基于类型的无源联邦学习与客户端自选择

Kaan Okumus*、Khac-Hoang Ngo†、Unnikrishnan Kunnath Ganesan*、
Giuseppe Durisi*、Erik G. Str¨om*和Shashi Raj Pandey‡
*电气工程系，查尔姆斯理工大学，41296哥德堡，瑞典
†电气工程系，Link¨oping大学，58183 Link¨oping，瑞典
‡电子系统系，奥尔堡大学，9220奥尔堡，丹麦

arXiv:2602.06601v1 [cs.IT] 6 Feb 2026

摘要——**我们研究数据异质性条件下无线网络中联邦学习的客户端选择问题。现有的客户端选择方法通常依赖服务器端对客户端特定信息的了解，从而损害隐私。为克服这一问题，我们提出一种客户端自选择策略，该策略仅基于本地计算的训练损失与集中更新的选择阈值之间的比较。此外，为支持客户端更新在无线信道上的鲁棒聚合，我们将该客户端自选择策略集成到最近提出的、面向分布式多输入多输出（D-MIMO）网络的基于类型的无源多址框架中。所得方案完全无源：服务器无需知道客户端的身份。此外，无论在客户端还是服务器端，都不需要信道状态信息。在分布式多输入多输出无线网络上进行的仿真结果表明，所提出的自选择策略与一种可比的先进服务器端选择方法性能相当，并始终优于随机客户端选择。**

一、引言

未来的无线网络有望通过海量联网设备支持智能服务。一项关键应用是空中联邦学习：多个设备（也称为客户端）在中央服务器的协调下协同训练模型，同时保护隐私。在每个训练轮次中，客户端接收全局模型，执行本地训练，并通过共享无线介质向服务器发送更新。由于通信资源有限且连接具有间歇性，这些传输并非持续进行。联邦平均（FedAvg） [1] 是一种基础方法，通过多个轮次迭代执行上述步骤。

当客户端数量较大时，在每一轮次中选择其中一部分变得至关重要，以降低通信与计算开销。随机客户端选择简单且被广泛使用。然而，在数据异质性下它会变得低效，因为它会导致“客户端漂移”：随着训练轮次的推进，每次本地更新都会被拉向局部最优，从而导致收敛缓慢且不稳定 [2]。它还会带来参与差距 [3]，，该差距刻画了训练期间观测到的数据与全局数据分布之间的分布偏移。

SCAFFOLD算法 [2] 使用客户端与服务器之间交换的控制变量（例如本地梯度和聚合梯度）来缓解客户端漂移。然而，该方法使每轮通信开销加倍；而且，在参与不定期的情况下，其性能会下降，因为本地控制变量往往会过时 [4]。如果所有客户端都与服务器共享留出数据，就可以估计参与差距，从而实现能够改善收敛性的客户端选择算法 [5]。然而，共享留出数据会侵犯客户端隐私。择优策略（PoC） [6] 通过让服务器根据上报的训练损失选择客户端，在一定程度上缓解了这一问题。但这些损失仍会泄露客户端特定信息。如何设计真正保护隐私的客户端选择机制，仍是一项有待解决的挑战。

迄今为止所述的所有工作都假设客户端与服务器之间的通信信道无误差。现有的空中联邦学习研究主要依赖模拟空中计算（AirComp）技术，例如 [7]–[10]，这类技术利用波形叠加进行模型聚合，但要求严格同步、信道预均衡以及专用模拟硬件。数字空中计算方法，例如 [11]–[13]，，保留了数字调制，但仍受限于低分辨率标量量化，因此难以随模型维度和客户端数量的增加而扩展。为克服这些局限，Qiao 等人等 [14] 提出了大规模数字（MD）空中计算框架，将本地更新的传输与聚合表述为一个无源多址接入（UMA） [15] 问题。在该框架中，所有客户端使用同一个编码器，服务器需要恢复（量化后的）本地更新集合。具体而言，服务器需要估计不同本地更新的数量及其出现次数，即其类型。这项任务属于基于类型的无源多址接入（TUMA） [16]框架。在 MD-空中计算 [14]，中，类型估计由近似消息传递（AMP）解码器执行，该解码器要求客户端具备信道状态信息（CSI），以进行信道预均衡。文献 [17]提出了一种无CSI的基于AMP的类型估计器，用于分布式多输入多输出（D-MIMO）网络中的通信。

本研究部分由瑞典研究理事会通过项目资助（项目编号：2021-04970 和 2022-04471）以及瑞典战略研究基金会资助。K.-H. Ngo 的研究部分得到了 Link¨oping–Lund 信息技术卓越中心（ELLIIT）的资助。S.R. Pandey 的研究部分得到了 DFF-Forskningsprojekt1“NETML”项目的资助，项目编号为 4286-00278B。

贡献： 本文探讨了在异构数据和客户端活动间歇发生的情况下，如何在分布式多输入多输出网络上的联邦学习中，同时实现保护隐私的客户端选择与可靠的本地更新

传输。我们首先提出一种客户端自选择策略，各客户端通过将本地训练损失与服务器广播的选择阈值进行比较，自主决定是否参与。服务器根据估计的参与客户端数量更新该阈值，而不依赖任何客户端特定信息。随后，我们将这一客户端选择策略整合到无信道状态信息的 TUMA 框架 [17] 中，用于分布式多输入多输出网络上的模型聚合。由此得到的方案完全无源：服务器无法获知客户端的身份信息。数值结果表明，我们的无信道状态信息的 TUMA 方案取得的测试准确率与理想通信下可达到的水平相当，且优于 MD-AirComp。此外，我们提出的客户端自选择方法，其收敛性能接近 PoC [6] ，并始终优于随机客户端选择。

记号：集合用花体字母表示（例如 $\mathcal{S}$），向量用小写粗体字母表示（例如 $\boldsymbol{x}$），矩阵用大写粗体字母表示（例如 $\boldsymbol{X}$）。我们用 $\otimes$表示克罗内克积，令 $[n]=\{1,\dots,n\}$，并使用 diag(x_1，…， x_n)表示对角线元素为 x_1，…， x_n的对角矩阵。

II. 系统模型

我们考虑一种由 K个客户端和一台服务器组成的联邦学习设置。每个客户端 $k\in[K]$ 持有一个大小为 $N_k=|\mathcal{B}_k|$的本地数据集 $\mathcal{B}_k$。目标是通过最小化全局经验损失，共同训练一个全局模型 $\boldsymbol{w}\in\mathbb{R}^W$

$$\boldsymbol{w}^\star=\arg\min_{\boldsymbol{w}\in\mathbb{R}^W}\sum_{k=1}^{K}\frac{N_k}{N_{\text{tot}}}f_k(\boldsymbol{w};\mathcal{B}_k),\tag{1}$$

其中 $N_{\text{tot}}=\sum_{k=1}^{K}N_k$，且 $f_k(\boldsymbol{w};\mathcal{B}_k)$表示客户端 k的本地损失。作为基线，我们采用 FedAvg 算法 [1]，同时将其扩展到如下所述的无源通信设置。

训练按轮次 $t\in[T]$进行。我们假设，在每轮 t开始时，每个客户端都会以概率 λ独立激活。服务器通过理想的下行信道向所有活跃客户端广播当前全局模型 $\boldsymbol{w}^{(t-1)}$。随后，每个活跃客户端根据本地选择策略决定是否参与该轮（并成为被选中客户端）。由此得到的参与客户端集合记为 $\mathcal{S}^{(t)}\subseteq[K]$，其基数记为 $L^{(t)}$。

每个被选中的客户端 $k\in\mathcal{S}^{(t)}$ 使用小批量随机梯度下降进行本地训练，训练 E轮，采用小批量大小 V和学习率 η:

$$\boldsymbol{w}_{k,\ell}^{(t)}=\boldsymbol{w}_{k,\ell-1}^{(t)}-\frac{\eta}{V}\sum_{n\in\mathcal{E}}\nabla f_k(\boldsymbol{w}_{k,\ell-1}^{(t)};n),\quad \ell\in[E],\tag{2}$$

其中，$\mathcal{E}\subset\mathcal{B}_k$ 是随机采样的小批量，$\boldsymbol{w}_k^{(t)}$. $\boldsymbol{w}^{(t-1)}$ E k $_0=$ 。经过若干训练轮数后，客户端获得本地模型 $\boldsymbol{w}_k^{(t)}=\boldsymbol{w}_{k\,E.}^{(t)}$，并计算本地模型更新

$$\boldsymbol{\Delta}_k^{(t)}=\boldsymbol{w}_k^{(t)}-\boldsymbol{w}^{(t-1)}.\tag{3}$$

客户端 k 随后将 $\boldsymbol{\Delta}_k^{(t)}$ 编码为码字，并通过无线信道传输该码字。

服务器接收信号 $\boldsymbol{Y}_{(t)}$，并利用它来估计传输的本地更新 $[\widehat{\boldsymbol{\Delta}}_1^{(t)},\dots,\widehat{\boldsymbol{\Delta}}_{\widehat{L}^{(t)}}^{(t)}]$。其中，$\widehat{L}^{(t)}$ 是服务器端对 $L^{(t)}$的估计。随后，服务器按如下方式更新全局模型：[1]

$$\boldsymbol{w}^{(t)}=\boldsymbol{w}^{(t-1)}+\eta_{\text{g}}\frac{1}{\widehat{L}^{(t)}}\sum_{i=1}^{\widehat{L}^{(t)}}\widehat{\boldsymbol{\Delta}}_i^{(t)},\tag{4}$$

其中 η_{g} 是全局学习率。

我们的总体目标是设计一种本地客户端选择策略以及一种通信方案，使得 $\boldsymbol{w}^{(t)}$ 快速收敛到 $\boldsymbol{w}^\star$。

III. 所提出的客户端自选择策略

在本节中，我们提出一种客户端选择算法，并在无差错通信条件下评估其性能。随后，我们将在第四节中研究在实际分布式多输入多输出部署下可达到的性能。

。

A. 自主选择算法

我们提出一种客户端自主选择机制，用以模拟PoC选择规则 [6]的行为。不过，与PoC不同，在PoC中服务器依据报告的效用选择客户端；而在这里，每个活跃客户端自主决定是否参与，且不向服务器透露任何客户端特定信息。具体而言，每个客户端通过将本地训练损失与服务器广播的阈值进行比较来决定是否参与。服务器更新该阈值，以调整预期的参与者数量。接下来将给出所提出策略的详细描述。

在轮次 t 开始时，每个活跃客户端都会接收服务器广播的当前选择阈值 $\theta^{(t)}$ 和全局模型 $\boldsymbol{w}^{(t-1)}$。

1) 候选集构建：每个活跃客户端以概率 $p_{\text{cand}}=d/(\lambda K)$ 加入候选集，因此候选集平均包含 d 个客户端。这确保每个轮次的决策都基于由多样化客户端组成的子集，从而防止过拟合，正如 [6] 中所述。

2) 本地评估与自主选择：每个候选客户端在接收到的全局模型上计算其本地损失，

$$\ell_k^{(t)}=f_k(\boldsymbol{w}^{(t-1)};\mathcal{B}_k),\tag{5}$$

并以概率

$$\Pr\{k\in\mathcal{S}^{(t)}\mid\ell_k^{(t)}\}=\sigma\Big(a\big(\ell_k^{(t)}-\theta^{(t)}\big)\Big),\tag{6}$$

其中，Sigmoid函数 $\sigma(x)=1/(1+e^{-x})$ 用于将本地损失改善 $\ell_k^{(t)}-\theta^{(t)}$ 转换为参与概率，而 $a>0$ 控制概率值从 0 到 1 变化的陡峭程度。具体而言，较小的a 会引入随机性，有助于避免反复选择相同的高损失客户端，从而防止过拟合。

[1]请注意，由于我们的传输策略（见第 IV 节）确保服务器无法获知客户端身份，因此式 (1) 中使用的按数据集大小加权平均无法应用于式 (4)。

算法1去中心化客户端自选择

1：服务器：广播（$\boldsymbol{w}^{(t-1)}$，$\theta^{(t)}$）
2：**对于**每个活跃客户端 k **执行**
3：　以概率 $p_{\text{cand}} = d/(\lambda K)$ 加入候选集
4：　**如果**候选客户端**则**
5：　　计算本地损失 $\ell_k^{(t)} = f_k(\boldsymbol{w}^{(t-1)}; \mathcal{B}_k)$
6：　　以一定概率自行选择　　$\sigma(a(\ell_k^{(t)} - \theta^{(t)}))$
7：　　**如果** 被选中，**则** 进行本地训练并传输 $\boldsymbol{\Delta}_k^{(t)}$
8：　**结束 if**
9：**结束 if**
10: **结束 for**
11: **服务器：** 估计 $\widehat{L}^{(t)}$ 并更新 $\theta^{(t+1)}$ 通过（7）

ran穹顶，$|\mathcal{S}(t)| = 99.7 \pm 9.1$

S 自我选择，$|\mathcal{S}(t)| = 97.2 \pm 20.3$

PoC，$|\mathcal{S}^{(t)}| = 100$

00 150 200 250 300 350 400 450 500

联邦学习 ro und t

3）阈值更新：聚合后，服务器会更新阈值，使每轮参与客户端的平均数量接近目标值 K_{tar}：

$$\theta^{(t+1)} = \theta^{(t)} + \xi(\widehat{L}^{(t)} - K_{\text{tar}}). \tag{7}$$

其中，$\xi > 0$ 是步长参数。估计的参与率 $\widehat{L}^{(t)}$ 高于 K_{tar} 时，阈值会提高，从而降低未来的参与概率。

整体策略总结见算法1。与本地训练相比，该策略的客户端计算开销可以忽略不计，因为每轮只需进行一次损失评估和一次标量阈值比较。

B. 完美通信下的性能评估

我们在无误差通信场景下评估所提出的自选择策略。在该场景中，本地更新无误差地传送至服务器。我们采用 FMNIST 数据集，该数据集包含 60 000 个样本，并按 0.8、0.1 和 0.1 的比例分别划分为训练集、验证集和测试集。模型为一个多层感知机，输入维度为 784，对应于展平后的 28×28FMNIST 图像。模型包含两个全连接隐藏层，大小分别为 64 和 30，且每层后均接一个 ReLU激活。由此得到的模型维度为 $W = 52\,500$。为提高泛化能力，在第二个隐藏层和输出层之间加入一个速率为 0.5 的丢弃层。最后一层的输出采用 LogSoftmax激活。损失函数为负对数似然。每个被选中的客户端每轮进行 $E = 30$ 轮本地训练，小批量大小为 $V = 64$，本地学习率为 $\eta = 0.001$，全局学习率为 $\eta_{\text{g}} = 1$。使用参数为$\alpha = 2$的狄利克雷分布将数据随机分配给各客户端，从而形成具有中等异质性的本地数据集。我们将客户端总数设为 $K = 1000$，激活概率设为 $\lambda = 0.8$。按照 [6], ，我们选择参与比例 $K_{\text{tar}}/K = 0.1$，由此得到 $K_{\text{tar}} = 100$。训练共进行 $T = 500$ 轮。

作为基准测试，我们考虑两种替代策略：

- **随机：** 每个活跃客户端以 $K_{\text{目标}}/(\lambda K)$ 的概率独立选择自身。

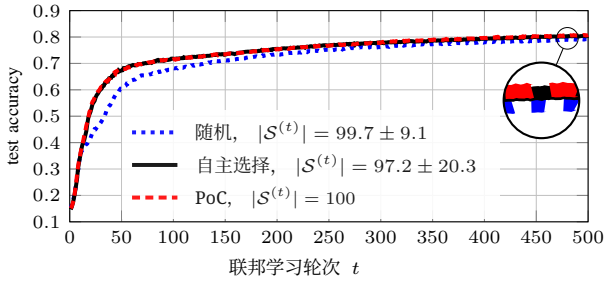

图 1. 在无误差通信条件下，不同选择策略对应的测试准确率随联邦学习轮次 t 的变化情况；这些策略产生的参与客户端平均数量 $K_{\text{tar}} = 100$ 大致相同。图例中列出了每种选择策略对应的实际参与客户端平均数量 $|\mathcal{S}^{(t)}|$ 及其标准差。

- **PoC：** 选择在服务器端进行，依据报告的最高本地损失[6]。我们将候选集的大小设置为 $d = 2 \times K_{\text{tar}} = 200$。

.2 ± 20.3

对于我们在 (6) 中设置 $a = 50$ 、在 (7) 中设置 $\theta^{(1)} = 2.32$，$\xi = 0.004$ 的情况，我们将这两种策略与所提出的自选择策略进行比较。[2] 与 PoC 类似，我们设定$d = 200$，因此 $p_{\text{cand}} = d/(\lambda K) = 0.25$。

图 1 展示了无误差通信条件下，三种选择策略的测试准确率随联邦学习轮次变化的情况。所提出的客户端自选择策略表现出与 PoC [1] 平相同的收敛行为，同时优于随机选择。这表明，客户端自选择能够在不损害隐私的情况下，达到与服务器端选择相当的性能。

IV. 基于类型的无源联邦学习

本节讨论如何通过无线信道传输本地模型更新。具体而言，我们考虑一种分布式多输入多输出场景，其中所有参与用户同时传输各自的更新，并使用 TUMA 解码器 [17] 来估计量化后的本地更新。

A. 本地更新量化

遵循 [14], [18], ，我们假设每个客户端都通过带误差累积的矢量量化，对其本地模型更新 $\boldsymbol{\Delta}_k^{(t)}$ 进行量化后再传输。设 $\mathcal{Q} = \{\boldsymbol{q}_1, \ldots, \boldsymbol{q}_M\}$ 为量化码本，其中每个$\boldsymbol{q}_m \in \mathbb{R}^Q$ 表示一个维度为 Q的量化码字。请注意，$Q = 1$ 对应标量量化。量化位数为 $J = \log_2 M$。在第 t轮，量化过程如下。设 $\boldsymbol{e}_k^{(t)}$ 表示客户端 k 在第 t轮的累积量化误差，其初始值为 $\boldsymbol{e}_k^{(0)} = \boldsymbol{0}$。客户端首先通过计算，将累积量化误差纳入模型更新中

$$\bar{\boldsymbol{s}}_k^{(t)} = \boldsymbol{\Delta}_k^{(t)} + \boldsymbol{e}_k^{(t-1)}. \tag{8}$$

[2]参数 a、$\theta^{(1)}$ 和 ξ 通过经验性超参数调优进行选择。

然后，将 $\bar{\boldsymbol{s}}_k^{(t)}$ $D = \lceil W/Q \rceil$ [3]
$\{\bar{\boldsymbol{s}}_k^{(d,t)}\}_{d=1}^{D}$
划分为连续的子向量，并根据 将每个子向量量化为最近的码字

$$M_k^{(d,t)} = \arg\min_{m\in[M]} \left\| \boldsymbol{q}_m - \bar{\boldsymbol{s}}_k^{(d,t)} \right\|_2, \quad d \in [D]. \tag{9}$$

量化后的子向量将重新拼接成长度为 W 的向量[4]，如下所示：

$$\boldsymbol{s}_k^{(t)} = \mathrm{concat}\left(\boldsymbol{q}_{M_k^{(1,t)}}, \ldots, \boldsymbol{q}_{M_k^{(D,t)}} \right). \tag{10}$$

最后，更新将延续到下一轮的量化误差：

$$\boldsymbol{e}_k^{(t)} = \bar{\boldsymbol{s}}_k^{(t)} - \boldsymbol{s}_k^{(t)}. \tag{11}$$

因此，每个客户端都会生成一系列量化索引 $\{M_k^{(d,t)}\}_{d=1}^{D}$，并通过 D 个子轮次进行传输。具体而言，这些索引将根据下文所述的 TUMA 编码方案映射为通信码字。

B. 分布式多输入多输出网络中的 TUMA

我们考虑一个分布式多输入多输出网络，其中 B 个接入点（AP）各配备 A 根天线（总计 $F = A \times B$ 根天线），并通过前传链路连接到服务器。接入点共同为随机分布在覆盖区域 $\mathcal{D}$内的单天线客户端提供服务。该区域被划分为 U 个互不重叠的区域 $\{\mathcal{D}_u\}_{u=1}^{U}$。每个区域由大尺度衰落系数（LSFC）相近的位置组成，因此同一区域内的用户到接入点的平均路径损耗近似相同。这种划分可减轻路径损耗变化的影响 [19]。我们将每个接入点 b 的位置记为 $\nu_b \in \mathcal{D}$。

所有客户端使用同一码本$\boldsymbol{C} \in \mathbb{C}^{N\times M}$，其中 N为码长，$M = U \times M$ 为码字总数。码本均匀划分为按区域划分的子码本：$\boldsymbol{C} = [\boldsymbol{C}_1, \ldots, \boldsymbol{C}_U]$，其中 $\boldsymbol{C}_u = [\boldsymbol{c}_{u,1}, \ldots, \boldsymbol{c}_{u,M}]$，且每个码字均满足平均功率约束$\|\boldsymbol{c}_{u,m}\|_2^2 \le 1$。量化索引到通信码字的映射是一一对应且取决于区域：区域 u中的客户端 k将量化索引 $M_{(d,\,t)}$ $_k \in [M]$映射为码字$\boldsymbol{c}_{u,\,M_k^{(d,t)}}$，并在第 t轮的第 d个子轮次中发送。

我们将第 t 轮的第 d 个子轮次中，区域 u 内发送码字 m 的客户端数量记为 $k_{u\,m}^{(d,t)}$。此外，我们用 $\boldsymbol{k}_u^{(d,t)} = [k_{u,1}^{(d,t)}, \ldots, k_{u,M}^{(d,t)}]^\top$ 表示相应的按区域划分的重数向量，并用$[k_1^{(d,t)}, \ldots, k_M^{(d,t)}]^\top$ $k_m^{(d,t)} = \sum_{u=1}^{U} k_{u\,m}^{(d,t)}$

（其中，）表示全局重数向量。注意，$\|\boldsymbol{k}^{(d,t)}\|_1$ 是第 t轮中参与客户端的数量。因此，类型 $\boldsymbol{t}^{(d,t)}$ 可表示为归一化重数向量，即 $\boldsymbol{t}^{(d,t)} = \boldsymbol{k}^{(d,t)}/\|\boldsymbol{k}^{(d,t)}\|_1$。该向量表示所有参与客户端所发送量化更新的经验分布。

[3]在进行此划分之前，如果 Q 不能整除 W，则会对 $\bar{\boldsymbol{s}}_k^{(t)}$ 进行零填充，以确保所有 D 子向量的长度 Q相等。[4]如果需要进行零填充，则会执行截断。

我们用 (u,k) 标记区域 u中参与的第 k个客户端。客户端 (u,k) 与接入点 b 之间的信道建模为空间白准静态瑞利衰落信道。具体而言，假设信道系数在不同天线、接入点和客户端之间相互独立，并且在每个码字期间保持不变。客户端 (u,k) 与所有接收天线之间的信道向量 $\boldsymbol{h}_u^{(d,t)}$，$_k \in \mathbb{C}^F$ 服从 $\mathcal{CN}(0, \Sigma(\rho_{u,k}))$ 分布，其中 $\rho_{u,k}$ 表示用户 (u,k) 的位置，$\Sigma(\rho_{u,k}) = \mathrm{diag}(\gamma_1(\rho_{u,k}) \ldots, \gamma_B(\rho_{u,k})) \otimes \boldsymbol{I}_A$，其中 $\gamma_b(\rho_{u,k})$ 为位置相关的大尺度衰落系数。当多个用户发送同一码字时，它们的贡献会在接收机处叠加。对于码字 $\boldsymbol{c}_{u,m}$，其有效信道向量定义为

$$\boldsymbol{x}_{u,m}^{(d,t)} = \begin{cases} \sum_{(u,k):\, M_k^{(d,t)}=m} \boldsymbol{h}_{u,k}^{(d,t)} & \text{if } k_{u,m}^{(d,t)} > 0, \\ \boldsymbol{0} & \text{if } k_{u,m}^{(d,t)} = 0. \end{cases} \tag{12}$$

我们定义有效信道矩阵 $\boldsymbol{X}_u^{(d,t)}$，它收集区域 u 中所有码字对应的有效信道向量，记为 $\boldsymbol{X}_u^{(d,t)} = [\boldsymbol{x}_{u,1}^{(d,t)}, \ldots, \boldsymbol{x}_{u,M}^{(d,t)}]^\top \in \mathbb{C}^{M\times F}$。所有接入点接收到的聚合信号为

$$\boldsymbol{Y}^{(d,t)} = \sqrt{NP} \sum_{u=1}^{U} \boldsymbol{C}_u \boldsymbol{X}_u^{(d,t)} + \boldsymbol{W}^{(d,t)}, \tag{13}$$

其中 $\boldsymbol{W}^{(d,t)}$ 为加性白高斯噪声，其元素独立地服从 $\mathcal{CN}(0,\ \sigma_w^2)$ 分布，而 P 表示平均发射功率。

接收矩阵 $\boldsymbol{Y}^{(d,t)}$ 作为 TUMA 解码器的输入。解码器在多个子轮次$d \in [D]$中运行，并在每个子轮次中给出对该子轮次所传输的量化本地更新类型的估计值 $\widehat{\boldsymbol{t}}^{(d,t)}$。根据估计出的类型，服务器对子块 d 执行全局模型更新，如下所示（见式（4））

$$\boldsymbol{w}_d^{(t)} = \boldsymbol{w}_d^{(t-1)} + \eta_{\mathrm{g}} \sum_{m=1}^{M} \widehat{t}_m^{(d,t)}\, \boldsymbol{q}_m. \tag{14}$$

处理完所有子轮次后，将基于子块更新的全局模型 $\{\boldsymbol{w}_d^{(t)}\}_{d=1}^{D}$ 拼接起来（并可能截断至长度 W），从而得到更新后的全局模型 $\boldsymbol{w}^{(t)}$。

作为TUMA解码器，我们采用 [17,算法 1], 中描述的集中式多源AMP算法，该算法通过贝叶斯去噪器迭代估计重数向量 $\widehat{\boldsymbol{k}}_u^{(d,t)}$。在缺乏客户端数据分布或参与统计的先验知识时，我们采用截断泊松先验

$$k_{u,m} \sim \mathrm{Pois}_{[0:K_{\max}]}\left(\frac{K_{\mathrm{tar}}}{U \times M} \right), \quad K_{\max} = \left\lceil \frac{K_{\mathrm{tar}}}{U} \right\rceil, \tag{15}$$

其中，$\mathrm{Pois}^{[0,a]}(\alpha)$ 表示均值为 α、截断至区间 $[0,a]$ 的泊松分布。该先验通过反映在给定轮次内每个区域中只有有限数量的码字处于激活状态，促进稀疏性。关于 TUMA 解码器的更多细节，请参阅 [17, 第 3] 节。

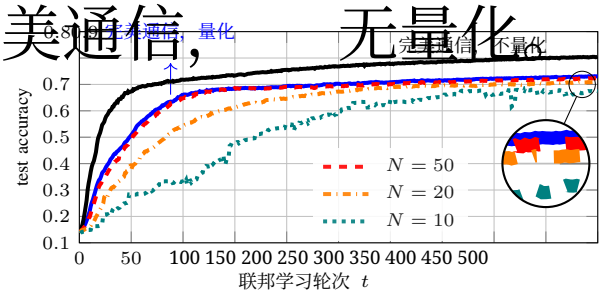

图2。TUMA解码器采用客户端自选择时，测试准确率随联邦学习轮次数量 t 的变化；我们考虑参数为 $J = 7$、 $Q = 30$ 的矢量量化，以及不同的传输块长值 $N \in \{10, 20, 50\}$。

V. 分布式多输入多输出网络上的性能

现在，我们将第III节提出的客户端选择策略与第IV节所述分布式多输入多输出网络中的基于TUMA的模型聚合相结合，并评估所得结果在第III-B节所考虑的相同联邦学习问题上的测试准确率。

为构建矢量量化码本，我们遵循 [14]中所述的过程。具体而言，我们假设服务器拥有一个单独的数据集，仅用于更新量化码本 $\mathcal{Q}$。[5] 在每一轮次中，服务器执行一个本地训练步骤，并获得一个带有误差累积的本地更新，然后将其划分为 D 个维度为 Q的子向量。服务器使用这些子向量作为样本，利用K-means++ 算法 [20]更新量化码本。具体而言，服务器找到 $M = 2^J$ 个聚类质心，并将其用作量化码字。更新后的码本与全局模型一起广播给客户端。

我们考虑一种简单的分布式多输入多输出部署，其采用 3×3 方形网格布局，每个正方形的边长为100 m。具体而言，我们假设 $B = 40$ 个接入点沿正方形边界均匀放置，每个接入点配备 $A = 4$ 根天线。因此，接收天线总数为 $F = 160$ 。与 [19]，类似，大尺度衰落系数建模为 $\gamma_b(\rho) = 1/(1 + (|\rho - \nu_b|/d_0)^\alpha)$，其中 $\alpha = 3.67$ 和$d_0 = 13.57$ m。覆盖区域划分为 $U = 9$个区域，每个区域对应一个正方形。我们将接收信噪比定义为 $\text{SNR}_{\text{rx}} = \text{SNR}_{\text{tx}}/(1 + (\varsigma/d_0)^\alpha)$，其中 $\varsigma = 50$ m 表示区域质心与距离其最近的接入点之间的距离。在所有仿真中，我们固定 $\text{SNR}_{\text{rx}} = 10$ dB，通信的平均每符号发射功率为 $P = 1$ mW。码字 $\{\boldsymbol{c}_{u,m}\}$ 独立生成：首先从高斯分布 $\mathcal{CN}(0, 1/N)$中独立采样每个元素，然后对每个码字进行归一化，使得$\|\boldsymbol{c}_{u,m}\|_2^2 = 1$。我们采用式（15）中的截断泊松先验，其中 $K_{\max} = 8$，并采用第 III-B 节中指定的学习超参数。[6]

[5]该数据集的存在不会侵犯隐私，因为它不同于客户端的本地数据集。

[6]用于获得所有结果的代码可在 https://github.com/ okumuskaan/ufl tuma 获取，以支持可复现研究。

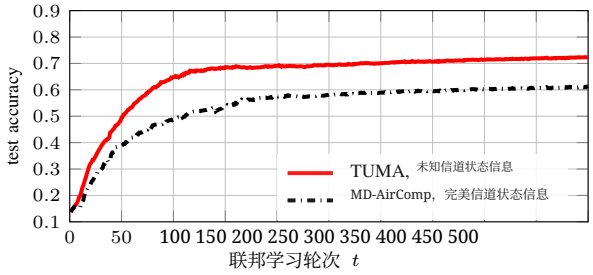

图3. 测试准确率随联邦学习轮次数量的变化；我们比较了在完美信道状态信息下的 TUMA 和 MD-AirComp，以及 $N = 50$、$J = 7$ 、 $Q = 30$ 和客户端自选择。

图2展示了使用 TUMA 解码器和我们提出的客户端自选择策略时，三个不同通信码长下的联邦学习轮次测试准确率，即 $N \in \{10, 20, 50\}$。我们设定 $J = 7$和 $Q = 30$。[7] 此外，我们还给出了完美通信情况下的曲线作为参考，分别对应使用和不使用矢量量化的情况，以分离所选参数下由矢量量化导致的性能损失。从图中可以看出，随着码长 N 增加，TUMA 下的测试准确率有所提升，并逐渐接近完美通信时的准确率。事实上，较长的码长会带来较低的速率，因此类型估计错误也更少。值得注意的是，系统工作在 $N \ll 2^J = 128$这一工作区间，因此无法使用正交码本。

在图3中，我们将采用TUMA解码器所达到的测试准确率，与将 [14]中提出的MD-AirComp方案扩展到分布式多输入多输出设置后所达到的测试准确率进行比较。我们设置 $N = 50$。对于MD-AirComp，我们假设CSI[8]完美，并按照 [14]执行发射端信道预均衡。此外，我们对发射信号进行归一化，使接收信噪比保持在固定值$\text{SNR}_{\text{rx}} = 10$dB，从而确保与TUMA进行公平比较。如图3所示，MD-AirComp遭受了显著的性能下降。

在图4中，对于所提出的客户端自选择方法，当 $N \in \{20, 50\}$ 和 Q =30时，我们展示了经过100轮后的测试准确率随量化位数 J 变化的情况。对于 $N = 50$，在图中所考虑的 J 取值范围内，测试准确率随 J单调增加。相反，对于 $N = 20$，其变化并不单调，且测试准确率在 $J = 8$处达到峰值。这种现象可以通过量化失真与通信错误之间的权衡来解释：前者在 J 增加时减小，而后者在 J 增加时增大。请注意，该最优工作点对应的 J 取值远大于使用正交码本时所允许的取值。

在表I中，我们总结了所提框架在不同通信条件和客户端选择策略下的端到端性能。对于 $N = 50$,

[7] Q取值较低、 J 取值较高时，量化精度更高，但通信开销也会增加，通信性能则会下降，因为较小的 Q 会导致更多通信子轮次，而较大的 J 会增加传输比特数，进而提高传输速率。8

请注意，TUMA编码方案则不需要任何瞬时信道状态信息。

表I 不同场景下的端到端性能

场景 / 编码方案	$(J,\ Q)$	N	选择	最终测试准确率 (%)	达到70%所需的轮数	$\lvert\mathcal{S}^{(t)}\rvert$
完美通信，无量化	–	∞	Self	80.5	73	97.2 ± 20.3
完美通信，量化	(7，30)	∞	Self	73.1	236	97.2 ± 21.6
TUMA	(7，30)	50	**Self**	72.3	295	94.4 ± 34.1
TUMA	(7，30)	50	PoC	72.8	284	100
TUMA	(7，30)	50	随机	70.0	478	99.7 ± 9.1
MD-AirComp	(7，30)	50	Self	61.4	–	89.6 ± 43.4
TUMA	(5，30)	50	Self	70.1	466	94.3 ± 35.3
TUMA	(7，30)	20	Self	70.9	430	90.1 ± 40.3

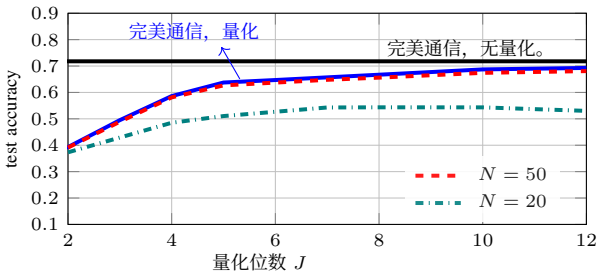

图4。在第 $t = 100$ 轮达到的测试准确率与量化位数 J 的关系，适用于TUMA解码器和客户端自选择； $Q = 30$。

所提出的自选择策略优于随机选择，尤其是在收敛速度方面；其性能接近 PoC，这证实了该策略能够在不牺牲准确性的前提下有效保护隐私。

VI. 结论

我们提出了一种用于空中联邦学习的客户端自选择策略，客户端根据本地计算的训练损失和集中更新的阈值决定是否参与。我们将该策略与基于 TUMA 的数字空中聚合相结合，并应用于分布式多输入多输出无线网络。所得到的方案不需要信道状态信息，并依据无源通信范式运行，在该范式中，服务器无需知道客户端的身份。我们的仿真结果表明，所提出的方法可达到与 PoC [6] 相匹配的测试准确率；PoC [6] 是一种基于客户端报告的训练损失的最先进的服务器端选择方法。在给定通信和计算约束的情况下，优化系统参数以最大化测试准确率，是一个值得研究的开放问题。

参考文献

[1] B. McMahan、E. Moore、D. Ramage、S. Hampson 和 B. Aguera y Arcas，“基于去中心化数据的深度网络通信高效学习”，载于国际人工智能与统计学会议论文集（*AISTATS*），美国佛罗里达州劳德代尔堡，2017年4月，第1273–1282页。[2] S. P. Karimireddy、S. Kale、M. Mohri、S. Reddi、S. Stich 和 A. T. Suresh，“SCAFFOLD：联邦学习的随机受控平均”，载于国际机器学习会议论文集（*ICML*），2020年7月，第5132–5143页。[3] H. Yuan、W. R. Morningstar、L. Ning 和 K. Singhal，“我们所说的联邦学习泛化究竟意味着什么？”，载于国际学习表征会议论文集（*ICLR*），2022年4月。[4] S. Reddi、Z. B. Charles、M. Zaheer、Z. Garrett、K. Rush、J. Kone˘cn´y、S. Kumar 和 B. McMahan，“自适应联邦优化”，载于国际学习表征会议论文集（*ICLR*），2021年5月。

[5] P. Singhal、S. R. Pandey 和 P. Popovski，“面向通信高效联邦学习的贪心 Shapley 客户端选择”，IEEE《网络通信快报》，第6卷，第2期，第134–138页，2024年2月。[6] Y. Jee Cho、J. Wang 和 G. Joshi，“理解联邦学习中的有偏客户端选择”，载于国际人工智能与统计学会议论文集（AISTATS），2022年3月，第10 351–10 375页。[7] M. M. Amiri 和 D. G¨und¨uz，“无线衰落信道上的联邦学习”，IEEE《无线通信汇刊》，第19卷，第5期，第3546–3557页，2020年5月。[8] G. Zhu、Y. Wang 和 K. Huang，“面向低时延联邦边缘学习的宽带模拟聚合”，IEEE《无线通信汇刊》，第19卷，第1期，第491–506页，2020年1月。[9] K. Yang、T. Jiang、Y. Shi 和 Z. Ding，“基于空中计算的联邦学习”，IEEE《无线通信汇刊》，第19卷，第3期，第2022–2035页，2020年3月。[10] Y.-S. Jeon、M. M. Amiri、J. Li 和 H. V. Poor，“面向大规模MIMO通信系统中联邦学习的压缩感知方法”，IEEE《无线通信汇刊》，第20卷，第3期，第1990–2004页，2021年3月。[11] G. Zhu、Y. Du、D. G¨und¨uz 和 K. Huang，“面向通信高效联邦边缘学习的一比特空中聚合：设计与收敛分析”，IEEE《无线通信汇刊》，第20卷，第3期，第2120–2135页，2021年3月。[12] A. S¸ahin，“非相干多数投票计算下无线网络中的分布式学习”，IEEE《无线通信汇刊》，第22卷，第11期，第8020–8034页，2023年11月。[13] ——，“基于平衡数制的联邦边缘学习空中计算”，IEEE《无线通信汇刊》，第23卷，第5期，第4564–4579页，2024年5月。[14] L. Qiao、Z. Gao、M. Boloursaz Mashhadi 和 D. G¨und¨uz，“面向通信高效联邦边缘学习的大规模数字空中计算”，IEEE《精选通信领域期刊》，第42卷，第11期，第3078–3094页，2024年11月。[15] Y. Polyanskiy，“关于大规模随机接入的观点”，载于IEEE国际信息论研讨会论文集（ISIT），德国亚琛，2017年6月，第2523–2527页。[16] K.-H. Ngo、D. P. Krishnan、K. Okumus、G. Durisi 和 E. G. Str¨om，“基于类型的免授权多址接入”，载于IEEE无线通信信号处理进展研讨会论文集（SPAWC），意大利卢卡，2024年9月，第911–915页。[17] K. Okumus、K.-H. Ngo、G. Durisi 和 E. G. Str¨om，“无蜂窝大规模MIMO衰落信道中的基于类型的免授权多址接入”，载于IEEE国际信息论研讨会论文集（ISIT），美国密歇根州安娜堡，2025年6月。[18] M. M. Amiri 和 D. G¨und¨uz，“无线边缘的机器学习：空中分布式随机梯度下降”，载于IEEE国际信息论研讨会论文集（ISIT），法国巴黎，2019年7月，第1432–1436页。[19] B. C¸akmak、E. Gkiouzepi、M. Opper 和 G. Caire，“无蜂窝用户中心无线网络中免授权随机接入的联合消息检测与信道估计”，IEEE《信息论汇刊》，第71卷，第5期，第3614–3643页，2025年5月。[20] D. Arthur 和 S. Vassilvitskii，“k-means++：精心选择初始中心的优势”，斯坦福信息实验室，技术报告2006-13，2006年6月。[在线]。可访问：http://ilpubs.stanford.edu:8090/778/