_m^{[r]}$ 中的每一个元素 $s_{m(t)}^{[r]}$, $t \in \{1, \dots, \xi\}$ 需要一个时隙进行传输,而空中计算的目标函数为

$$x_{(t)}^{[r]} = \sum_{m \in \mathcal{M}} \rho_m s_{m(t)}^{[r]}. \quad (4)$$

为了便于分析,将 $x_{(t)}^{[r]}$ 和 $s_{m(t)}^{[r]}$ 分别简化为 x 和 s . 基站接收到的空中计算信号为

$$y = \sum_{m \in \mathcal{M}} h_m a_m s_m + z, \quad (5)$$

其中 $h_m \in \mathbb{C}$ 为设备 m 与基站之间的信道系数, $a_m \in \mathbb{C}$ 为设备 m 的复值发射系数, $z \sim \mathcal{CN}(0, \sigma^2)$ 是加性高斯 (Gauss) 白噪声. x 在基站处的估计值是

$$\hat{x} = cy = c \sum_{m \in \mathcal{M}} h_m a_m s_m + cz, \quad (6)$$

其中 $c \in \mathbb{C}$ 是基站的去噪接收系数. 则可得 w 的估计为 $\hat{\mathbf{w}}^{[r]} = \Psi^{-1}(\hat{\mathbf{x}}^{[r]})$.

(4) 全局模型更新: 基站将全局模型更新为 $\hat{\mathbf{w}}^{[r]}$.

(5) 当满足预定的终止条件时, 停止训练. 否则, $r = r + 1$, 重复步骤 (1)~(5), 进入下一轮循环.

为了衡量空中计算带来的模型失真, 可以计算估计值 \hat{x} 和目标值 x 之间的 MSE, 根据文献 [18], 可得 MSE 的最小值为

$$\text{MSE}^*(\hat{x}, x) = \frac{\sigma^2}{P_0} \max_{m \in \mathcal{M}} \frac{|\rho_m|^2}{|h_m|^2}, \quad (7)$$

其中 P_0 为设备发射功率约束. 该结果表明空中计算联邦学习的模型聚合误差主要取决于信道质量差的小区边缘设备.

2.2 D2D 辅助的双阶段空中计算联邦学习

为了减小信道质量差的小区边缘设备对模型聚合误差的负面影响, 提出一种 D2D 辅助的双阶段空中计算聚合方案, 以提高基站的接收信号相对于空中计算的目标函数的估计精度. 如图 1(b) 所示, 将 M 个设备分为两类, 即 N 个辅助设备和 K 个边缘设备. 本文在仿真部分考虑了两种分类方式, 在场景一中根据设备与基站的距离进行辅助设备和边缘设备的分类, 在场景二中根据信道质量进行辅助设备和边缘设备的分类. 设 $\mathcal{N} = \{1, \dots, N\}$, $\mathcal{K} = \{1, \dots, K\}$ 分别表示辅助设备集和边缘设备集, 并且 $\mathcal{M} = \mathcal{N} \cup \mathcal{K}$, 假设所有设备都是半双工的. 相应地, 将每一轮通信中的局部模型上传分为两个阶段. 在第一阶段, 边缘设备通过空中计算链路并发地将模型信号发送给辅助设备和基站. 在第二阶段, 辅助设备通过空中计算链路并发地向基站发送自己的模型信号和接收到的信号.

在第一阶段, 所有边缘设备调整发射功率后将其信号 s_k 同时发送给所有辅助设备和基站, 设 $a_k \in \mathbb{C}$ 为第一阶段设备 k 处的复值发射系数, 该系数用于调节 s_k 的权重, 在辅助设备 n 处的接收信号和自有信号的混合信号为

$$r_n = \sum_{k=1}^K g_{k,n} a_k s_k + d_n s_n + z_n, \quad (8)$$

其中 $g_{k,n} \in \mathbb{C}$ 表示边缘设备 $k \in \mathcal{K}$ 与辅助设备 $n \in \mathcal{N}$ 之间的信道系数, d_n 为辅助设备 n 处自有信号的放缩因子, 该放缩因子用于调节自有信号 s_n 与接收到的信号 $\sum_{k=1}^K g_{k,n} a_k s_k$ 的比例, $z_n \sim \mathcal{CN}(0, \sigma^2)$ 是辅助设备 n 处的加性高斯白噪声. 同时, 基站处接收到的信号为

$$y_1 = \sum_{k=1}^K h_k a_k s_k + z_1, \quad (9)$$

其中 $h_k \in \mathbb{C}$ 表示边缘设备 k 与基站之间的信道系数, $z_1 \sim \mathcal{CN}(0, \sigma^2)$ 是第一阶段的加性高斯白噪声.

在第二阶段, 每个辅助设备 n 将混合信号乘以复值系数 $b_n \in \mathbb{C}$ 进行放大后发送给基站, b_n 用于调节自有信号 s_n 以及接收到的信号 $\sum_{k=1}^K g_{k,n} a_k s_k$ 的权重, 则基站的接收信号为

$$y_2 = \sum_{n=1}^N f_n b_n r_n + z_2 = \sum_{n=1}^N f_n b_n \left(\sum_{k=1}^K g_{k,n} a_k s_k + d_n s_n + z_n \right) + z_2, \quad (10)$$

其中 $f_n \in \mathbb{C}$ 表示辅助设备 n 与基站之间的信道系数, $z_2 \sim \mathcal{CN}(0, \sigma^2)$ 是第二阶段的加性高斯白噪声. 由于每个设备都有最大功率限制, 边缘设备和辅助设备的功率约束分别为

$$\mathbb{E}(|a_k s_k|^2) = |a_k|^2 \leq P_0, \forall k \in \mathcal{K}, \quad (11)$$

$$\mathbb{E}(|b_n r_n|^2) = |b_n|^2 \left(\sum_{k=1}^K |g_{k,n}|^2 |a_k|^2 + |d_n|^2 + \sigma^2 \right) \leq P_0, \forall n \in \mathcal{N}. \quad (12)$$

基站对 y_1 和 y_2 分别施加去噪接收系数 $c_1 \in \mathbb{C}$ 和 $c_2 \in \mathbb{C}$, 得到加权和 x 的估计为 \hat{x}' , 表示为

$$\begin{aligned} \hat{x}' &= c_1 y_1 + c_2 y_2 \\ &= \sum_{k=1}^K \left(c_1 h_k a_k + c_2 a_k \sum_{n=1}^N f_n b_n g_{k,n} \right) s_k + \sum_{n=1}^N c_2 f_n b_n d_n s_n + c_1 z_1 + c_2 z_2 + c_2 \sum_{n=1}^N f_n b_n z_n. \end{aligned} \quad (13)$$

将上述方案称为 D2D 辅助的双阶段空中计算模型聚合. \hat{x}' 与 x 之间的均方误差为

$$\begin{aligned} \text{MSE} &= \mathbb{E}(|\hat{x}' - x|^2) \\ &= \sum_{k=1}^K \left| c_1 h_k a_k + c_2 a_k \sum_{n=1}^N f_n b_n g_{k,n} - \rho_k \right|^2 + \sum_{n=1}^N |c_2 f_n b_n d_n - \rho_n|^2 \\ &\quad + \left(|c_1|^2 + |c_2|^2 + |c_2|^2 \sum_{n=1}^N (|f_n|^2 |b_n|^2) \right) \sigma^2. \end{aligned} \quad (14)$$

由上式可知, MSE 是 $K + N + 1$ 项的线性组合, 其中前 K 项表征了 K 个边缘设备的局部模型由信道衰落引起的误差, 中间 N 项体现了 N 个辅助设备的局部模型由信道衰落引起的误差, 最后一项表示两阶段模型上传方案中由通信噪声引起的误差.

2.3 问题建模

如文献 [23] 所述, 模型聚合的 MSE 在收敛速度和模型误差 (实际模型与理想模型的差异) 两方面对联邦学习性能都有重要影响. 较小的 MSE 可以实现较快的收敛速度和较小的模型误差. 基于此, 通过优化设备的发射系数 $\{a_k, \forall k; b_n, d_n, \forall n\}$ 和基站的去噪因子 $\{c_1, c_2\}$, 以最小化 (14) 中的 MSE, 即

$$\min_{a_k, \forall k; b_n, d_n, \forall n; c_1, c_2} \text{MSE} \quad (15a)$$

$$\text{s.t.} \quad |a_k|^2 \leq P_0, \forall k \in \mathcal{K}, \quad (15b)$$

$$|b_n|^2 \left(\sum_{k=1}^K |g_{k,n}|^2 |a_k|^2 + |d_n|^2 + \sigma^2 \right) \leq P_0, \forall n \in \mathcal{N}. \quad (15c)$$

由于目标函数 MSE 和约束 (15c) 中都存在耦合变量, 问题 (15) 是非凸的, 导致该问题难以直接求解, 因此本文提出一种低复杂度的算法来有效地解决这个问题.

3 交替优化以最小化 MSE

为了降低问题的复杂度, 提出了一种交替最小化方法, 交替优化第一阶段边缘设备的发射系数 $\{a_k\}$ 与第二阶段辅助设备的发射系数 $\{b_n\}$, 第一阶段辅助设备的加权因子 $\{d_n\}$, 基站的第一阶段降噪接收系数 $\{c_1\}$ 与第二阶段降噪接收系数 $\{c_2\}$, 实现耦合变量的解耦.

3.1 给定 $\{b_n, c_1, c_2\}$, 优化 $\{a_k, d_n\}$

在本小节中, 对于给定的 $\{b_n, c_1, c_2\}$, 优化第一阶段中边缘设备上的发射系数 $\{a_k\}$ 和辅助设备的信号加权因子 $\{d_n\}$. 在这种情况下, 将式 (15) 中的 MSE 最小化问题简化为

$$\min_{a_k, \forall k; d_n, \forall n} \sum_{k=1}^K \left| c_1 h_k a_k + c_2 a_k \sum_{n=1}^N f_n b_n g_{k,n} - \rho_k \right|^2 + \sum_{n=1}^N |c_2 f_n b_n d_n - \rho_n|^2 \quad (16a)$$

$$\text{s.t.} \quad |a_k|^2 \leq P_0, \forall k \in \mathcal{K}, \quad (16b)$$

$$|b_n|^2 \left(\sum_{k=1}^K |g_{k,n}|^2 |a_k|^2 + |d_n|^2 + \sigma^2 \right) \leq P_0, \forall n \in \mathcal{N}. \quad (16c)$$

问题 (16) 是一个凸二次约束二次规划问题, 可以使用标准凸优化求解工具, 如 CVX^[31] 进行求解.

3.2 给定 $\{a_k, d_n, c_1, c_2\}$, 优化 $\{b_n\}$

对于给定的 $\{a_k, d_n, c_1, c_2\}$, 通过求解以下简化的 MSE 最小化问题来优化第二阶段辅助设备上的发射系数, 即 $\{b_n\}$:

$$\min_{b_n, \forall n} \sum_{k=1}^K \left| c_1 h_k a_k + c_2 a_k \sum_{n=1}^N f_n b_n g_{k,n} - \rho_k \right|^2 + \sum_{n=1}^N |c_2 f_n b_n d_n - \rho_n|^2 + \left(|c_2|^2 \sum_{n=1}^N |f_n|^2 |b_n|^2 \right) \sigma^2$$

$$\text{s.t.} \quad |b_n|^2 \left(\sum_{k=1}^K |g_{k,n}|^2 |a_k|^2 + |d_n|^2 + \sigma^2 \right) \leq P_0, \forall n \in \mathcal{N}. \quad (17a)$$

问题 (17) 是一个凸二次约束二次规划问题, 可以使用 CVX 进行求解.

3.3 给定 $\{a_k, b_n, d_n\}$, 优化 $\{c_1, c_2\}$

对于给定 $\{a_k, b_n, d_n\}$, 通过求解以下简化的 MSE 最小化问题来优化去噪接收系数 $\{c_1, c_2\}$:

$$\min_{c_1, c_2} \sum_{k=1}^K \left| c_1 h_k a_k + c_2 a_k \sum_{n=1}^N f_n b_n g_{k,n} - \rho_k \right|^2 + \sum_{n=1}^N |c_2 f_n b_n d_n - \rho_n|^2$$

算法 1 Alternating minimization for problem (15)

```

1: Input:  $J_{\max}$  and  $\varepsilon$ ;
2: Initialization  $a_k^{(0)} = 1, \forall k; b_n^{(0)} = 1, d_n^{(0)} = 1, \forall n; c_1^{(0)} = 1, c_2^{(0)} = 1; j = 1$ ;
3: Compute  $\text{MSE}^{(0)}$  by  $\{a_k^{(0)}, \forall k; b_n^{(0)}, d_n^{(0)}, \forall n; c_1^{(0)}, c_2^{(0)}\}$ ;
4: while  $j \leq J_{\max}$  do
5:   Compute  $\{a_k^{(j)}, d_n^{(j)}\}$  by solving (16) with  $\{b_n^{(j-1)}, c_1^{(j-1)}, c_2^{(j-1)}\}$ ;
6:   Compute  $\{b_n^{(j)}\}$  by solving (17) with  $\{a_k^{(j)}, d_n^{(j)}, c_1^{(j-1)}, c_2^{(j-1)}\}$ ;
7:   Compute  $\{c_1^{(j)}, c_2^{(j)}\}$  by solving (18) with  $\{a_k^{(j)}, b_n^{(j)}, d_n^{(j)}\}$ ;
8:   Compute  $\text{MSE}^{(j)}$  by  $\{a_k^{(j)}, \forall k; b_n^{(j)}, d_n^{(j)}, \forall n; c_1^{(j)}, c_2^{(j)}\}$ ;
9:   if  $\frac{|\text{MSE}^{(j)} - \text{MSE}^{(j-1)}|}{|\text{MSE}^{(j)}|} < \varepsilon$  then
10:     break;
11:   end if
12:    $j++$ ;
13: end while
14:  $a_k^* = a_k^{(j)}, \forall k; b_n^* = b_n^{(j)}, \forall n; d_n^* = d_n^{(j)}, \forall n; c_1^* = c_1^{(j)}; c_2^* = c_2^{(j)}$ ;
15: Output:  $\{a_k^*, \forall k; b_n^*, d_n^*, \forall n; c_1^*, c_2^*\}$ ;
    
```

$$+ \left(|c_1|^2 + |c_2|^2 + |c_2|^2 \sum_{n=1}^N (|f_n|^2 |b_n|^2) \right) \sigma^2. \quad (18)$$

问题 (18) 是一个无约束二次规划问题. 当关于 $\{c_1, c_2\}$ 的梯度为 0 时, 得到最优解, 即

$$\sum_{k=1}^K \left(c_1 h_k a_k + c_2 a_k \sum_{n=1}^N f_n b_n g_{k,n} - \rho_k \right) \bar{h}_k \bar{a}_k + c_1 \sigma^2 = 0, \quad (19)$$

$$\sum_{k=1}^K \left(c_1 h_k a_k + c_2 a_k \sum_{n=1}^N f_n b_n g_{k,n} - \rho_k \right) \bar{a}_k \sum_{n=1}^N \bar{f}_n \bar{b}_n \bar{g}_{k,n} + \sum_{n=1}^N (c_2 f_n b_n d_n - \rho_n) \bar{f}_n \bar{b}_n \bar{d}_n + c_2 \sigma^2 \left(1 + \sum_{n=1}^N (|f_n|^2 |b_n|^2) \right) = 0, \quad (20)$$

其中 \bar{h}_k 表示 h_k 的共轭. 求解 (19) 和 (20) 组成的二元一次方程组即可得到最优解 c_1^* 和 c_2^* .

3.4 交替优化算法分析

算法 1 给出了求解问题 (15) 的总体交替最小化算法. 初始化 $\{a_k, \forall k; b_n, d_n, \forall n; c_1, c_2\}$. 当达到最大迭代次数 J_{\max} 或对目标值的改进小于预定阈值 ε 时, 该算法终止, 得到参数输出 $\{a_k^*, \forall k; b_n^*, d_n^*, \forall n; c_1^*, c_2^*\}$. 算法 1 的收敛性如下所示: 定义第 j 轮迭代的目标值为 $\text{MSE}\{a_k^{(j)}, d_n^{(j)}, b_n^{(j)}, c_1^{(j)}, c_2^{(j)}\}$, 在第 j 轮迭代时, 由于问题 (16)~(18) 是凸优化问题, 可分别得到其最优解, 因而有如下关系式:

$$\text{MSE}\{a_k^{(j-1)}, d_n^{(j-1)}, b_n^{(j-1)}, c_1^{(j-1)}, c_2^{(j-1)}\} \geq \text{MSE}\{a_k^{(j)}, d_n^{(j)}, b_n^{(j-1)}, c_1^{(j-1)}, c_2^{(j-1)}\} \quad (21)$$

$$\geq \text{MSE}\{a_k^{(j)}, d_n^{(j)}, b_n^{(j)}, c_1^{(j-1)}, c_2^{(j-1)}\} \quad (22)$$

$$\geq \text{MSE}\{a_k^{(j)}, d_n^{(j)}, b_n^{(j)}, c_1^{(j)}, c_2^{(j)}\}. \quad (23)$$

式 (21)~(23) 分别对应算法 1 的第 5~7 行, 由式 (21)~(23) 可知, 算法 1 在每次迭代中都是非递增的. 此外, 算法 1 的目标值 $\text{MSE} = \mathbb{E}(|\hat{x}' - x|^2) \geq 0$ 具有一个有限下界, 使得算法 1 是收敛的 [32].

由于 CVX 采用内点法求解凸二次约束二次规划问题, 因此算法 1 的第 5 和 6 行的复杂度上限分别为 $\mathcal{O}(K^3 + N^3)$ 和 $\mathcal{O}(N^3)$ [18], 由于 c_1^* 和 c_2^* 可以通过直接求解式 (19) 和 (20) 组成的二元一次方程

组得到其解的表达式,因此算法1的第7行的复杂度为 $\mathcal{O}(1)$.综合以上分析,本文提出的交替优化算法的总体复杂度上限为 $\mathcal{O}(J_{\max}(K^3 + N^3))$,这表明算法1能够在多项式时间内解决原始优化问题.由于原始优化问题的非凸性使得最优解难以分析求解,因此算法1在迭代过程中会趋于稳定点.大量实验结果表明,这个稳定点的性能较好.后续的仿真进一步验证了该算法性能要优于基线算法.本文所提出的交替优化算法及使用CVX解决子问题只是解决该优化问题的一种途径.也可以结合其他优化方法(如自适应学习率、正则化项等)以进一步提升算法性能.

3.5 D2D 辅助的双阶段空中计算模型聚合方案的性能分析

本小节将给出所提方案性能的理论分析,证明其MSE相较于直接进行空中计算模型聚合的传统方案更小.为了便于分析,本小节只考虑一个辅助设备,即 $N=1$,则 $f_n, b_n, g_{k,n}$ 分别简化为 f, b, g_k ,此外,假设所有设备具有相同的本地训练数据量,即 $\rho_k = \rho_n = \rho, \forall k \in \mathcal{K}, \forall n \in \mathcal{N}$.由于问题(15)中的优化变量 $\{a_k, \forall k; b_n, d_n, \forall n; c_1, c_2\}$ 高度耦合,难以直接分析,因此本小节通过约束 $c_1=0$ 来分析问题(15)的次优解^[18].本节通过分析问题(15)最小值的上界 MSE_{d2d} ,并证明该上界不大于传统的空中计算模型聚合方案的 $\text{MSE}_{\text{no-d2d}} = \frac{\sigma^2}{P_0} \max_{m \in \mathcal{M}} \frac{|\rho_m|^2}{|h_m|^2}$.则优化问题(15)简化为

$$\min_{a_k, \forall k; b, d, c_2} \sum_{k=1}^K |c_2 a_k f b g_k - \rho|^2 + |c_2 f b d - \rho|^2 + (|c_2|^2 + |c_2|^2 |f|^2 |b|^2) \sigma^2 \quad (24a)$$

$$\text{s.t.} \quad |a_k|^2 \leq P_0, \forall k \in \mathcal{K}, \quad (24b)$$

$$|b|^2 \left(\sum_{k=1}^K |g_k|^2 |a_k|^2 + |d|^2 + \sigma^2 \right) \leq P_0. \quad (24c)$$

由于优化变量的耦合,该优化问题仍然是非凸的.考虑如下不等式:

$$\sum_{k=1}^K |c_2 a_k f b g_k - \rho|^2 + |c_2 f b d - \rho|^2 + (|c_2|^2 + |c_2|^2 |f|^2 |b|^2) \sigma^2 \geq (|c_2|^2 + |c_2|^2 |f|^2 |b|^2) \sigma^2. \quad (25)$$

等号成立的充分条件为 $a_k = \frac{\rho}{c_2 f b g_k} = \frac{d}{g_k}, \forall k \in \mathcal{K}, b = \frac{\rho}{c_2 f d}$,令 $\theta = |c_2|^2, \zeta = \frac{1}{|d|^2}$,则优化问题转换为

$$\min_{a_k, \forall k; b, d, c_2} (\theta + \zeta \rho^2) \sigma^2 \quad (26a)$$

$$\text{s.t.} \quad \zeta \geq \frac{1}{|g_k|^2 P_0}, \forall k \in \mathcal{K}, \theta \geq \frac{K \rho^2 + \rho^2 + \rho^2 \sigma^2 \zeta}{P_0 |f|^2}. \quad (26b)$$

该问题为关于变量 θ, ζ 的凸优化问题,其最优解为

$$\zeta^* = \frac{1}{P_0} \max_{\forall k \in \mathcal{K}} \frac{1}{|g_k|^2}, \theta^* = \frac{K \rho^2 + \rho^2 + \rho^2 \sigma^2 \zeta^*}{P_0 |f|^2}. \quad (27)$$

因此,若满足 $\text{MSE}_{\text{d2d}} \leq \text{MSE}_{\text{no-d2d}}$,则要求

$$\begin{aligned} \min_{a_k, \forall k; b, d, c_2} (\theta + \zeta \rho^2) \sigma^2 &= \left(\frac{K \rho^2 + \rho^2 + \rho^2 \sigma^2 \zeta^*}{P_0 |f|^2} + \frac{\rho^2}{P_0} \max_{\forall k \in \mathcal{K}} \frac{1}{|g_k|^2} \right) \sigma^2 \\ &= \left(\frac{K \rho^2 + \rho^2 + \frac{\rho^2 \sigma^2}{P_0} \max_{\forall k \in \mathcal{K}} \frac{1}{|g_k|^2}}{P_0 |f|^2} + \frac{\rho^2}{P_0} \max_{\forall k \in \mathcal{K}} \frac{1}{|g_k|^2} \right) \sigma^2 \leq \frac{\sigma^2 \rho^2}{P_0} \max_{k \in \mathcal{K}} \frac{1}{|h_k|^2}. \end{aligned} \quad (28)$$

表 2 系统参数的默认配置
Table 2 Default value of system parameters

Parameter	Description	Default value
K (or N)	The number of edge (or assisting) devices	20 (or 20)
P_0	The maximum transmit power	1 W
σ^2	The noise power	-100 dBW
α	The path loss exponent	4
R_n (or R_e)	The disk radius of assisting (or edge) devices located in scenario 1	250 m (or 500 m)
R_d	The outer radius of devices distributed in scenario 2	500 m
β, ω	The shape and scaling parameters of Nakagami model	1, 1
ϵ	The termination threshold of the algorithm	10^{-4}
J_{\max}	The maximum iteration number	100

上式等价于 $\frac{\sigma^2}{|f|^2 P_0} (K + 1 + \frac{\sigma^2}{P_0 \min_{\forall k \in \mathcal{K}} |g_k|^2}) \leq \frac{\sigma^2}{P_0 \min_{\forall k \in \mathcal{K}} |h_k|^2} - \frac{\sigma^2}{P_0 \min_{\forall k \in \mathcal{K}} |g_k|^2}$. 设 $\text{SNR}_{N\text{-BS}}$ 为辅助设备到基站链路的信噪比, $\text{minSNR}_{K\text{-BS}}$ 和 $\text{minSNR}_{K\text{-N}}$ 分别为边缘设备到基站和辅助设备链路的最小信噪比, 则

$$\text{SNR}_{N\text{-BS}} \geq \frac{\text{minSNR}_{K\text{-BS}} \text{minSNR}_{K\text{-N}}}{\text{minSNR}_{K\text{-N}} - \text{minSNR}_{K\text{-BS}}} \left(K + 1 + \frac{1}{\text{minSNR}_{K\text{-N}}} \right). \quad (29)$$

综上所述, 在满足式 (29) 的条件时, 本文提出的 D2D 辅助的双阶段空中计算模型聚合方案的 MSE_{d2d} 不大于直接进行空中计算模型聚合的传统方案的 $\text{MSE}_{\text{no-d2d}}$, 该条件要求辅助设备到基站链路的接收信噪比足够好. 需要注意的是, 该条件是 $\text{MSE}_{\text{d2d}} \leq \text{MSE}_{\text{no-d2d}}$ 的充分不必要条件. 在本文后续的仿真中也通过大量的实验验证了上述结论. 虽然所提聚合方案能够显著降低聚合误差, 提高基于空中计算的联邦学习性能, 但相对于直接进行空中计算模型聚合而言, 需要两倍的时隙资源, 同时在模型聚合的两个阶段都需要严格的时间同步, 以保证各个设备的模型参数能够对齐. 另外要求提供辅助的设备愿意提供 D2D 转发服务.

4 数值结果

本文选择 MSE 作为主要性能指标, 描述了无线环境得到的实际模型与理想模型的差异. 通过最小化 MSE, 提高接收信号估计的精度, 确保实际模型能够准确反映真实模型. 同时, MSE 作为一种通用性能指标, 在多个领域都有广泛应用. 大量文献也选择 MSE 来衡量实际模型与理想模型的差异 [15, 18, 33~35]. 此外, 除了 MSE, 本文还以联邦学习的测试精度作为评估实际模型的性能指标, 通过与理想模型预测精度随训练轮次的变化曲线对比, 来衡量实际模型与理想模型的误差. 本节通过与 2.1 小节中传统的空中计算联邦学习 (即 No D2D 方案) 对比, 进行数值实验来评估所提方案的性能. 系统参数的默认取值如表 2 所示. 后文所展示的所有实验结果均为 100 次独立重复试验结果的平均值.

本文考虑两种不同的辅助和边缘设备分布场景. 具体来说, 考虑一个 $1000 \text{ m} \times 1000 \text{ m}$ 的二维平面, 基站位于仿真区域的中心, 假设为坐标原点. 在场景一中, 辅助设备和边缘设备分别均匀分布在半径为 R_n 和 R_e 的圆周上, 如图 2 所示. 场景二为所有设备随机分布在外径为 R_d , 内径为 1 m 的圆环上, 如图 3 所示, 然后基于各个设备信道质量选择 N 个具有较高信道增益的用户作为辅助设备, 即 $\mathcal{N} = \max_{[N]} \{|h_1|, \dots, |h_M|\}$, 其中 $\{h_1, \dots, h_M\}$ 为每个设备到基站的信道系数. 在实际应用中, 基

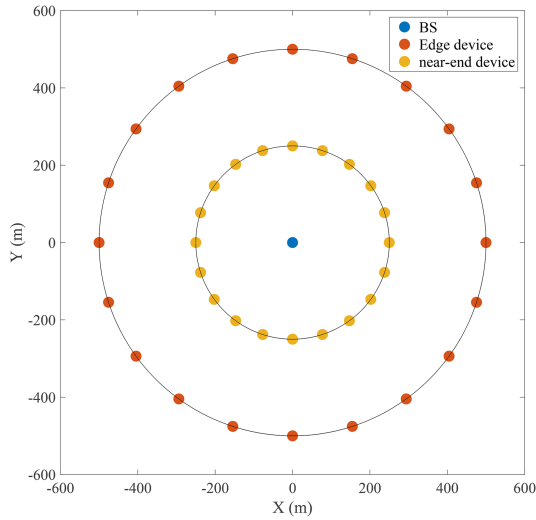


图 2 (网络版彩图) 设备均匀分布在内外两圆周上

Figure 2 (Color online) Devices are evenly distributed on the inner and outer circumferences

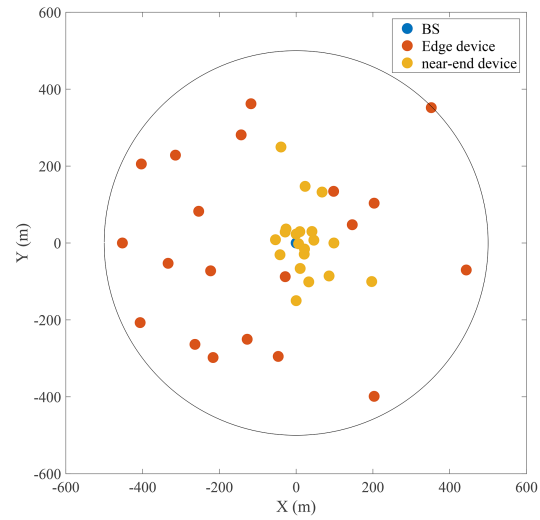


图 3 (网络版彩图) 设备在圆内随机分布

Figure 3 (Color online) Devices are randomly distributed within the disk

站可以根据所有设备的信道情况, 设置相应的阈值进行辅助设备和边缘设备的分类, 或者使用机器学习等算法进行二分类. 为了评估该方案性能, 采用 MNIST^[36] 数据集构建了一个基于空中计算的联邦学习任务, 该数据集包含从数字“0”到“9”的 10 个类别的 60000 个训练和 10000 个测试手写样本. 假设所有设备均分总数据样本, 每个设备都使用一个 3 层全连接的神经网络来训练模型. 网络的输入层为 784 像素值, 隐藏层为 Tanh 激活的 100 个神经元的全连接层, 输出层为 softmax 确定的 10 个类别的概率. 为了评估算法的稳定性, 本文对所提算法在场景一中进行了 100 次实验, 其中每次实验使用随机生成的不同的信道信息 $h_m, h_k, g_{k,n}, f_n$. 虽然不同的初始条件导致每次实验的迭代轮次和 MSE 优化结果不尽相同. 但这 100 次实验的 MSE 均成功优化至 -21.13 dB 及以下, 且均值为 -22.06 dB, 方差为 5.01×10^{-7} , 体现了算法在各种初始条件下的稳定性.

图 4 展示了场景一中所提出的交替优化算法在不同的设备数量下的收敛特性. 在 4 种设备数量设置情况下, 用户数量越多使得优化变量越多, 收敛速度则会减慢, MSE 分别在第 12 轮, 16 轮, 16 轮和 21 轮迭代后满足收敛条件. 在收敛后, 该算法在 MSE 的数值上明显降低, 即 MSE 分别降低为 $-20.48, -22.36, -23.91, -25.19$ dB.

图 5 和 6 分别研究了不同噪声功率下设备最大发射功率对于 MSE 的影响. 实验结果表明所提方案的 MSE 明显小于传统的空中计算联邦学习方案. 当 $P_0 = 1$ W 且 $\sigma^2 = -100$ dBW 时, 所提方案的 MSE 比传统方案分别小 11.39 和 10.78 dB. 同时, 较大噪声功率和较小发射功率会导致较小的接收信噪和较大的 MSE. 随着噪声功率的增大, D2D 辅助的空中计算方案的 MSE 增幅明显小于非 D2D 方案, 以图 5 为例, $P_0 = 1$ W, σ^2 从 -100 dBW 增加到 -80 dBW 时, D2D 辅助方案 and 传统方案的 MSE 分别增加了 5.36 和 20.01 dB; 随着 P_0 的减小, D2D 辅助方案的 MSE 增幅明显小于非 D2D 方案, 例如 $\sigma^2 = -100$ dBW, P_0 从 1 W 减小到 0.1 W 时, D2D 辅助方案 and 传统方案的 MSE 分别增加了 3.57 和 10.01 dB. 这是因为 D2D 方案中的通信质量差的设备可以利用多条传输路径的分集增益, 有效改善设备的通信条件.

图 7 研究了场景一中 4 种多径衰落级别下设备最大发射功率 P_0 对于 MSE 的影响. 实验结果表明,

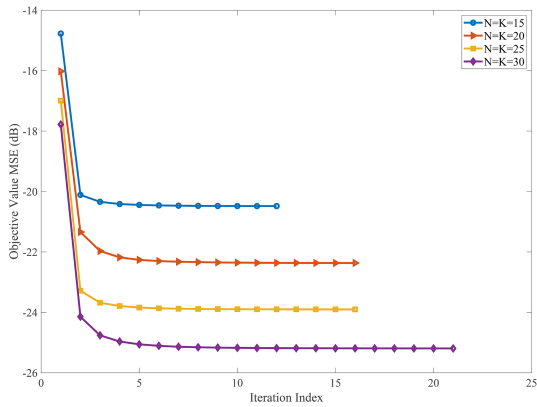


图 4 (网络版彩图) 算法的收敛性

Figure 4 (Color online) Convergence of the algorithm

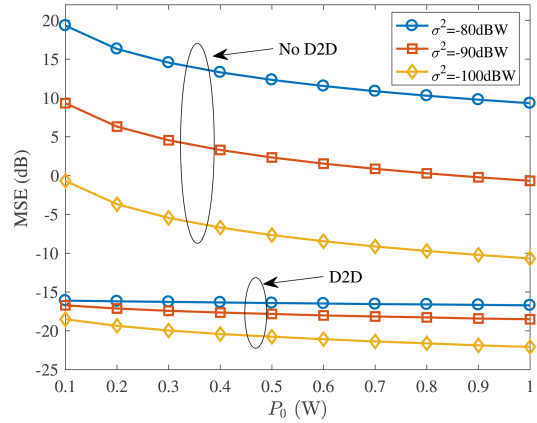


图 5 (网络版彩图) 场景一中 P_0 对 MSE 的影响

Figure 5 (Color online) Effect of P_0 on MSE in scenario 1

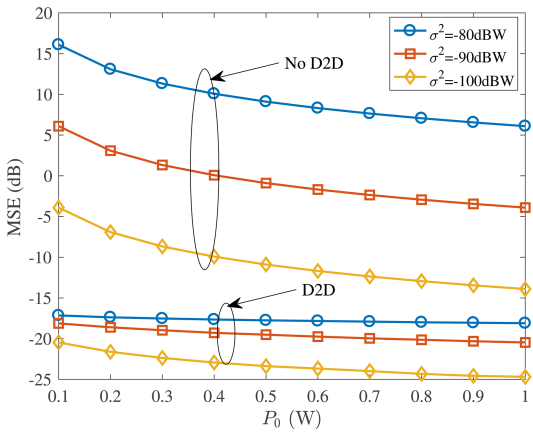


图 6 (网络版彩图) 场景二中 P_0 对 MSE 的影响

Figure 6 (Color online) Effect of P_0 on MSE in scenario 2

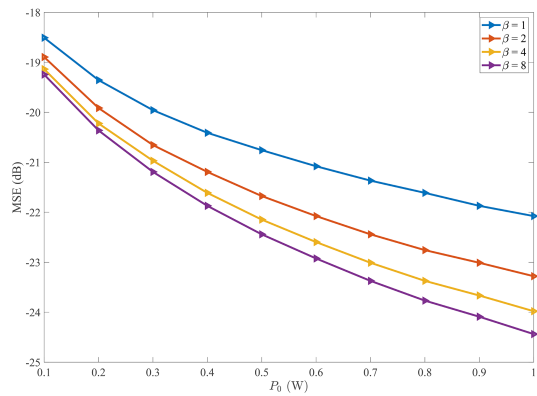


图 7 (网络版彩图) 多径衰落级别对 MSE 影响

Figure 7 (Color online) Impact of multipath fading level on MSE

不同多径衰落情况下所提方案都能保证较低的 MSE, 如本实验中 MSE 维持在 -18.5 dB 以下. 通过比较不同的性能曲线可以看出, β 越大, MSE 越小; 设备最大发射功率越大, 接收信噪比越大, MSE 越小. 当 $\beta = 1$ 时, Nakagami 模型退化为瑞利 (Rayleigh) 衰落, 没有直达径, 只有多径散射情况; 当 $\beta > 1$ 时, Nakagami 模型表征了直达径和多径共存的衰落情况, 同时 β 越大多径数量越少, 对应信道条件越好, 所提方案性能越好.

图 8 研究了场景一中不同 R_e 和 R_n 对于 MSE 的影响. 结果表明 R_e 越大 MSE 越大, 且边缘设备与基站的信道条件都会随着 R_e 的增大而变差, 从而导致 MSE 增大. 此外, D2D 方案的 MSE 随着 R_n 变大而变大, 原因在于 R_n 变大, 辅助设备信道条件变差, 从而导致辅助设备信号和 D2D 信号畸变程度增加. 图 9 研究了场景一中不同总设备数量和辅助设备数量对于 MSE 的影响. 实验结果表明当 N/M 越大, MSE 越小. 一方面, 辅助设备越多, D2D 方案中边缘设备通过辅助设备的转发次数

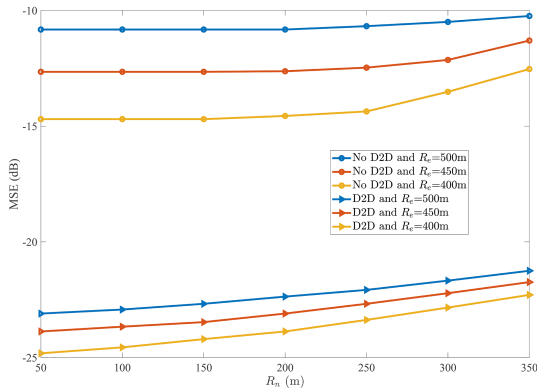


图 8 (网络版彩图) 设备位置对 MSE 的影响
 Figure 8 (Color online) Effect of device location on MSE

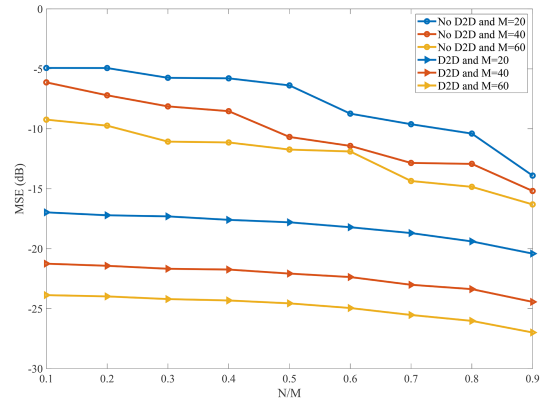


图 9 (网络版彩图) 设备数量对 MSE 的影响
 Figure 9 (Color online) Effect of the number of devices on MSE

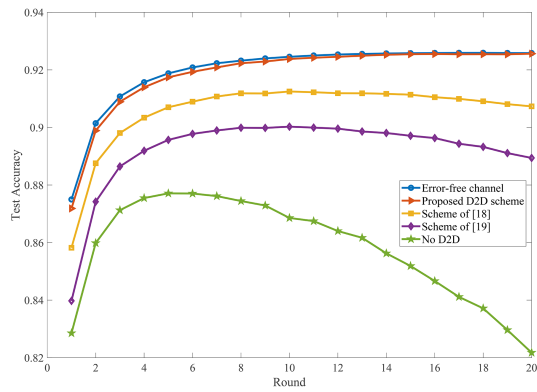


图 10 (网络版彩图) 场景一中 D2D 辅助的双阶段空中计算联邦学习性能
 Figure 10 (Color online) Performance of D2D-assisted two-phase over-the-air computation based federated learning in scenario 1

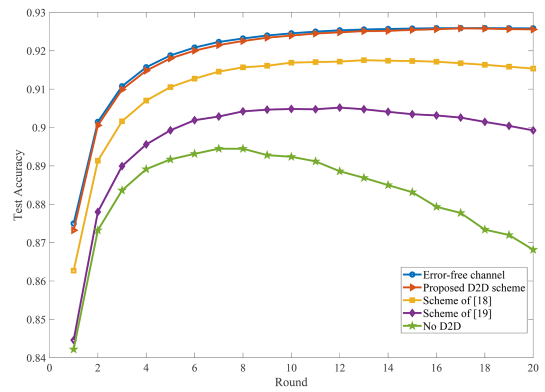


图 11 (网络版彩图) 场景二中 D2D 辅助的双阶段空中计算联邦学习性能
 Figure 11 (Color online) Performance of D2D-assisted two-phase over-the-air computation based federated learning in scenario 2

越多, 分集增益越大. 另一方面, 当 M 不变时, N 越大, 信道质量差的边缘设备数量减少, 误差有所减小.

图 10 和 11 分别展示了两种传统的空中计算联邦学习方案、本文方案、基于中继的方案 [18] 和基于调度的方案 [19] 的模型预测精度在训练过程中的变化. 具体如下所述.

(1) 基于中继的方案: 部署多个半双工中继, 通过中继辅助空中计算方式进行模型聚合. 在阶段一, 所有设备采用空中计算方式向中继和基站发送本地模型, 在阶段二, 所有设备和中继通过空中计算向基站发送本地模型. 与本文所提的 D2D 辅助方案的主要区别在于该方案需要引入中继设备. 此外, 该方案的中继设备仅用来辅助其他设备转发模型数据, 本身不需要向基站上传模型数据, 而在本文所提方案中, 辅助设备还需兼顾自身模型数据的上传. 本文仿真部分为中继方案部署了 4 个均匀分布在半径为 250 m 圆周上的中继设备.

(2) 基于调度的方案: 基站通过估计每个设备的信道增益, 对所有设备的信道增益进行排序, 调度

一定比例的信道条件较好的设备通过空中计算进行模型上传. 该方案直接放弃信道质量差的设备. 若调度设备过多, 则模型聚合误差增大. 若调度设备过少, 则放弃过多数据量导致聚合的联邦学习模型存在偏离. 这都将导致联邦学习模型的预测精度不佳, 因此在综合考虑后选择了 75% 这个性能相对较好的比例. 然而, 本文所提方案是 D2D 辅助所有设备都参与训练, 使得聚合误差显著降低.

传统空中计算方案会产生较大模型聚合误差. 考虑到网络的复杂性和安装维护成本, 中继数量较少, 导致基于中继方案的联邦学习所得模型预测精度有所下降. 训练中丢弃设备往往会减少参与模型训练的数据数量, 导致基于调度的方案所得的模型预测精度有所下降. 由于 D2D 的作用, D2D 辅助的双阶段空中计算方案具有更小的聚合误差, 取得了显著的精度提高, 并且最终模型的预测精度很接近无差错理想模型聚合的联邦学习. 随着训练的进行, 由于对比方案的误差远大于本文所提方案的误差, 而在联邦学习的训练过程中, 后一轮的训练都是在前一轮的聚合模型的基础上进行的, 这将导致模型误差的累积, 较大的累积聚合误差导致对比方案的全局模型愈发不准确, 从而降低了测试精度^[18].

5 结论

为了减小无线通信中空中计算联邦学习的模型聚合误差, 本文提出一种 D2D 辅助的双阶段空中计算模型聚合方案, 其中信道质量差的设备首先同时向多个信道质量好的用户和基站上传模型, 随后信道质量好的设备聚合本设备的模型和接收的模型再发送给基站, 基站聚合两个阶段接收的模型. 为了降低模型聚合误差, 分析了影响 MSE 的关键因素, 设计了联合优化问题, 并提出了一种交替优化算法用于优化设备发射系数、加权因子和基站的降噪接收系数以有效降低模型聚合误差. 在此基础上, 设计了两种不同的设备分布场景以及构建了一个基于空中计算的联邦学习任务进行实验. 结果表明, 该算法能够显著减小模型聚合误差, 并且能够很快得到收敛. 同时, D2D 辅助的双阶段空中计算联邦学习在不同的实验参数设置下的 MSE 均显著优于传统的空中计算聚合方案. 所得模型性能优于目前基于调度的方案和基于中继的方案, 并且最终模型的预测精度趋近于无差错的理想模型. 此外, 通过理论分析, 本文确定了使所提方案的 MSE 不大于传统方案的充分条件, 从而在理论上证明了所提方案的性能优势. 本文工作为在蜂窝网络中部署基于空中计算的联邦学习提供了有益支撑.

参考文献

- 1 Stoica I, Song D, Popa R A, et al. A Berkeley view of systems challenges for AI. 2017. ArXiv:1712.05855
- 2 LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature*, 2015, 521: 436–444
- 3 Zhu G, Liu D, Du Y, et al. Toward an intelligent edge: wireless communication meets machine learning. *IEEE Commun Mag*, 2020, 58: 19–25
- 4 Park J, Samarakoon S, Bennis M, et al. Wireless network intelligence at the edge. *Proc IEEE*, 2019, 107: 2204–2239
- 5 Dinh T Q, La Q D, Quek T Q S, et al. Learning for computation offloading in mobile edge computing. *IEEE Trans Commun*, 2018, 66: 6353–6367
- 6 Dinh T Q, Tang J, La Q D, et al. Offloading in mobile edge computing: task allocation and computational frequency scaling. *IEEE Trans Commun*, 2017, 65: 3571–3584
- 7 Han S, Mao H, Dally W J. Deep compression: compressing deep neural networks with pruning, trained quantization and Huffman coding. 2015. ArXiv:1510.00149
- 8 Cheng Y, Wang D, Zhou P, et al. Model compression and acceleration for deep neural networks: the principles, progress, and challenges. *IEEE Signal Process Mag*, 2018, 35: 126–136
- 9 Lin Y, Han S, Mao H, et al. Deep gradient compression: reducing the communication bandwidth for distributed training. 2017. ArXiv:1712.01887

- 10 Wang S, Tuor T, Salonidis T, et al. When edge meets learning: adaptive control for resource-constrained distributed machine learning. In: Proceedings of IEEE Conference on Computer Communications, 2018. 63–71
- 11 Karakus C, Sun Y, Diggavi S, et al. Straggler mitigation in distributed optimization through data encoding. In: Proceedings of Advances in Neural Information Processing Systems, 2017. 30
- 12 Yang Q, Liu Y, Chen T, et al. Federated machine learning: concept and applications. *ACM Trans Intell Syst Technol*, 2019, 10: 1–19
- 13 Smith V, Chiang C K, Sanjabi M, et al. Federated multi-task learning. In: Proceedings of Advances in Neural Information Processing Systems, 2017. 30
- 14 Tran N H, Bao W, Zomaya A, et al. Federated learning over wireless networks: optimization model design and analysis. In: Proceedings of IEEE Conference on Computer Communications, 2019. 1387–1395
- 15 Yang K, Jiang T, Shi Y, et al. Federated learning via over-the-air computation. *IEEE Trans Wireless Commun*, 2020, 19: 2022–2035
- 16 Nazer B, Gastpar M. Computation over multiple-access channels. *IEEE Trans Inform Theor*, 2007, 53: 3498–3516
- 17 Goldenbaum M, Boche H, Stanczak S. Harnessing interference for analog function computation in wireless sensor networks. *IEEE Trans Signal Process*, 2013, 61: 4893–4906
- 18 Lin Z, Liu H, Zhang Y J A. Relay-assisted cooperative federated learning. *IEEE Trans Wireless Commun*, 2022, 21: 7148–7164
- 19 Ma X, Sun H, Wang Q, et al. User scheduling for federated learning through over-the-air computation. In: Proceedings of IEEE 94th Vehicular Technology Conference, 2021. 1–5
- 20 Zhu G, Wang Y, Huang K. Broadband analog aggregation for low-latency federated edge learning. *IEEE Trans Wireless Commun*, 2019, 19: 491–506
- 21 Zhang F, Wan J J, Wang K L, et al. Joint receiver design and user scheduling for over-the-air aggregation in federated learning. In: Proceedings of the 5th International Conference on Communications, Information System and Computer Engineering (CISCE), 2023. 20–25
- 22 Du J, Jiang B Q, Jiang C X, et al. Gradient and channel aware dynamic scheduling for over-the-air computation in federated edge learning systems. *IEEE J Sel Areas Commun*, 2023, 41: 1035–1050
- 23 Liu H, Yuan X, Zhang Y J A. Reconfigurable intelligent surface enabled federated learning: a unified communication-learning design approach. *IEEE Trans Wireless Commun*, 2021, 20: 7595–7609
- 24 Qu Z, Guo S, Wang H, et al. Partial synchronization to accelerate federated learning over relay-assisted edge networks. *IEEE Trans Mobile Comput*, 2021, 21: 4502–4516
- 25 Aygün O, Kazemi M, Gündüz D, et al. Hierarchical over-the-air federated edge learning. In: Proceedings of IEEE International Conference on Communications, 2022. 3376–3381
- 26 Zhang F, Chen J, Wang K, et al. Device scheduling for relay-assisted over-the-air aggregation in federated learning. *IEEE Trans Veh Technol*, 2024, 73: 7412–7417
- 27 Feng S, Niyato D, Wang P, et al. Joint service pricing and cooperative relay communication for federated learning. In: Proceedings of International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData), 2019. 815–820
- 28 Lien S Y, Chien C C, Tseng F M, et al. 3GPP device-to-device communications for beyond 4G cellular networks. *IEEE Commun Mag*, 2016, 54: 29–35
- 29 McMahan B, Moore E, Ramage D, et al. Communication-efficient learning of deep networks from decentralized data. In: Proceedings of the 20th International Conference on Artificial Intelligence and Statistics, 2017. 1273–1282
- 30 Zhang R, Deng N, Wei H. Inter-cell interference coordination for aircomp federated learning in cellular networks. In: Proceedings of IEEE/CIC International Conference on Communications in China (ICCC), 2023. 1–5
- 31 Grant M, Boyd S. CVX: Matlab software for disciplined convex programming, version 2.1. 2014
- 32 Hua M, Wang Y, Wu Q, et al. Energy-efficient cooperative secure transmission in multi-UAV-enabled wireless networks. *IEEE Trans Veh Technol*, 2019, 68: 7761–7775
- 33 Wang F, Xu J. Optimized amplify-and-forward relaying for hierarchical over-the-air computation. In: Proceedings of IEEE Globecom Workshops (GC Wkshps), 2020. 1–6
- 34 Chen L, Qin X, Wei G. A uniform-forcing transceiver design for over-the-air function computation. *IEEE Wireless*

Commun Lett, 2018, 7: 942–945

35 Cao X, Zhu G, Xu J, et al. Cooperative interference management for over-the-air computation networks. IEEE Trans Wireless Commun, 2020, 20: 2634–2651

36 LeCun Y. The MNIST database of handwritten digits. 1998. <http://yann.lecun.com/exdb/mnist/>

D2D-assisted two-stage model aggregation scheme based on over-the-air computation

Ranqiang ZHANG¹, Na DENG^{1*}, Haichao WEI², Chengwen XING³ & Nan ZHAO¹

1. *School of Information and Communication Engineering, Dalian University of Technology, Dalian 116024, China;*

2. *School of Information Science and Technology, Dalian Maritime University, Dalian 116026, China;*

3. *School of Information and Electronics, Beijing Institute of Technology, Beijing 100081, China*

* Corresponding author. E-mail: dengna@dlut.edu.cn

Abstract In wireless federated learning, the channel quality heterogeneity of the devices at different locations results in model aggregation errors dominated by the device with the worst channel quality. To address this, a device-to-device (D2D) assisted over-the-air computation (OAC) federated learning scheme is proposed, where the devices with good channel quality are selected to assist edge devices in updating their local models to the server. An optimization problem is formulated to minimize the mean squared error and an alternating optimization algorithm is proposed to optimize the operational parameters of all devices and the base station with low complexity. To assess the performance of the proposed scheme, this paper validates its advantages over traditional OAC-based model aggregation schemes through theoretical analysis. Meanwhile, two different device distribution scenarios are designed to construct federated learning experiments based on neural networks and real datasets. The results show that the proposed algorithm converges quickly. Compared with traditional OAC schemes and existing scheduling or relay-based OAC schemes, the proposed scheme significantly reduces model aggregation errors and improves the prediction accuracy of federated learning.

Keywords D2D, federated learning, over-the-air computation, alternating optimization algorithm