

面向边缘智能的 通信计算一体化研究



Integrated Communication and Computation for Edge Intelligence

江炳青/JIANG Bingqing¹, 杜军/DU Jun¹,
王劲涛/WANG Jintao¹, 牟林/MU Lin^{2,3}

(1. 清华大学电子工程系, 中国 北京 100084;
2. 中兴通讯股份有限公司, 中国 深圳 518057;
3. 移动网络和移动多媒体技术国家重点实验室, 中国 深圳 518055)
(1. Department of Electronic Engineering, Tsinghua University, Beijing 100084, China;
2. ZTE Corporation, Shenzhen 518057, China;
3. State Key Laboratory of Mobile Network and Mobile Multimedia Technology, Shenzhen 518055, China)

DOI: 10.12142/ZTETJ.2024S1003

网络出版地址: <http://kns.cnki.net/kcms/detail/34.1228.TN.20240724.1044.006.html>

网络出版日期: 2024-07-25

收稿日期: 2023-12-06

摘要: 为了进一步提高无线数据聚合效率, 空中计算技术通过利用无线信道波形叠加特性允许模型更新信息在空中“一次性”完成聚合, 实现通信网与算力网的“网媒融合”。然而在这过程中, 信道衰落和噪声可能会带来聚合失真。此外, 更新数据的质量以及边缘设备的传输能耗也可能影响模型聚合以及收敛效率。为此提出了基于空中计算的联邦学习系统, 并针对其存在的信道干扰、高效数据传输和数据失真问题建立动态设备调度机制, 在满足接收端信噪比条件下选择适当数量质量较高的设备参与模型训练。该机制利用梯度重要性、信道条件和传输能耗衡量设备质量并保留累积未被选择设备的梯度以加速收敛。基于李雅普诺夫优化理论进行问题建模和求解, 仿真结果表明该机制具有较高训练精度和较快收敛速度, 同时针对不同噪声功率具有一定鲁棒性。

关键词: 空中计算; 联邦学习; 设备调度; 设备质量; 鲁棒性

Abstract: Over-the-air computation (AirComp) technology leverages the waveform superposition characteristics of wireless channels to further enhance the efficiency of wireless data aggregation, enabling model update information to be aggregated “in one shot”. This achieves a convergence of communication networks and computational power networks, exemplifying the concept of “network and computation fusion”. However, channel fading and noise may introduce aggregation distortion during this process. Additionally, the quality of update information and the transmission energy consumption of edge devices can impact model aggregation and convergence efficiency. Therefore, we establish an AirComp enabled federated learning system and propose a dynamic device scheduling mechanism to address issues related to channel interference, efficient data transmission, and data distortion. Specifically, an appropriate number of higher-quality devices are selected to participate in model training while satisfying receiving signal-to-noise ratio conditions. It utilizes gradient importance, channel conditions, and transmission energy consumption to assess device quality and retains and accumulates gradients from unselected devices to accelerate convergence. The problem is modeled and solved based on the Lyapunov optimization theory. Simulation results demonstrate that this mechanism achieves higher training accuracy, faster convergence speed, and a certain level of robustness against varying noise power levels.

Keywords: over-the-air computation; federated learning; device scheduling; device quality; robustness

引用格式: 江炳青, 杜军, 王劲涛, 等. 面向边缘智能的通信计算一体化研究 [J]. 中兴通讯技术, 2024, 30(S1): 16-23. DOI: 10.12142/ZTETJ.2024S1003

Citation: JIANG B Q, DU J, WANG J T, et al. Integrated communication and computation for edge intelligence [J]. ZTE technology journal, 2024, 30(S1): 16-23. DOI: 10.12142/ZTETJ.2024S1003

随着移动互联网由“万物互联”发展为“万物智联”, 智能化已经成为新的主要需求和发展趋势。这一巨大变革推动了新的智能化应用, 同时对于超可靠低延迟通信、

能量效率、智能与安全也提出了更加严格的要求^[1]。在智能化引领发展的阶段中, 人工智能和机器学习技术被广泛应用于移动互联网领域及其发展中。得益于特有的普适性、自主性以及迭代优化等特性, 人工智能和机器学习技术通过数据处理得到更加严谨、稳固的模型和推演结果以提供先进的计

基金项目: 国家自然科学基金项目 (U23A20281、61971257)

算能力，支持越来越多的智能应用^[2]。

传统机器学习范式通常采用依赖于中央服务器的集中式学习方式，然而智能应用的爆炸式增长以及网络边缘产生的海量数据使得计算密集型任务在数据传输与模型计算方面产生较大时延与带宽负担。此外，由于集中式学习需要聚集所有的原始数据，尽管可以获得性能更优的训练效果，但这一过程中可能会遭到恶意攻击或者窃听，产生了隐私安全问题。随着隐私安全意识的逐渐加强，这一缺点对于部分应用会造成致命的影响，例如金融、医疗等行业中隐私高度敏感的应用。

受益于边缘设备日渐增长的存储与计算能力，分布式学习架构的提出解决了隐私安全和带宽资源问题。中央服务器将数据限制在各个边缘设备处进行模型训练，并将最终的模型汇聚到服务器进行汇总和存储^[3]。分布式学习架构实现了将任务分布到不同边缘设备进行训练，避免了原始数据的上传，有效保护了数据安全。但是由于不同边缘设备仅拥有部分数据且彼此之间缺乏数据或模型方面的“沟通”，系统内数据无法交汇融合从而导致训练模型缺乏全局性和泛化能力，这一问题也被称为“数据孤岛”问题^[4]。

针对以上问题，由谷歌提出的联邦学习框架使得智能边缘协同模型训练成为现实^[5]。在不共享边缘设备本地数据集的情况下，通过边缘服务器进行协作训练^[6-7]。典型的联邦学习是通过通信与计算相分离的方式实现的。具体来说，每个边缘设备都使用边缘服务器广播的当前全局模型进行本地模型训练，通过运行随机梯度下降算法计算本地模型参数；本地训练结束后，服务器聚合边缘设备上传的本地模型参数

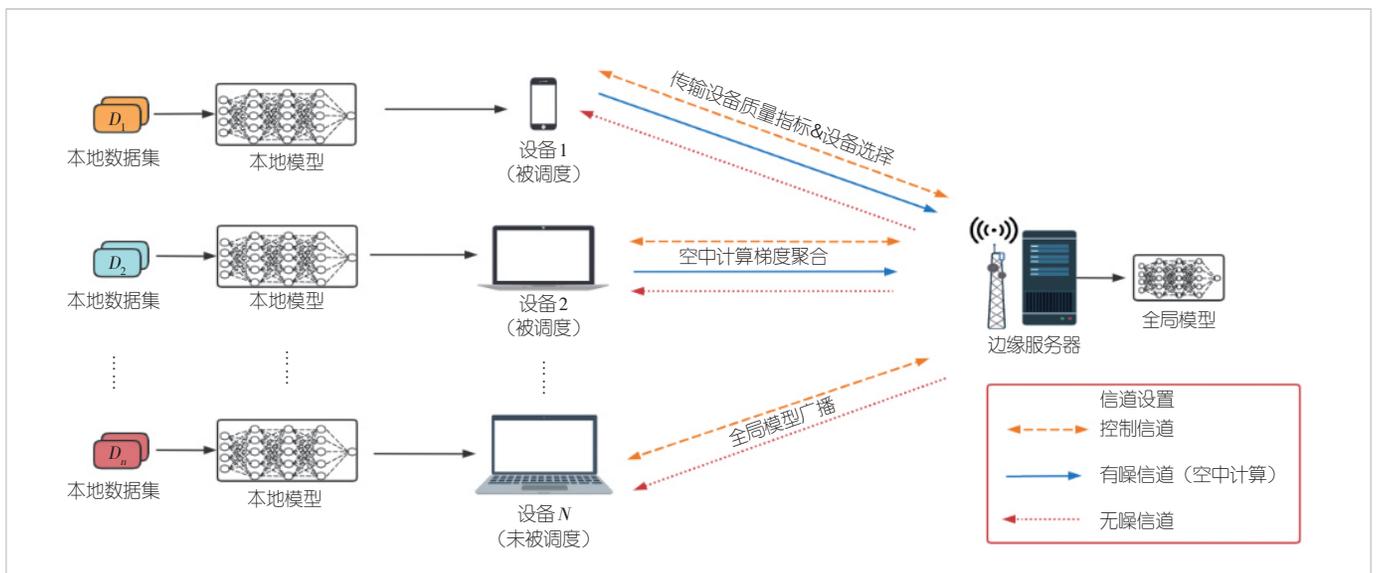
及其梯度，并利用聚合值的平均值进行全局模型更新。服务器将得到的全局模型再次广播给边缘设备以开始下一轮训练，这一过程是较为经典的联邦平均算法。

然而在通信与计算相分离的方式下，尽管本地模型的数据量相比于数据本身而言已经小得多，但向无线信道上上传本地模型参数或其梯度仍然需要消耗大量频谱资源，造成了“通信瓶颈”问题；另一方面，服务器仅需要本地模型参数的加权平均值而非其本身来训练全局模型，而传统的模型聚合利用编解码辅助准确接收模型参数，这会导致不必要的通信和计算资源浪费。因此，将通信网与算力网进行“网媒融合”，实现通信计算一体化是十分必要的。受此启发，空中计算的引入为“通信瓶颈”问题提供了有效的解决方案，且已被文献[8]证明是一种利用无线信道叠加特性计算和训练的有效通信方法。通过实现通信与计算一体化融合，空中计算可以很好地满足联邦学习系统的高通信效率和低延迟需求，辅助联邦学习在无需精准接收模型参数的情况下实现快速、高效的无线数据聚合。

本文将研究基于空中计算的联邦学习系统，其中在边缘服务器的协同下，边缘设备协作训练并高效通信以完成全局训练任务。本文从收敛性分析着手，综合考虑不同边缘设备间对于全局模型的贡献程度差异，设计了动态设备调度策略以追求高模型精度、能量效率以及通信效率。

1 系统模型

如图1所示，基于空中计算的联邦学习系统由边缘服务器和 N 个边缘设备组成，设备本地数据集为 D_n 并采用小批



▲图1 基于空中计算的联邦学习系统架构图

量($D_{n,t}^m \subseteq D_n$)随机梯度下降算法计算本地模型梯度。本文定义一次全局模型更新为一个通信轮次,索引表示为 $t \in \{1, \dots, T\}$ 。由于通信资源限制每轮只有部分设备上传信息, S_t 为被选设备集合, $|S_t| = \sum_{n=1}^N a_{n,t}$, 其中 $a_{n,t} \in \{0, 1\}$ 为设备 n 是否被选择的二元变量。

1.1 联邦学习模型

定义设备 n 的局部损失函数为 $f_n(\mathbf{w}) = \frac{1}{|D_n|} \sum_{(s_i, q_i) \in D_n} l(\mathbf{w}; s_i, q_i)$, 其中 $l(\mathbf{w}; s_i, q_i)$ 为数据样本 (s_i, q_i) 的损失函数, 则衡量全局模型对整体数据集平均拟合度的全局损失函数为

$$f(\mathbf{w}) = \frac{1}{|\bigcup_{n \in \mathbb{N}} D_n|} \sum_{n=1}^N |D_n| f_n(\mathbf{w}). \quad (1)$$

联邦学习旨在寻找最佳模型参数 $\mathbf{w}^* = \arg \min_{\mathbf{w}} E[f(\mathbf{w}_T)]$ 。应用伸缩和法则, 优化问题化为

$$P_1: \min \sum_{t=1}^T E[f(\mathbf{w}_t)] - E[f(\mathbf{w}_{t-1})], \quad (2a)$$

$$\text{s.t. } \frac{1}{T} \sum_{t=1}^T a_{n,t} E_{n,t} \leq \bar{E}_n, \quad (2b)$$

$$a_{n,t} \in \{0, 1\}, \quad (2c)$$

其中, \bar{E}_n 为设备 n 的平均传输能量约束。由于神经网络模型参数无法预测, $E[f(\mathbf{w}_t)] - E[f(\mathbf{w}_{t-1})]$ 难以写出闭式形式, 因此后文将通过收敛性分析得到其闭式上界进行近似简化。

1.2 空中计算传输聚合模型

假设信道增益遵循独立同分布准静态瑞利衰落, $h_{n,t} \sim CN(0, 1)$, 本文提出基于信道反转法的功率控制以抵抗信道衰落^[9]。设备 n 的发射功率为

$$P_{n,t} = \sigma_t \frac{h_{n,t}^H}{|h_{n,t}|^2}, \quad (3)$$

其中功率扩展因子 σ_t 决定服务器接收端信噪比。尽管 $h_{n,t}$ 会引入相关复数操作, 但发射功率的计算是不受影响的。令设备发射信号 $\mathbf{x}_{n,t} = \Phi(\mathbf{g}_{n,t})$, $\Phi(\cdot)$ 为预处理操作以确保 $\mathbf{x}_{n,t}$ 具有零均值和方差 $P_{n,t}$ 以便能量控制^[9], 则聚合信号为

$$\mathbf{y}_t = \sum_{n \in S_t} h_{n,t} \mathbf{x}_{n,t} + \boldsymbol{\varepsilon}_t, \quad (4)$$

其中 $\boldsymbol{\varepsilon}_t$ 为 0 均值方差 σ_0^2 的高斯噪声。服务器完成后处理后更新全局模型:

$$\mathbf{w}_t = \mathbf{w}_{t-1} - \eta_t \left(\frac{\sum_{n \in S_t} \mathbf{g}_{n,t}}{|S_t|} + \hat{\boldsymbol{\varepsilon}}_t \right), \quad (5)$$

其中 η_t 为学习速率, $\hat{\boldsymbol{\varepsilon}}_t = \Phi^{-1} \left(\frac{\boldsymbol{\varepsilon}_t}{\sigma_t |S_t|} \right)$ 。根据公式 (4), 接收

端信噪比 $\gamma_s = \frac{|h_{n,t}|^2 E[\|\mathbf{x}_{n,t}\|_2^2]}{\sigma_0^2}$, 令 γ_{thr} 为预设信噪比阈值, 要求服务器处满足 $\gamma_s \geq \gamma_{thr}$, 则 σ_t^2 为

$$\sigma_t^2 = \gamma_{thr} \sigma_0^2. \quad (6)$$

2 基于梯度与信道感知的动态设备调度机制

2.1 动态残差反馈机制

由于设备选择机制的存在, 未被选择设备的本地更新信息无法用于全局模型更新, 从而会导致“数据失真”问题。目前大多数研究中未被选择设备的本地更新信息直接被丢弃, 这种粗糙的操作可能会导致训练偏差, 甚至改变梯度的方向和大小并进一步影响模型训练的稳定性和准确性^[10]。因此, 有必要保留未被选择设备的训练信息, 并在数据和信道质量都满足条件而被调度时联合累积的梯度一同传输至服务器。为此, 本文提出了一种动态残差反馈方案, 该方案允许边缘设备传输过去未发送的累积局部梯度, 而未被选择的设备在本地保存本轮模型梯度并等待下一次传输。定义被选择的边缘设备 n 在第 t 轮的本地更新信息 $\tilde{\mathbf{g}}_{n,t}$ 为当前通信轮次的梯度向量 $\mathbf{g}_{n,t}$ 与累积残差 $\mathbf{r}_{n,t}$ 的组合, 即

$$\tilde{\mathbf{g}}_{n,t} = \mathbf{g}_{n,t} + \mathbf{r}_{n,t}, \quad (7a)$$

$$\mathbf{r}_{n,t} = \begin{cases} 0, n \in S_{t-1}, t \geq 2 \\ \xi \mathbf{g}_{n,t-1}, n \notin S_{t-1}, t \geq 2, \end{cases} \quad (7b)$$

其中, $0 \leq \xi \leq 1$ 表示累积残差对于当前训练的重要程度, $\mathbf{r}_{n,1} = 0$ 。

2.2 基于收敛性分析的问题转化

本文首先引入以下假设以利于分析收敛性^[11]:

假设 1 (l -smooth): 损失函数 $f_1(\mathbf{w}), \dots, f_N(\mathbf{w})$ 具有 L -光滑性, 即对于 $\forall \mathbf{x}, \mathbf{y} \in \mathcal{R}^{L_s}$ 以及 $\forall n \in N, \exists l < \infty$ 使得损失函数满足

$$f_n(\mathbf{y}) \geq f_n(\mathbf{x}) + \nabla f_n(\mathbf{x})^T (\mathbf{y} - \mathbf{x}) + \frac{l}{2} \|\mathbf{y} - \mathbf{x}\|^2. \quad (8)$$

假设 2 (μ -strongly): 损失函数 $f_1(\mathbf{w}), \dots, f_N(\mathbf{w})$ 具有 μ -鲁棒性, 即对于 $\forall \mathbf{x}, \mathbf{y} \in \mathcal{R}^{L_s}$ 以及 $\forall n \in N, \exists \mu < \infty$ 使得损

失函数满足

$$f_n(\mathbf{y}) \leq f_n(\mathbf{x}) + \nabla f_n(\mathbf{x})^T(\mathbf{y} - \mathbf{x}) + \frac{\mu}{2} \|\mathbf{y} - \mathbf{x}\|^2. \quad (9)$$

假设3 (无偏估计&方差有界): 本地随机梯度具有无偏估计性且方差有界, 即 $\forall n \in \mathbb{N}, \forall t \in T$, 对随机数据抽样取期望后满足

$$E[\mathbf{g}_{n,t}]|_{D_{n,t}^m \subseteq D_n} = \mathbf{g}_t, \quad (10a)$$

$$E\left[\|\mathbf{g}_{n,t} - \mathbf{g}_t\|_2^2\right]|_{(s_t, q_t) \in D_n} \leq G^2, \quad (10b)$$

其中, G^2 表示局部梯度的方差上界且可以认为是常数。基于以上假设, 本文首先在以下引理1中推导了单轮平均累积残差的期望用于进一步的收敛性分析。

引理1: 给定第 t 轮的全局模型梯度 \mathbf{g}_t 和本地随机梯度的方差上界 G^2 , 单轮平均累积残差的上界为

$$\left[\frac{1}{N} \sum_{n=1}^N r_{n,t} \right] \leq \frac{P_r(G^2 + \|\mathbf{g}_t\|_2^2)}{1 - P_r}, \quad (11)$$

其中 P_r 表示第 t 轮中设备未被选择的平均概率, $0 < P_r < 1, \forall t \in \{0, \dots, T-1\}$ 。相关证明见文献[12]附录。

随后基于假设和引理1, 本文通过推导 $E[f(\mathbf{w}_t)] - E[f(\mathbf{w}_{t-1})]$ 的上界进行单轮收敛性分析, 此处综合考虑了发射功率、信道噪声、平均累积残差上界以及随机梯度来进行分析, 分析结果如引理2。

引理2: 给定第 t 轮的调度设备集合 S_t 和小批量大小 $D_{n,t}^m$, 通信的收敛速度由公式(12)给出。

$$E[f(\mathbf{w}_t)] - E[f(\mathbf{w}_{t-1})] \leq \frac{l\eta_t^2}{2} \|\mathbf{g}_t\|_2^2 + \frac{l\eta_t^2 + \eta_t}{2} m^2 + \frac{l\eta_t^2 + \eta_t}{2} \frac{P_r(G^2 + \|\mathbf{g}_t\|_2^2)}{1 - P_r} + \frac{l\eta_t^2}{2} \left[\frac{\delta^2}{\gamma_{thr}|S_t|^2} + \frac{G^2}{|S_t||D_{n,t}^m|} \right], \quad (12)$$

其中, m 是全局梯度的估计平均值, δ^2 是全局梯度的估计方差值。证明见文献[12]附录。

观察引理2可知, 每轮所选择的设备数量, 即 $|S_t| = \sum_{n \in S_t} a_{n,t}$, 在抵抗信道噪声影响方面起到了至关重要的作用。因此, 高效的设备调度机制需要根据每一轮中的信道噪声水平来优化设备选择策略 $a_{n,t}$ 。与最大化平均每轮所选择的设备数量不同, 引理2提供了一个更加合理的优化方向, 从收敛速率着手, 最小化 $E[f(\mathbf{w}_t)] - E[f(\mathbf{w}_{t-1})]$ 的上界以抵抗信道扰动的影响。此外, 由于本文使用单轮收敛速率的上界进行近似得到闭式表达式, 使得操作可执行化。对引理

2做进一步观察, \mathbf{g}_t 是定义在整个数据集上的全局梯度, 这表明对于固定数据集, \mathbf{g}_t 也是固定的。同时引理2中公式(3) - (11)的前三项与调度策略无关, 可以视作常数对待。因此, 忽略常数项 $l\eta_t^2/2$, 令

$$U_t \triangleq \frac{\delta^2}{\gamma_{thr}|S_t|^2} + \frac{G^2}{|S_t||D_{n,t}^m|}. \quad (13)$$

进而问题 P_1 可以优化转化为 P_2 进行求解

$$\begin{aligned} P_2 \min \quad & \sum_{t=1}^T U_t, \\ \text{s.t.} \quad & \frac{1}{T} \sum_{t=1}^T a_{n,t} E_{n,t} \leq \bar{E}_n, \\ & a_{n,t} \in \{0, 1\}. \end{aligned} \quad (14)$$

2.3 设备质量指标设计

在联邦平均算法以及当前大多数研究中, 服务器随机选择边缘设备参与全局模型训练, 忽略了不同边缘设备之间的差异。然而, 上传更新变化不大的局部模型参数梯度对全局模型性能提升十分有限, 因此降低了传输资源的利用效率。此外, 传输所带来的能量消耗以及信道条件都会影响模型聚合的效率和准确性。综合考虑上述因素, 本小节旨在设计量化合适的设备质量指标以测量不同边缘设备对全局模型的潜在影响。

本地更新信息重要性由其 l_2 -范数 $\|\tilde{\mathbf{g}}_t\|_2^2$ 度量并定义为数据状态信息, 较大数据状态信息可提供较大模型变化^[13]; 本文用 $|h_{n,t}|$ 衡量信道条件并定义为信道状态信息, 信道状态信息越大信道条件越好, 避免拖后腿问题^[9]。之后归一化数据

与信道状态信息 $v_{DSI_{n,t}} = \frac{\|\tilde{\mathbf{g}}_t\|_2^2}{g_{\max}}$, $v_{CSI_{n,t}} = \frac{|h_{n,t}|}{h_{\max}}$ 以统一范围并定义设备重要性, 即:

$$V_{n,t} = \rho_1 v_{DSI_{n,t}} + \rho_2 v_{CSI_{n,t}}, \rho_1 + \rho_2 = 1. \quad (15)$$

由于90%能耗来自于传输过程, 因此选择机制需要考虑设备传输能耗 $E_{n,t} = \sigma_t^2 / |h_{n,t}|^2$ 以追求高能效^[14]。综合考虑能耗损失和设备重要性增益, 定义设备质量指标以衡量不同设备对全局模型的潜在影响:

$$I_{n,t} = \lambda_V V_{n,t} - \lambda_E E_{n,t}, \lambda_E + \lambda_V = 1. \quad (16)$$

设备能本地计算 $I_{n,t}$ 并传输到服务器, 服务器每轮选择具有较大 $I_{n,t}$ 的设备参与训练。

2.4 基于李雅普诺夫漂移优化算法问题建模和动态设备调度算法

基于上述分析, 所选设备数量和设备质量是调度机制中的两大关键要素, 因此本文旨在设计一种设备调度机制以决定所选设备数量减少噪声干扰, 并确保在有限资源下所被选择设备质量尽可能高以充分利用其训练性能。基于李雅普诺夫漂移惩罚算法的设备选择优化问题建立为

$$P_3 \min_{a_{n,t}} \alpha U_t - \sum_{n=1}^N a_{n,t} I_{n,t}, \quad (17)$$

约束条件公式(2b)和(2c), 其中, α 是平衡收敛上界可调整项 U_t 和设备质量 $I_{n,t}$ 之和的李雅普诺夫因子。此外, 参数 α 的选择还需要考虑平均能量约束 (2b) 是否满足。

算法 1 给出了基于李雅普诺夫优化理论的求解算法, 算法复杂度为 $O(N \log N)$ 。尽管传输 $I_{n,t}$ 带来额外通信开销, 但相比于传输梯度几乎可忽略不计。此外, $I_{n,t}$ 的定义十分灵活, 算法具有很强的适应性。

算法 1: 基于梯度和信道条件感知的动态设备调度机制

```

输入:  $\lambda_E, \lambda_V, \alpha, \gamma_{thr}$ 
输出: 设备调度决策  $a_{n,t}$ 
初始化  $w_0$ 
for  $t=1 \text{---} t=T$  do
    服务器广播全局模型参数  $w_{t-1}$ 
    for  $n=1 \text{---} n=N$  do
        本地模型训练计算  $g_{n,t}$ 
        计算  $I_{n,t}$  并通过控制信道传输到服务器
    end for
    服务器降序排序  $I_{n,t}$ 
    for  $k=1 \text{---} k=N$  do
        服务器计算相应惩罚项  $p_t(k) = \alpha U_t - \sum_{n=1}^k I_{n,t}$ 
    end for
     $k^* = \arg \min p_t(k)$ 
    for  $n=1 \text{---} n=N$  do
        if  $I_{n,t} > I_{k^*,t}$  then
             $a_{n,t} = 1$ 
        else
             $a_{n,t} = 0$ 
        end if
    end for
end for
for  $n=1 \text{---} n=N$  do

```

根据式(7)更新本地残差

if $a_{n,t} = 1$ then

根据式(3)设置传输功率并通过空中计算向服务器
传输 $\tilde{g}_{n,t}$

end if

end for

服务器根据式(5)更新全局模型

end for

3 仿真结果与分析

在仿真中, 本文研究基于空中计算的联邦学习系统, 该系统由一个边缘服务器和 $N=100$ 个边缘设备组成。本文假设每个边缘设备到边缘服务器的无线信道遵循独立同分布准静态瑞利衰落, 建模为 $h_{n,t} \sim CN(0,1)$, 且边缘服务器和边缘设备都可以观察到准确的信道增益。发射功率根据公式 (6) 进行计算, 其中信道噪声的方差 σ_0^2 分别在 $\{0.5, 1, 3\}$ 中取值以表示不同的信道条件, 通过调整不同 σ_0^2 的大小验证所提出的设备调度机制的鲁棒性。本文设置平均能量阈值为 $\bar{E}_n=1$ ($\forall n \in \mathbb{N}$)。在联邦学习训练设置中, 本文通过使用两个 5×5 卷积层的卷积神经网络、一个带有 512 个线性整流单元 (ReLU) 激活的全连接层以及一个 SoftMax 输出层来执行图像分类的学习任务^[5]。此外, 本文使用了 MNIST 数据集, 其中设置了 60 000 个被标记的训练数据样本和 10 000 个测试数据样本, 同时考虑了独立同分布以及非独立同分布数据分布。联邦学习的超参数设置为: 动量优化器为 0.5, 本地模型训练轮次为 2, 总共训练 $T=50$ 轮, 本地小批量的大小为 $|D_{n,t}^m|=10$, 学习速率 $\eta_t=0.01$ 。此外, 本文在预训练中将 G^2 设置为所有局部梯度方差的最大值^[11]。李雅普诺夫因子设置为 $\alpha=5\ 000$, 并令 $\rho_1=\rho_2=0.5$, $\lambda_E=\lambda_V=0.5$ 以保证公平性^[10]。

本文研究系统将与以下基线算法进行性能对比。1) 理想基线: 使用所有无噪梯度平均值更新全局模型以提供基准精度; 2) 基线 1^[9]: $|h_{n,t}| \geq |h_{thr}| = 1$ 被调度; 3) 基线 2^[13]: 首选 $|h_{n,t}|$ 最大的 $k_c = 50$ 个设备, 再选 $\|g_t\|_2^2$ 最大的 $k = 20$ 个设备; 4) 基线 3^[10]: 设备根据公式 (18) 决定是否被调度。

$$P_n^t = \begin{cases} \left(\frac{|h_{n,t}|}{|h_{n,t}|^2 + c} \right)^2, & \frac{\|\tilde{g}_t\|_2^2 |h_{n,t}|^2}{|h_{n,t}|^2 + c} \geq c P_{on} \\ 0, & \text{else} \end{cases}, \quad (18)$$

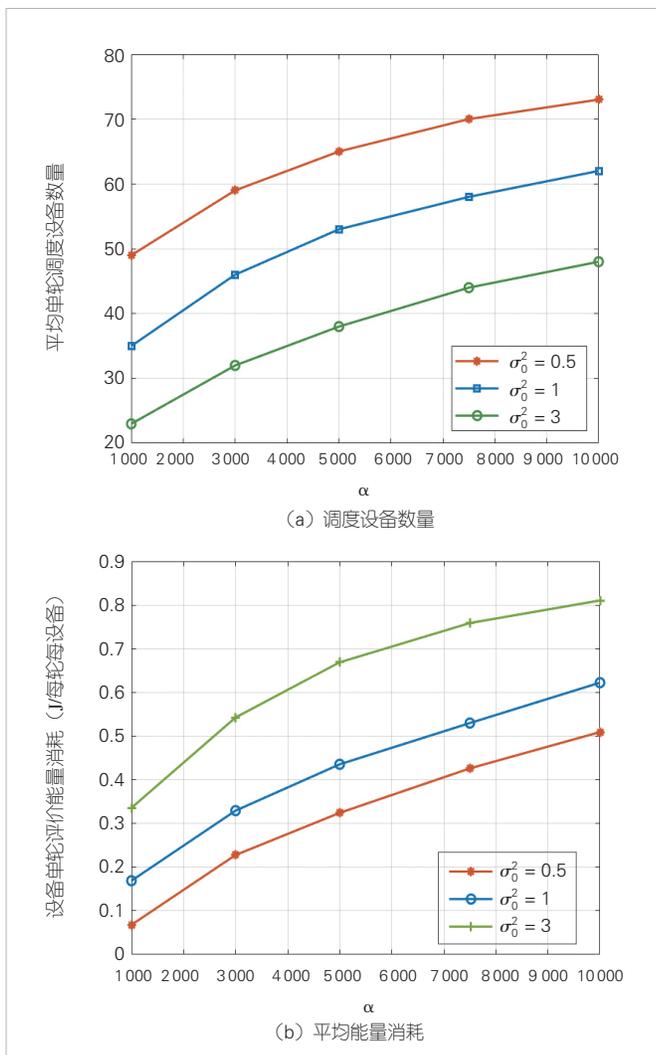
其中, 功率消耗代价 $c = 1$, 设备激活功率 $P_{on} = 4$ 。

3.1 李雅普诺夫因子 α 的影响

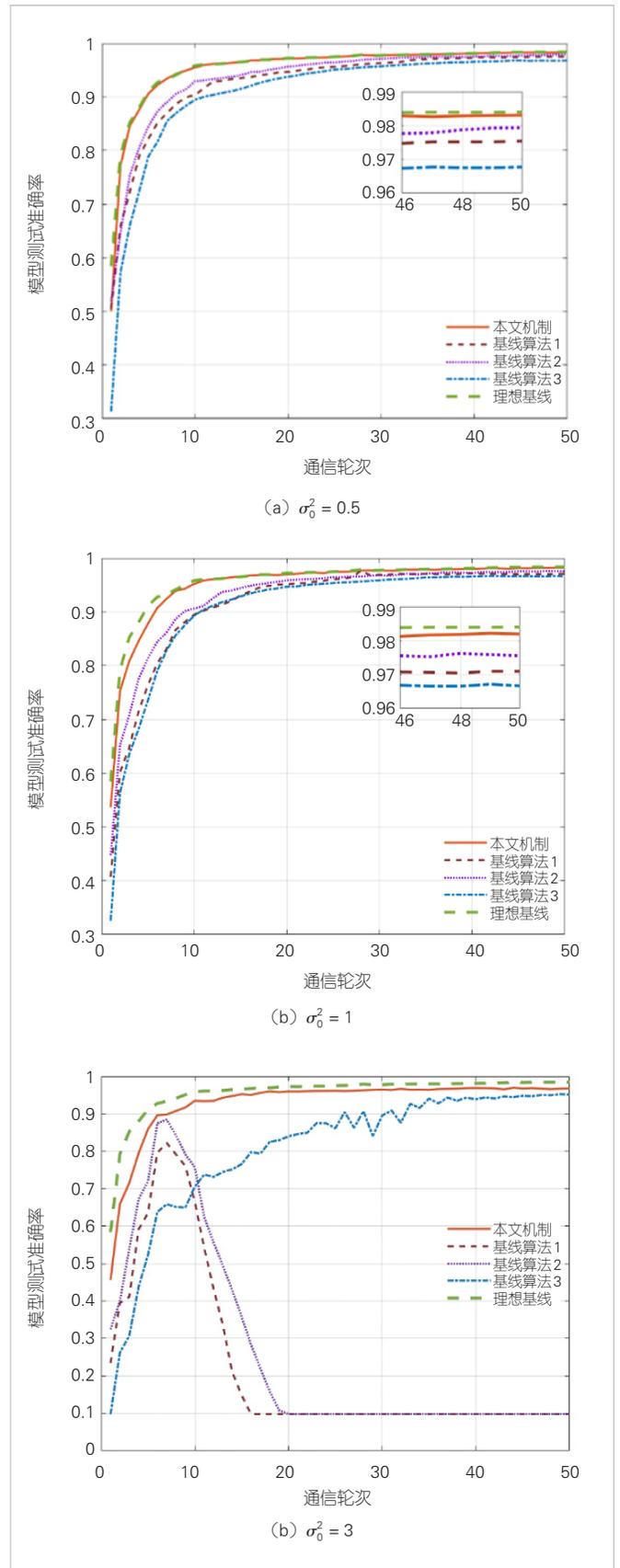
本文首先分析了李雅普诺夫因子 α 对于调度设备数量和单轮设备平均传输能耗的影响，仿真结果如图2所示。图2 (a) 表明，调度设备数量随着 α 的增加而增加，这是因为根据设备调度问题的优化目标、公式 (13) 以及 $I_{n,l}$ 的定义，边缘服务器倾向于选择更多设备以减少噪声的影响，但是相应代价是增加系统的传输能耗，如图2 (b) 所示。具体的，根据优化目标可知，增加 α 意味着服务器更倾向于通过增加调度设备数量以降低收敛上界可调整项 U_l ，而适当降低由 $I_{n,l}$ 定义的设备质量权重。另一方面，随着 α 的降低，在可选设备数量降低的条件下，服务器更倾向于选择那些具有“良好”设备质量（即 $I_{n,l}$ 值较大）的设备以抵抗信道噪声影响。

3.2 算法性能对比

图3为算法性能比较图，结果表明得益于动态残差反



▲图2 不同 α 和 σ_0^2 下的调度设备数量、平均能量消耗



▲图3 不同信道条件下的模型测试精度

馈、基于信道反转法的功率控制以及基于梯度和信道感知的动态设备调度，该机制实现了最高的测试精度和最快的收敛速度，并针对不同信道条件下表现出一定鲁棒性。当信道条件相对较好时 ($\sigma_0^2 = 0.5$ 和 $\sigma_0^2 = 1$) 被测模型均可达到理想精度。然而当 $\sigma_0^2 = 3$ 时，基线1和2性能迅速恶化，因为它们均没有综合考虑信道状态与数据状态，在信道逐渐恶化时无法提供较好更新质量导致性能崩溃；基线3虽然综合考虑了信道状态和数据状态，但其传输功率无法有效适应信道变化，没有很好地消除信道衰落影响，此外基线3由于未考虑不同噪声功率下的动态调度设置而缺乏一定鲁棒性。综上所述，与基线算法相比，本文所提出的设备调度机制在不同信道条件下表现出最佳的训练精度和鲁棒性，并且能够达到接近基准算法所示的最佳测试精度。

3.3 非独立同分布数据分布性能测试

如图4所示，在非独立同分布数据分布下，噪声对于精度的影响比独立同分布数据分布下更大。但当 σ_0^2 增大时，受益于适应不同噪声功率的动态调度和辅助全面利用数据集的累积残差，该机制仍然优于基线算法。

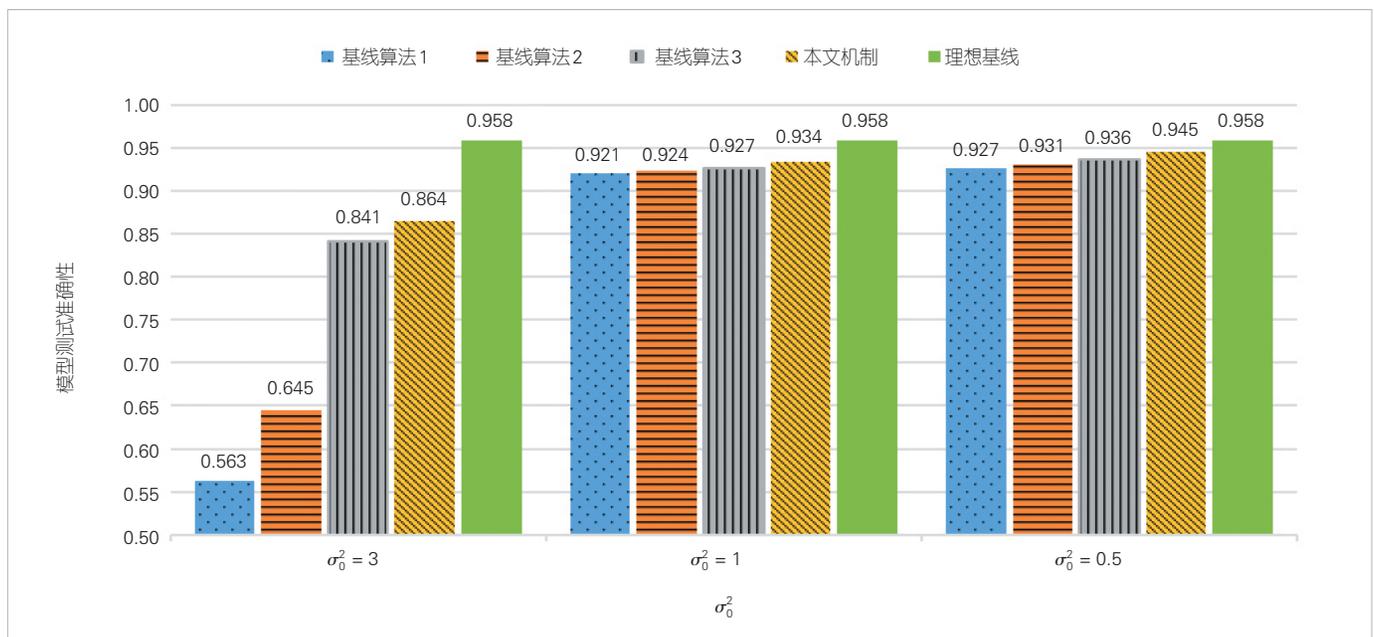
4 总结

本文研究了基于空中计算的联邦学习系统，该系统利用多址接入信道的叠加特性完成了高效数据聚合，实现了通信网与算力网的一体融合。此外，综合考虑本地模型更新信息

重要性、信道条件以及能耗水平，本文设计了基于梯度和信道感知的动态设备调度机制，在每轮训练中服务器将选择合适数量、具有较高设备质量的边缘设备参与全局模型更新过程。在该机制中，还使用了基于信道反转法的功率控制方案以抵抗信道衰落和噪声的影响。为了提高全局模型更新的效率和准确性，本文还提出了基于动态累积残差反馈的梯度传输机制。最后，为了在线求解设备调度优化问题，本文提出了基于李雅普诺夫漂移优化理论的求解算法并进行了收敛性分析。通过与不同基线算法比较，仿真结果验证了本文所提出的调度机制能够有效提高模型测试精度和收敛速度，并且针对不同信道条件下具有一定鲁棒性。

参考文献

- [1] DU J, JIANG C X, WANG J, et al. Machine learning for 6G wireless networks: carrying forward enhanced bandwidth, massive access, and ultrareliable/low-latency service [J]. IEEE vehicular technology magazine, 2020, 15(4): 122-134. DOI: 10.1109/MVT.2020.3019650
- [2] 贺倩. 人工智能技术在移动互联网发展中的应用 [J]. 信息通信技术与政策, 2017(2): 1-4
- [3] 杨强, 童咏昕, 王晏晟, 等. 群体智能中的联邦学习算法综述 [J]. 智能科学与技术学报, 2022, 4(1): 29-44. DOI: 10.11959/j.issn.2096-6652.202218
- [4] LIU Y, MA Z, YAN Z, et al. Privacy-preserving federated k-means for proactive caching in next generation cellular networks [J]. Information sciences, 2020, 521: 14-31. DOI: 10.1016/j.ins.2020.02.042
- [5] MCMAHAN H B, MOORE E, RAMAGE D, et al. Communication-



▲图4 非独立同分布数据分布下不同信道条件模型精度对比

- efficient learning of deep networks from decentralized data [EB/OL]. (2017-02-28) [2024-05-16]. <https://arxiv.org/abs/1602.05629>
- [6] ZHU Z Q, WAN S, FAN P Y, et al. Federated multiagent actor-critic learning for age sensitive mobile-edge computing [J]. IEEE Internet of Things journal, 2022, 9(2): 1053 - 1067. DOI: 10.1109/JIOT.2021.3078514
- [7] KHAN L U, SAAD W, HAN Z, et al. Federated learning for Internet of Things: recent advances, taxonomy, and open challenges [J]. IEEE communications surveys & tutorials, 2021, 23(3): 1759-1799. DOI: 10.1109/COMST.2021.3090430
- [8] NAZER B, GASTPAR M. Computation over multiple-access channels [J]. IEEE transactions on information theory, 2007, 53(10): 3498-3516. DOI: 10.1109/TIT.2007.904785
- [9] 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. DOI: 10.1109/TWC.2019.2946245
- [10] SU L Q, LAU V K N. Data and channel-adaptive sensor scheduling for federated edge learning via over-the-air gradient aggregation [J]. IEEE Internet of things journal, 2022, 9(3): 1640 - 1654. DOI: 10.1109/JIOT.2021.3096570
- [11] 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. DOI: 10.1109/JSAC.2021.3126078
- [12] 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 [J]. IEEE journal on selected areas in communications, 2023, 41(4): 1035 - 1050. DOI: 10.1109/JSAC.2023.3242727
- [13] AMIRI M M, GÜNDÜZ D, KULKARNI S R, et al. Convergence of update aware device scheduling for federated learning at the wireless edge [J]. IEEE transactions on wireless communications, 2021, 20(6): 3643-3658. DOI: 10.1109/TWC.2021.3052681
- [14] TAİK A, MLIKA Z, CHERKAoui S. Data-aware device scheduling for federated edge learning [J]. IEEE transactions on cognitive communications and networking, 2022, 8(1): 408-421. DOI: 10.1109/TCCN.2021.3100574

作者简介



江炳青，清华大学电子工程系在读硕士研究生；主要研究领域为无线通信与人工智能。



杜军，清华大学电子工程系助理研究员、硕士生导师；研究方向为异构网络智能协同与组网优化；入选中国科协青年人才托举工程，曾获中国电子学会科学技术奖技术发明一等奖；发表论文 80 余篇，授权国家发明专利 10 余项，出版专著 6 部。



王劲涛，清华大学电子工程系教授、博士生导师；研究方向为数字多媒体广播、无线光异构融合通信、AI 增强的智能通信信号处理技术；曾获国家科技进步奖一等奖、日内瓦国际发明展金奖、北京市科学技术奖一等奖；发表学术论文 180 余篇，授权国家发明专利 50 余项，出版专著 5 部。



牟林，中兴通讯移动网络和移动多媒体技术国家重点实验室高级研发总监、技术规划与行业趋势资深专家；研究方向为移动网络和移动多媒体。