边缘计算中通过彩票票据剪枝实现隐私增强的高效联邦学习

施一帆、韦康、*IEEE*会员，沈力、李军、*IEEE*高级会员，王雪倩，*IEEE*会员，袁波，*IEEE*高级会员，以及郭嵩，*IEEE*会士

arXiv:2305.01387v1 [cs.DC] 2 May 2023

摘要—联邦学习（FL）是一种针对移动终端（MTs）分散私有数据的协同学习范式。然而，它面临通信、MT资源和隐私方面的问题。现有的隐私保护型FL方法通常采用实例级差分隐私（DP），虽能提供严格的隐私保证，却存在若干瓶颈：性能严重下降、传输开销，以及边缘设备（如MT）的资源限制。为克服这些缺点，我们提出Fed-LTP，这是一种结合**彩票假**设（彩票假设（LTH）和零集中差分**隐私**（zCDP）的高效隐私增强型FL框架。该框架在服务器端生成剪枝后的全局模型，并在客户端从头开始进行基于zCDP的稀疏到稀疏训练。在服务器端，我们提出两种剪枝方案：（i）基于权重的剪枝（LTH）确定剪枝后的全局模型结构；（ii）迭代剪枝进一步缩减剪枝模型的参数规模。同时，通过基于拉普拉斯机制的模型验证，Fed-LTP的性能也得到提升。在客户端，我们采用稀疏到稀疏训练来解决资源限制问题，并提供更紧密的隐私分析以降低隐私预算。我们在多个真实世界数据集上，在独立同分布（IID）和非独立同分布设置下评估Fed-LTP的有效性。结果清楚地证明，与当前最先进（SOTA）方法相比，Fed-LTP在通信、计算和内存效率方面更具优势，同时实现了更优的效用-隐私权衡。

索引词—联邦学 习、差分隐私、彩票假设、零集中差分隐私、移动边缘计算 g

1 引言

联邦学习（FL） [1] 允许分布式客户端（例如移动终端（MTs））在云端的协调下协作训练共享模型，而无需共享本地数据[1]。然而，FL 面临若干关键挑战，例如计算资源、内存、通信带宽和隐私泄露 [3]。近期大多数研究主要关注通信成本 [2],[4]–[6] 或移动终端的资源开销 [7]–[10]。此外，好奇的服务器也可能通过精心设计的生成模型和/或影子模型推断移动终端的隐私信息，例如成员身份和数据特征 [11]–[15]。为解决隐私问题，差分隐私（DP） [16], 作为联邦学习中的事实标准，能够保护移动设备数据集中的每个样本以及移动终端之间的信息（样本级差分隐私 [17]–[20]），或者不太严格地说，仅保护移动终端之间的信息（客户端级差分隐私 [21]–[26]）。例如，银行需要采用实例级差分隐私方法，以防止每位客户的每条数据记录被识别；而移动设备中的语言预测模型只需保护数据的所有权，因此客户端级差分隐私已足够。然而，所有差分隐私方法都会引入与模型大小成正比的额外随机噪声，这可能导致严重的性能下降，尤其是对于实例级差分隐私而言。

为缓解性能下降和通信效率问题，现有的实例级差分隐私技术 [17], [18] 会在上传前采用本地更新稀疏化方法，在降低通信成本的同时改善效用-隐私权衡。然而，它们仍存在以下缺点：1）仅降低了上传（客户端到服务器）的通信成本，未考虑服务器到客户端的成本；2）移动设备的计算开销和内存占用均未改变；3）它们只关注差分隐私训练，而未考虑模型验证，因此所得模型不一定是最优的；4）稀疏化方法 [18] 具有较大的随机性，在稀疏度较高时可能导致性能下降。因此，一个关键问题是：如何设计隐私保护算法，在提升模型效用的同时，适当平衡边缘设备的计算开销、内存效率和通信效率？

为回答这一问题，我们提出了一种高效且增强隐私保护的**联邦**学习框架，

• 施一帆和王雪倩就职于清华大学深圳国际研究生院人工智能与机器人中心，中华人民共和国深圳市*518055*（电子邮件：*shiyf21@mails.tsinghua.edu.cn*；*wang.xq@sz.tsinghua.edu.cn*）。

• 沈力任职于中国北京的京东探索研究院。（电子邮箱：*mathshenli@gmail.com*）

• 韦康和郭嵩任职于香港理工大学计算学系，中国香港 *999077*（电子邮箱：{*song.guo, adam-kang.wei*}*@polyu.edu.hk*）。

• *JunLi*就职于中国南京南京理工大学电气与光学工程学院。电子邮箱：*jun.li@njust.edu.cn*。

• 袁波任职于深圳市智策科技有限公司，中国深圳市*518055*（电子邮箱：*boyuan@ieee.org*）。

1. 例如，作为联邦学习中的经典方法，联邦平均算法 [2] 使用随机梯度下降，对以分布式方式选出的移动终端进行多轮并行训练，然后聚合各移动设备的模型更新，以提升全局模型的性能。

该框架结合了**彩票假设（LTH）**和零集中差分**隐私**（zCDP）方法，命名为 Fed-LTP。Fed-LTP 的主要创新在于：**（i）**设计了两种服务器端剪枝方案，以获得高质量的初始模型：一种基于权重的剪枝方案，用于创建预训练模型，也称为中奖彩票（WT）；另一种进一步的迭代剪枝方案，用于创建剪枝程度各异的异构移动设备模型，从而进一步降低计算和通信开销；**（ii）**设计了服务器端 WT 广播机制，以确保全局模型的稳定性和收敛性，同时减轻边缘设备的巨大计算开销和内存占用；**（iii）**使用 zCDP 训练本地剪枝模型，以缓解隐私预算压力；并基于私有验证数据集使用拉普拉斯机制，在客户端获取验证分数，然后将其上传至服务器进行模型验证，以选出最佳全局模型并防止过拟合。

总而言之，我们的主要贡献有四点：

- 我们首次将彩票假设引入结合差分隐私的联邦学习，并提出一种高效且隐私增强的联邦学习框架（Fed-LTP），在兼顾双向通信成本的同时，有效缓解客户端在内存和计算方面的资源限制，并保持模型效用。
- 我们提出了两种服务器端剪枝方案：基于权重的剪枝方案和进一步的迭代剪枝方案，以优化效用、通信成本与边缘设备资源开销之间的平衡。
- 我们针对每台移动设备中的训练数据和验证数据，提出了一种新颖且紧致的隐私分析（零集中差分隐私），以提升隐私保护水平，同时保持模型效用和性能，从而优化效用与隐私之间的权衡。
- 与在独立同分布和非独立同分布设置下的各种真实世界数据集上的最先进方法相比，实验结果验证了我们框架的有效性和优越性。

第二节回顾了联邦学习中实例级差分隐私和彩票假设的相关工作。第三节介绍联邦学习和差分隐私的背景。第四节详细介绍所提出的 Fed-LTP，第五节进行隐私分析。第六节展示了大量实验评估。第七节总结本文，并提出未来工作的方向。

2 相关工作

联邦学习中的实例级差分隐私。近年来，实例级差分隐私 [18]–[20] 已成为联邦学习领域的新兴课题。Fed-SPA [18] 将随机稀疏化与梯度扰动相结合，以实现更优的效用—隐私权衡并降低通信成本。同时，它采用加速技术来缓解收敛缓慢的问题。联邦模型蒸馏框架 FEDMD-NFDP [19] 在异构模型架构下能够提升性能，并通过共享模型预测结果消除白盒推断攻击的风险。文献 [20] 研究了本地差分隐私（LDP）的模型聚合，并提出了一种经验性解决方案，以实现将 LDP 应用于联邦学习时的严格隐私保证。

彩票假设。彩票假设 [27] 是集中式机器学习中的一种流行剪枝方法。它通过迭代剪枝生成中奖彩票（WT），使模型在相同训练轮次下能够快速收敛至接近原始模型的性能。近年来，彩票假设 [28], [29], 相关研究主要沿着两条与本文相关的方向发展：将彩票假设扩展至联邦学习，以及将差分隐私扩展至集中式机器学习。例如，LotteryFL [30]是一种个性化且通信高效的联邦学习框架，通过在非 IID 数据集上利用彩票假设实现这一目标。HeteroFL [31]可用于应对配备了计算和通信能力差异显著的移动终端。文献 [32] 使用无标签公共数据预训练模型，随后利用彩票假设压缩模型，以降低通信成本，同时不影响性能。CELL [33] 通过利用下行广播扩展 LotteryFL，从而提高通信效率。与上述研究相比，我们的工作与 PrunFL [9]更为接近。PrunFL采用自适应分布式参数剪枝，考虑了边缘设备资源有限的情况，在保持与原始模型准确率相近的同时，降低通信和计算开销，并尽量缩短整体训练时间。相比之下，彩票假设与差分隐私的结合尚未得到充分探索。DPLTM [34] 采用“高质量中奖彩票”并通过定制的评分函数进行选择，以改善隐私与效用之间的权衡。实验研究表明，DPLTM能够快速收敛，从而可以提前停止训练，减少隐私预算消耗，并将噪声影响降低至与 DPSGD [35]相当的水平。

与现有研究不同，我们的工作首次将彩票假设引入联邦学习，并利用公共数据获取初始全局模型和网络结构。通过这种方式，稀疏结构能够降低系统开销（传输和计算），并缓解注入随机噪声导致的性能下降。同时，我们提出了基于拉普拉斯机制的模型验证，以保证最终模型的性能。

3 预备知识

3.1 联邦学习

考虑一个由 U 个移动终端组成的通用联邦学习系统，其中每个客户端都拥有一个本地数据集。设 $\mathcal{D}_i^{\text{train}}$、 $\mathcal{D}_i^{\text{val}}$和 $\mathcal{D}_i^{\text{test}}$分别表示客户端 i持有的训练数据集、验证数据集和测试数据集，其中 $i \in \mathcal{U} = \{1, 2, \ldots, U\}$。形式上，该联邦学习任务可表述为：

$$\boldsymbol{\theta}^{\star} = \arg\min_{\boldsymbol{\theta}} \sum_{i \in \mathcal{U}} p_i F(\boldsymbol{\theta}, \mathcal{D}_i^{\text{train}}), \tag{1}$$

其中， $F(\cdot)$ 是损失函数， $p_i = |\mathcal{D}_i^{\text{train}}|/|\mathcal{D}^{\text{train}}| \geq 0$满足 $\sum_{i \in \mathcal{U}} p_i = 1$； $|\mathcal{D}_i^{\text{train}}|$ 分别表示训练数据集$\mathcal{D}_i^{\text{train}}$的大小以及训练数据集总大小。对于第 i个客户端，基于训练数据 $\mathcal{D}^{\text{train}}$ 学习本地模型的更新过程可表示为：

$$\boldsymbol{\theta}_i = \boldsymbol{\theta}_i^t - \alpha \nabla F(\boldsymbol{\theta}_i, \mathcal{D}_i^{\text{train}}). \tag{2}$$

表 1 主要符号汇总

$\mathcal{D}_{pub}$, $\mathcal{D}_{pr}$	公开数据集和隐私数据集
$\mathcal{M}$	用于差分隐私的随机机制
$\mathcal{D}, \mathcal{D}'$	相邻数据库
ϵ, δ	与差分隐私相关的参数
α, ρ	与 R´ényi-差分隐私相关的参数
$\mathcal{D}_i^{\text{train}}$, $\mathcal{D}_i^{\text{val}}$, $\mathcal{D}_i^{\text{test}}$	分别由第 i个用户/客户端持有的训练数据库、验证数据库和测试数据库
S_i , s_{val}	验证分数和验证函数
$\lvert\cdot\rvert$	集合的基数
$\mathcal{U}, U$	所有移动终端的集合及其总数
$\mathcal{K}, K$	所有选定的MT的集合及其总数
t, T	第 t个通信轮次的索引以及通信轮次的数量
$\boldsymbol{w}$	所有中奖彩票的模型参数
$\boldsymbol{w}_j$	第 j个中奖彩票的模型参数
$\boldsymbol{\theta}$	全局模型的模型参数
$\boldsymbol{\theta}_i^t$	第 i轮通信中第 i个用户/客户端的模型参数
$\hat{\boldsymbol{\theta}}_i^t$	第 t轮通信中进一步剪枝后的第 i个用户/客户端的模型参数
$\hat{\boldsymbol{\theta}}^t$	第 t轮通信中的全局模型
$F(\boldsymbol{\theta}_i)$	第 i个用户的全局损失函数
$\boldsymbol{M}$	剪枝模型的掩码向量
Pr	生成中奖彩票的剪枝程度/比例
M	中奖彩票数量
k	训练中奖彩票的迭代次数
P_1^{fip}	联邦迭代剪枝方案中初始选定的中奖彩票保留率 $P_1^{fip} = Pr$
P_2^{fip}	联邦迭代剪枝方案中的进一步剪枝程度
p	模型的（平均）最终保留率或压缩率
γ	联邦迭代剪枝方案中的自适应折扣因子剪枝
Δ_i^t	通信轮次 t中第 i个用户的本地模型更新

通常，损失函数 $F(\cdot)$ 由经验风险给出，并且在各移动终端上具有相同的表达式。随后， U相关联的移动终端基于训练数据$\mathcal{D}_i^{\text{train}}$学习全局模型 $\boldsymbol{\theta}$ ， $\forall i \in \mathcal{U}$。给定服务器通过聚合得到的全局模型参数 $\boldsymbol{\theta}$ ，每个客户端 i 都可以根据其验证数据集 $\mathcal{D}_i^{\text{val}}$验证模型，并获得验证分数:

$$S_i = s_{val}(\boldsymbol{\theta}, \mathcal{D}_i^{\text{val}}), \tag{3}$$

其中， s_{val} 是用于计算使用训练好的模型 $\boldsymbol{\theta}$进行预测时正确预测数量的验证函数。

3.2 差分隐私

差分隐私 [16] 是一种用于衡量隐私风险的严格隐私概念。本文考虑差分隐私定义的两种放宽版本：R´ényi差分隐私（ RDP） [36] 和零集中差分隐私（zCDP） [37]。

定义 1。 $(\epsilon,\ \delta)$-差分隐私 *[16]*。 给定隐私参数 $\epsilon > 0$和 $0 \leq \delta < 1$，若随机机制 $\mathcal{M}$ 对于任意一对相邻数据集 $\mathcal{D}$、 $\mathcal{D}'$ 以及任意输出子集$O \subseteq range(\mathcal{M})$均满足以下条件，则称其满足$(\epsilon,\ \delta)$-差分隐私:

$$\Pr[\mathcal{M}(\mathcal{D}) \in O] \leq e^{\epsilon} \Pr\left[\mathcal{M}\left(\mathcal{D}'\right) \in O\right] + \delta. \tag{4}$$

相邻数据集通过添加或删除任意一条记录构造；$(\epsilon,\ \delta)$-差分隐私是指 ϵ-差分隐私；当 $\delta = 0$时，则为纯差分隐私。

定义 2. 联邦学习的实例级差分隐私 *[18]*。 **</style>**若随机算法 满足**(**ϵ, δ)-差分隐私，则对于任意两个相邻数据集 $\mathcal{I}$、$\mathcal{I}'$ （它们通过在任意客户端的数据集中添加或删除任意一条记录构造而成），以及任意可能的输出子集 O:

$$\Pr[\mathcal{M}(\mathcal{I}) \in O] \leq e^{\epsilon} \Pr\left[\mathcal{M}\left(\mathcal{I}'\right) \in O\right] + \delta. \tag{5}$$

定义 3. ***R´ényi*** 差分隐私 *[36]*。 **</style}**给定实数 **(***1*, ∞)和隐私参数 $\rho \geq 0$，若随机机制 $\mathcal{M}$满足(α, ρ)-***RDP***，则对于任意两个仅相差一条记录的相邻数据集 $\mathcal{D}$、 $\mathcal{D}'$，***R´ényi*** α-散度在$\mathcal{M}(\mathcal{D})$ </style}与$(\mathcal{D}')$ **</style}**之间满足:

$$D_{\alpha}\left[\mathcal{M}(\mathcal{D}) \| \mathcal{M}\left(\mathcal{D}'\right)\right] := \frac{1}{\alpha - 1} \log \mathbb{E}\left[\left(\frac{\mathcal{M}(\mathcal{D})}{\mathcal{M}\left(\mathcal{D}'\right)}\right)^{\alpha}\right] \leq \rho, \tag{6}$$

其中，期望是对 $\mathcal{M}(\mathcal{D}')$的输出取值。

为定义 ρ-零集中差分隐私，我们首先引入隐私损失随机变量。对于输出范围($\mathcal{M}$)中的一个输出 $o \in$ ，机制 $\mathcal{M}$ 的隐私损失随机变量 Z 定义如下:

$$Z = \log \frac{\Pr\left[\mathcal{M}(\mathcal{D}) = o\right]}{\Pr\left[\mathcal{M}(\mathcal{D}') = o\right]}. \tag{7}$$

定义 4。 ρ-零集中差分隐私 *[37]*。 ρ-零集中差分隐私对隐私损失 Z 的矩母函数施加上界，并要求其集中在零附近。形式上，它需要满足:

$$e^{D_{\alpha}(\mathcal{M}(\mathcal{D}) \| \mathcal{M}(\mathcal{D}'))} = \mathbb{E}\left[e^{(\alpha-1)Z}\right] \leq e^{(\alpha-1)\alpha\rho}. \tag{8}$$

本文使用以下零集中差分隐私组合结果。

引理 1. 如果 $\mathcal{M}$ 满足 ϵ-差分隐私，那么 $\mathcal{M}$ 满足$\left(\frac{1}{2}\epsilon^2\right)$-零集中差分隐私 *[37]*。

4 所提出的方法

本节将详细介绍 Fed-LTP。其整体框架总结于算法 2，其客户端视角下的工作流程如图 1 所示。

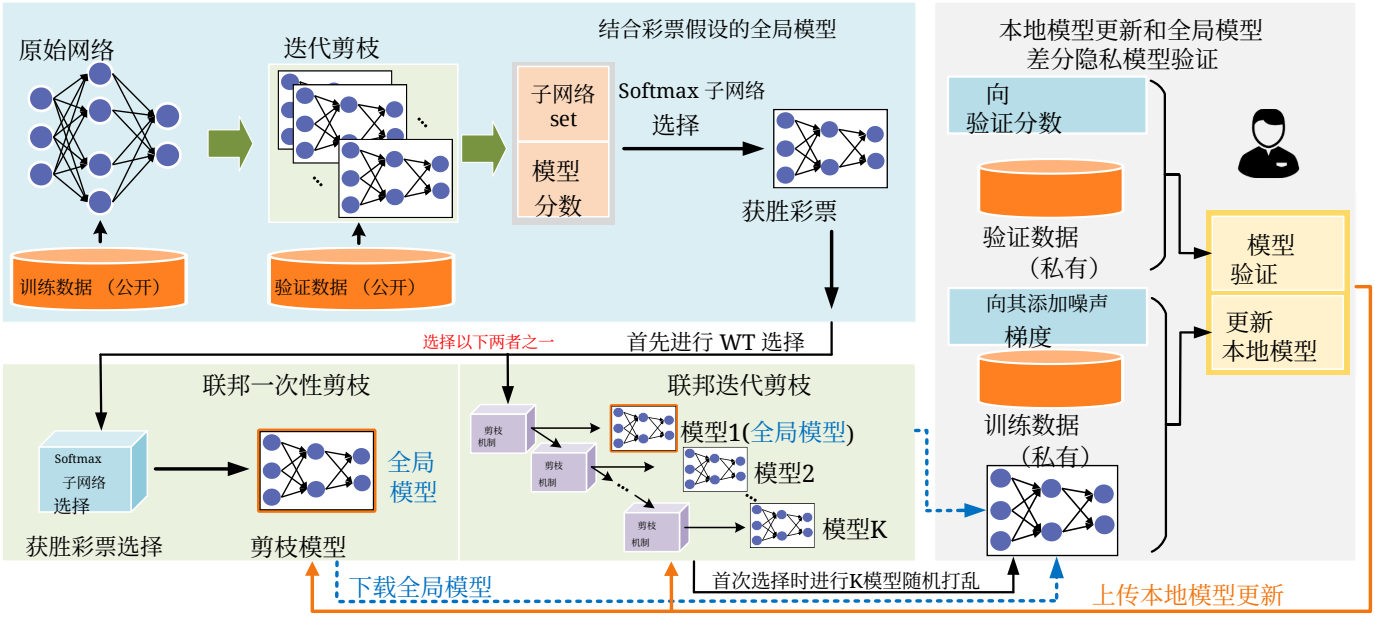


图 1。从客户端视角概述包含三个组成部分的 Fed-LTP，其中客户端是边缘计算系统中的 MT。服务器端：1）使用彩票假设对原始网络进行剪枝和训练，以生成多个 WT，并选择一个 WT 作为候选全局模型；2）借助 WT 广播机制，候选全局模型要么通过联邦一次性剪枝保持当前的剪枝程度，要么通过联邦迭代剪枝进行进一步剪枝。客户端：3）在每个客户端的私有数据上结合差分隐私训练本地模型，并采用拉普拉斯机制进行模型验证。

4.1 基于LTH的全局模型生成

受 DPLTM [34], 启发，我们采用彩票假设生成候选全局模型，其中包含两个主要步骤（算法1）。

1）使用彩票假设在公开数据集上生成 *WT*。该过程与彩票假设完全相同，唯一的区别是我们使用公开数据集生成 WT。

为了找到测试准确率更高的轻量级网络，在生成中奖彩票时使用迭代剪枝方法。在第 j 次剪枝过程中，当剪枝程度为 Pr 时，如果模型 $\boldsymbol{w}_j$ 被剪枝，则将掩码向量 $\boldsymbol{M}(\boldsymbol{w}_j)$ 设为零；如果未被剪枝，则设为一。令 $\boldsymbol{w}_{j,j}$ 表示模型 $\boldsymbol{w}_j$ 第 j 层中的权重。注意，对于第 j 层，操作定义如下：

$$\boldsymbol{M}(\boldsymbol{w}_{j,j}) = \begin{cases} 1 & if\ |w| > Pr\,|w_{max}| \\ 0 & otherwise \end{cases}, \tag{9}$$

其中，$w \in \boldsymbol{w}_{j,\ j}$ 和 $|w_{max}|$ 表示 $\boldsymbol{w}_{j,\ j}$中的最大值。因此，通过对每一层应用公式 (9)，构建模型 $\boldsymbol{w}_j$ 的掩码矩阵。

2) 使用 *softmax* 函数进行中奖彩票选择。与 DPLTM [34], 中的中奖彩票选择方法不同，这里有 M 个候选中奖彩票，每张中奖彩票都有一个得分 $V(\boldsymbol{w}_j, \mathcal{D}_{pub})$，该得分等于推理时预测正确的样本数量。为了调整准确率与网络剪枝程度之间的权衡，我们对偏好值 $V(\boldsymbol{w}_j, \mathcal{D}_{pub})$（也称为得分）使用 softmax 函数 [38] 来选择 WT，从而确保探索所有 WT:

$$P_j = \sigma\left(V\left(\boldsymbol{w}_j, \mathcal{D}_{pub}\right)\right) = \frac{e^{V\left(\boldsymbol{w}_j, \mathcal{D}_{pub}\right)}}{\sum_{j=1}^{M} e^{V\left(\boldsymbol{w}_j, \mathcal{D}_{pub}\right)}}, \tag{10}$$

其中，j 表示可能的多个中奖彩票之一，P_j 是其对应的概率。值得注意的是，我们使用彩票假设在公共数据集上生成 WT，而每个客户端都使用差分隐私在其私有数据上执行训练过程。与 DPLTM [34], 不同，我们的方法不涉及隐私预算，因为生成模型结构时不会使用私有数据。

算法1： 基于彩票假设的全局模型生成

输入： 公开数据集 $\mathcal{D}_{pub}$，获胜者数量 票数 M，训练的迭代次数 中奖彩票 k，剪枝率 Pr。
输出： 选中的中奖彩票 j。

过程 ***1***：在公开数据集上生成 ***WT***，使用 ***LTH***
对于 $j = 0$ **到** M-1，**执行**
  随机初始化一个神经网络 $f(\boldsymbol{w}_0)$。
  在 $\mathcal{D}_{pub}$ 上训练网络 k 迭代次数，以获得 $\boldsymbol{w}_j$ 第 j 个中奖彩票。
  剪除 $Pr\%$ 的参数，得到 $\boldsymbol{w}_j$，并创建一个掩码 $\boldsymbol{M}$，依据式 (9)。
  将剩余参数重置为其在 $\boldsymbol{w}_0$中的取值，从而生成中奖彩票和一个剪枝模型 $f\ (\boldsymbol{M} \odot \boldsymbol{w}_0)$。
结束
保存剪枝模型及其分数 $V(\boldsymbol{w}_j,\ \mathcal{D}_{pub})$。
过程***2***：使用 ***softmax***函数选择 ***WT***
以概率 P_j 选择一张中奖彩票 j，依据公式（10）根据分数。

4.2 服务器端 WT 广播机制

我们提出了一种服务器端 WT 广播机制，包含两个主要步骤（算法 2）。它将本地模型和

算法 2： Fed-LTP算法

输入： 通信轮次总数 T、本地更新周期 τ、每轮选定的MT数量 K、裁剪阈值 C、学习率 η 和隐私预算总量 ϵ。

输出： 最佳全局模型 $\hat{\boldsymbol{\theta}}$（通过使用拉普拉斯机制进行模型验证获得）。

服务器执行：
初始化模型流程： 首先，选择一个初始剪枝后的模型 $\boldsymbol{\theta}_0$ 通过彩票假设全局模型生成（算法1）。其次，通过**服务器端**模型剪枝获取模型（$\hat{\boldsymbol{\theta}}_0^t$，……，$\hat{\boldsymbol{\theta}}_{K-1}^t$）**模型剪枝**。
对于 $t=1$ **到** T **执行**
  均匀随机抽取 K 客户端/移动终端，不进行替换。
  将模型通过**客户端模型选择** 分配给选中的MTs.
  对于每个选中的客户端 i 并行**执行**
    $\Delta_i^t \leftarrow$ **客户端更新**$(\hat{\boldsymbol{\theta}}_i^t, \Delta^t)$
  end
  $\hat{\boldsymbol{\theta}}^{t+1} \leftarrow \hat{\boldsymbol{\theta}}^t + \frac{1}{K}\sum_{i\in\mathcal{K}} \Delta_i^t$
结束
客户端更新$(\hat{\boldsymbol{\theta}}_i^t, \Delta^t)$: $\boldsymbol{\theta}_i^{t,0} \leftarrow \hat{\boldsymbol{\theta}}_i^t$
对于 s=0 到 $\tau-1$，**执行**
  计算一个小批量随机梯度 $\boldsymbol{g}^t{}_i{}^s$
  $\boldsymbol{\theta}_i^{t,s+1} \leftarrow \boldsymbol{\theta}_i^{t,s} - \eta(\boldsymbol{g}^t{}_i{}^s \times \min(1, C/\|\boldsymbol{g}^t{}_i{}^s\|_2) + \boldsymbol{b}^t{}_i{}^s)$
  其中 $\boldsymbol{b}^t{}_i{}^s \sim \mathcal{N}(0, (\sigma^2 C^2)\cdot \boldsymbol{I}_d)$.
结束
$\Delta_i^t \leftarrow \boldsymbol{\theta}^{t,\tau} - \hat{\boldsymbol{\theta}}_i^t$
返回 Δ_i^t
服务器端模型剪枝：
如果 联邦迭代剪枝**则**
  对于 模型索引 i，从 0 到 $K-1$ **执行**
    计算 $\hat{\boldsymbol{\theta}}^t$ 根据（12）。
  end
  返回（$\hat{\boldsymbol{\theta}}_0^t$，……$\hat{\boldsymbol{\theta}}_{K-1}^t$）
结束
如果联邦一次性剪枝**则**
  生成 $\hat{\boldsymbol{\theta}}^t$ 由式（11）给出。
  返回 $\hat{\boldsymbol{\theta}}^t$
结束
客户端模型选择：
如果 联邦迭代剪枝**则**
  生成更多剪枝模型（$\hat{\boldsymbol{\theta}}_0^t$,…, $\hat{\boldsymbol{\theta}}_{K-1}^t$）。
  客户端 i首次选择时的模型随机打乱，并且不同移动终端的模型满足式（14）。
结束
如果 联邦一次性剪枝**则**
  广播一个全局模型 $\hat{\boldsymbol{\theta}}^t$ 对所有选定的MT。
结束

全局模型限制在同一模型类别内，以稳定全局模型聚合。

1) 服务器端模型剪枝。我们提出了以下两种不同的模型剪枝策略，最终保留率记为 p。

联邦一次性剪枝。遵循彩票假设[27] 和 [39], 的惯例，令 $i \in \mathcal{K} = \{1, 2, ..., K\}$ 表示选定的客户端。根据算法1，在生成中奖彩票时进行基于权重的剪枝，并可将由softmax函数选出的中奖彩票用作全局模型：

$$\hat{\boldsymbol{\theta}}^t = \boldsymbol{\theta}_0 = \boldsymbol{w}_j, \tag{11}$$

由此得到 $p = 1 - Pr$。

联邦迭代剪枝。与联邦一次性剪枝相比，为了在测试准确率、通信成本和边缘设备资源开销之间取得更好的平衡，联邦迭代剪枝基于彩票假设进一步剪枝全局模型。然后，我们生成具有不同剪枝
遵循 HeteroFL 的程度，其中本地模型具有相似的架构，但可以在同一全局模型类别内缩减模型大小。 [31],

如图2所示， p 由两个剪枝因素决定：(i). 最初选定的中奖彩票的保留率 $P_1^{fip} = Pr$；(ii). 进一步剪枝程度 P_2^{fip}。在进一步剪枝阶段，我们反复应用式(9)中的基于权重的剪枝方法，通过迭代剪枝生成异构模型，迭代次数与每次选定的移动终端数量相同。形式上，模型 i 在第 t 轮训练中的模型参数 $\hat{\boldsymbol{\theta}}_i^t$ 定义如下：

$$\begin{cases} \hat{\boldsymbol{\theta}}_i^t = \hat{\boldsymbol{\theta}}_{i-1}^t \odot \boldsymbol{M}\left(\hat{\boldsymbol{\theta}}_{i-1}^t\right) \\ \hat{\boldsymbol{\theta}}_1^t = \boldsymbol{\theta}_0^t \odot \boldsymbol{M}\left(\boldsymbol{\theta}_0^t\right) \end{cases}, \tag{12}$$

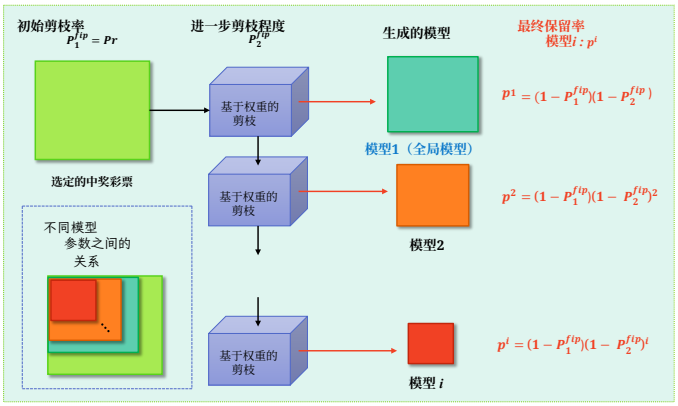


图 2. 联邦迭代剪枝方案概览。

其中， $\odot$ 表示逐元素乘积。最后，全局模型由下式生成：

$$\hat{\boldsymbol{\theta}}^t = \boldsymbol{\theta}_0^t \odot \boldsymbol{M}\left(\boldsymbol{\theta}_0^t\right), \tag{13}$$

其中，$\boldsymbol{\theta}_0^t = \boldsymbol{\theta}_0 = \boldsymbol{w}_j$ 表示全局训练轮次 $t=0$。请注意，聚合参数是对每个参与客户端中的未剪枝参数取平均得到的，且模型结构与保留率最高的客户端模型一致。

2）客户端模型选择。在联邦一次性剪枝策略中，仅将一个全局模型 $\hat{\boldsymbol{\theta}}^t$ 广播给所有移动终端；而在联邦迭代剪枝方案中，首次为客户端 i选择时，通过模型随机打乱（无放回）获得本地模型 $\hat{\boldsymbol{\theta}}_i$ 。需要注意的是，在后续训练轮次中，已被选中的移动终端不参与模型随机打乱，且首次选中的其他移动终端的模型与已被选中的移动终端的模型不同，即：

其他移动终端在首次选择时选中的模型与已被选中的移动终端的模型不同，即：

$$\boldsymbol{M}\left(\hat{\boldsymbol{\theta}}_i^t\right) \neq \boldsymbol{M}\left(\hat{\boldsymbol{\theta}}_j^t\right), \forall i, j \in \mathcal{K} = \{1, ..., K\}. \tag{14}$$

为了通过式（12）展示联邦迭代剪枝中最终保留率p 与不同客户端模型之间的更多细节，客户端 i的自适应折扣因子 γ^i 和 p^i 定义如下：

$$\gamma^i = (1 - P_2^{fip})^i, \text{and} \quad p^i = \gamma^i(1 - P_1^{fip}), \tag{15}$$

其中， $P_1^{fip} = Pr$和 P_2^{fip}也表示本地模型之间的变化程度。由于各移动终端的剪枝程度不同， p 被设为 p^i的平均值：

$$p = \frac{1}{K}\sum_{i=1}^{K} p^i \tag{16}$$

4.3 使用差分隐私进行本地模型训练

参与的移动终端从服务器端下载初始剪枝模型后，便开始使用自己的私有数据进行受差分隐私保护的本地模型训练。由于初始模型已在服务器端 WT 广播机制中经过剪枝，在第 t 轮通信中，每个参与的客户端 i 执行本地模型迭代更新。在第 s个本地迭代步骤中，根据一个小批量私有数据计算小批量随机梯度 $\boldsymbol{g}^t{}_{,i}{}^s$。然后对其进行裁剪 $\boldsymbol{g}^t{}_{,i}{}^s$ 并添加差分隐私随机噪声 $\boldsymbol{b}^t{}_{,i}{}^s$，其中噪声服从高斯分布 $\mathcal{N}(0, (\sigma^2 C^2) \cdot \boldsymbol{I}_d)$。因此，本地迭代的执行方式如下：

$$\boldsymbol{\theta}_i^{t,s+1} \leftarrow \boldsymbol{\theta}_i^{t,s} - \eta(\boldsymbol{g}_i^{t,s} \times \min(1, C/\left\|\boldsymbol{g}_i^{t,s}\right\|_2) + \boldsymbol{b}_i^{t,s}).$$

完成 τ 本地迭代步骤后，我们计算本地模型更新：

$$\Delta_i^t \leftarrow \boldsymbol{\theta}_i^{t,\tau} - \hat{\boldsymbol{\theta}}_i^t.$$

最后，每个参与的客户端 i 将本地模型更 Δ_i^t 新发送给服务器。更多细节还总结在算法 2 的第 13–19 行中。需要注意的是，在私有数据上进行本地模型训练，也意味着执行知识迁移，因为本地模型是在服务器端提供的公共数据上训练的。下面，我们将讨论公共数据与私有数据之间的知识迁移。

公共数据与私有数据之间的知识迁移。由于在差分隐私文献中使用公共数据十分常见 [39]–[42],按照 [9], [39], [42], 的惯例，我们使用带标签的公共数据和服务器的计算能力，而非边缘设备的有限资源。在实践中，我们在公共数据上生成若干个预训练 WT，然后将选定 WT 的架构保存为全局模型，并重新初始化未剪枝参数的值。请注意，与 DPLTM [34], 不同，Fed-LTP 不需要额外的隐私预算。

4.4 使用拉普拉斯机制进行模型验证

参与的移动终端的所有本地模型更新上传至服务器后，为了选择最佳全局模型并防止过拟合，服务器在每个通信轮次中，根据经过差分隐私处理的本地模型更新后得到的本地验证数据集计算验证分数。然而，传统方案通常使用网格搜索 [18], 调整超参数，由于需要观察测试隐私数据，这违反了真实系统的规则。因此，为了保护真实系统中的数据隐私，必须保护测试私有数据，这也会消耗隐私预算。本文采用拉普拉斯机制，在验证过程中实现差分隐私保证。

在每个通信轮次开始时，服务器通过聚合获得全局模型 $\hat{\boldsymbol{\theta}}^t$。然后，它将该全局模型发送给所有移动终端。随后，每个客户端根据其本地验证数据集对接收到的模型进行验证，以获得验证分数 S_i^t，并将其发送给服务器。为保护本地验证数据集的隐私，每个客户端需要按以下方式扰动验证分数：

$$\widetilde{S}_i^t = S_i^t + \text{Lap}\left(\Delta_1(s_{val})\lambda_{val}\right), \tag{17}$$

其中， $\Delta_1(s_{val})$ 是差分隐私敏感度， λ_{val} 是拉普拉斯分布的参数。显然，单个样本导致的分数最大变化被限定为 $\Delta_1(s_{val}) = 1$。服务器获得全局模型的验证分数如下：

$$S^t = \sum_{i\in=1}^{K} \widetilde{S}_i^t = \sum_{i=1}^{K} S_i^t + \sum_{i=1}^{K} \text{Lap}(\Delta_1(s_{val})\lambda_{val}), \tag{18}$$

在收到所有移动终端的分数后。最后，当联邦学习训练终止时，服务器选择验证分数最高的模型作为全局模型，即：

$$\hat{\boldsymbol{\theta}}^f = \hat{\boldsymbol{\theta}}^{\arg\max_{t\in[T]} S^t}. \tag{19}$$

根据（18），随着每轮参与者数量增加，聚合噪声的方差会减小，系统因此能够获得更可靠的验证分数。

5 隐私分析

在本节中，我们基于零集中差分隐私进行紧致的隐私分析，以计算随着通信轮次 T 增加而产生的隐私损失/预算 ϵ。

如图 1 和算法 2 所示，在给定 δ 的情况下，隐私预算 ϵ 可分为两部分：差分隐私本地模型训练和基于差分隐私的模型验证。

定理1. 所提出算法在第 t个通信轮次后的累积隐私损失可表示为

$$\epsilon^t = \rho_s^t + \frac{(t+1)\alpha(\alpha-1)}{2\lambda_{val}^2} + \frac{\log\left(\frac{1}{\delta}\right) - \log(\alpha)}{\alpha - 1} + \log\left(1 - \frac{1}{\alpha}\right), \tag{20}$$

其中

$$\rho_s^t = \frac{t+1}{(\alpha-1)} \log \mathbb{E}_{z\sim\mu_0(z)}\left[\left(1 - \tilde{q} + \tilde{q}\frac{\mu_1(z)}{\mu_0(z)}\right)^{\alpha}\right]. \tag{21}$$

在 (21) 中， $\mu_0(z) = \mathcal{N}(0, \sigma^2)$ 表示高斯概率密度函数（*PDF*）， $\mu_1(z) = \tilde{q}\mathcal{N}(1, \sigma^2) + (1 - \tilde{q})\mathcal{N}(0, \sigma^2)$

是两个高斯分布混合后的概率密度函数，$\tilde{q}$ 是本地模型训练中本地梯度（小批量）的采样率。

证明。 根据文献 [36] 和 [43],，隐私损失可表示为:

$$\epsilon^t = \rho^t + \frac{\log\left(\frac{1}{\delta}\right) - \log(\alpha)}{\alpha - 1} + \log\left(1 - \frac{1}{\alpha}\right), \tag{22}$$

其中，$\rho^t = \rho_s^t + \rho_v^t$ 是 Rényi α-散度，ρ_s^t 和 ρ_v^t 分别由差分隐私本地模型训练和基于差分隐私的模型验证引起。具体而言，根据 [35] 和 [36],，我们可以按式 (21) 计算 ρ_s^t，其中使用 Rényi 距离来估计隐私损失。我们用 $\mathcal{M}_{val}$ 表示模型验证中使用的随机机制。根据验证过程的**引理 1**，如果 $\mathcal{M}_{val}$ 满足 ϵ_{val}-DP，则也满足 $\frac{1}{2}\epsilon_{val}^2$-零集中差分隐私。因此，有:

$$D_\alpha(\mathcal{M}_{val}(\mathcal{D}) \| \mathcal{M}_{val}(\mathcal{D}')) \leq \frac{1}{2}\alpha\epsilon_{val}^2(\alpha - 1) = \frac{\alpha(\alpha-1)}{2\lambda_{val}^2}. \tag{23}$$

然后，考虑 (t+1) 个通信轮次，可得:

$$\rho_v^t \leq \frac{(t+1)\alpha(\alpha-1)}{2\lambda_{val}^2}. \tag{24}$$

根据 (22) 和 (23)，隐私损失按 (20) 累积。

□

在训练过程中，每个 epoch 都会更新累积隐私损失；一旦累积隐私损失超过固定的隐私预算 ϵ，训练便会终止。为了在给定总隐私预算的情况下实现预期的训练时间，我们可以在训练前确定这些调度方案的超参数值。请注意，对于基于验证数据的调度方案，由于需要访问验证数据集，因此还需将额外的隐私成本考虑在内。

6 实验评估

本节旨在评估 Fed-LTP 在常用基准数据集上采用不同最终保留率时的性能，并将其与其他基线方法进行比较，以展示我们框架的优越性。

6.1 实验设置

基线方法。 为了评估 Fed-LTP 的性能，我们将其与几种基线方法进行比较：1) DP-Fed：该基线在联邦平均算法 [2] 中加入实例级差分隐私；2) Fed-SPA [18]：该基线结合随机稀疏化与梯度扰动，并采用加速技术来提高收敛速度。

数据集与数据划分。我们在四个数据集上评估 Fed-LTP：MNIST [44], FEMNIST [45], CIFAR-10 [46],和 Fashion-MNIST [47], ，其中 FEMNIST 和 CIFAR-10 被视为公共数据。我们设置了两组实验，分别采用不同的私有数据：Fashion-MNIST 和 MNIST。请注意，本文介绍了采用 **Fashion-MNIST 私有数据**的实现细节和结果。同时，关于采用 MNIST 私有数据的详细实验评估见**附录**D。我们为所有算法考虑两种不同设置:

表 2 Fashion-MNIST 私有数据上，不同最终保留率 p 下，Fed-LTP 在移动终端 R_{loc} 上存储的模型压缩率，包括联邦迭代剪枝和基线方法。

方法	公共数据	p （平均）	R_{loc} （选定的MT之间）				
Fed-LTP	FEMNIST	0.20	0.44	0.27	0.16	0.10	0.06
		0.29	0.52	0.36	0.25	0.18	0.12
		0.40	0.59	0.47	0.38	0.30	0.24
		0.54	0.66	0.60	0.54	0.48	0.44
	CIFAR-10	0.28	0.60	0.36	0.22	0.13	0.08
		0.39	0.70	0.49	0.34	0.24	0.17
		0.53	0.79	0.64	0.51	0.41	0.33
		0.73	0.89	0.80	0.72	0.65	0.59
Fed-SPA	–	All	1.00	1.00	1.00	1.00	1.00
DP-Fed	–	1.00	1.00	1.00	1.00	1.00	1.00

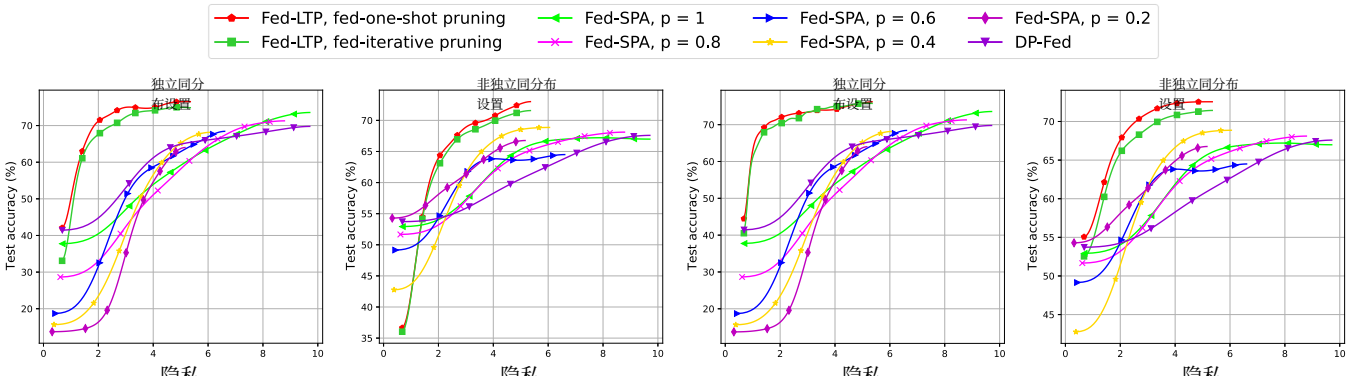
联邦客户端（例如 MT）之间的数据分布既可以是相同的（IID），也可以是不同的（非 IID）；其中，非 IID 指本地 MT 的数据分布具有异质性，这会增加全局模型的训练难度。我们根据狄利克雷分布**Dir(α)** 为每个客户端划分训练数据 [48] ，并按照相同的分布为每个客户端生成相应的验证数据和测试数据，其中 $\alpha > 0$ 是控制 MT 之间分布均匀程度的浓度参数。对于非独立同分布设置，我们设定 $\alpha = 1$，此时 MT 可以拥有数量不同、类别随机选取的样本。

实现细节。 所有实验中，移动终端的数量为 $U = 50$，服务器随机选择一组移动终端，其采样率为移动终端的 $q = 0.1$ ，参与训练。所有实验中，我们设置隐私失效概率$\delta = 10^{-3}$ 和本地迭代次数 $\tau = 300$，并采用固定的裁剪阈值 $C = 10$ 和噪声乘数$\sigma = 1.4$ 。对于移动终端上的本地优化器，我们使用动量随机梯度下降，并将本地动量系数设为 0.5；同时，将联邦学习过程中的学习率 η 设为 0.01 ，衰减率设为 0.99 。与此同时，生成中奖彩票的学习率设为 1.2×10^{-3}。在所有实验的联邦迭代剪枝方案中，P_2^{fip} 设为 0.1 。对于 Fashion-MNIST 私有数据，采用卷积神经网络模型，通信轮次$T = 100$ 和批量大小 $B = 15$。更多实现细节见**附录** B。此外，所有算法的隐私损失均使用[43]中提供的 API 计算。基线方法的通信成本按照与 Fed-SPA [18]相同的方式计算。Fed-LTP 中的每个客户端也使用 $p \times d \times 32 \times T \times q$ 位，其中 d 是需要更新至服务器的未剪枝模型参数数量。

6.2 实验结果

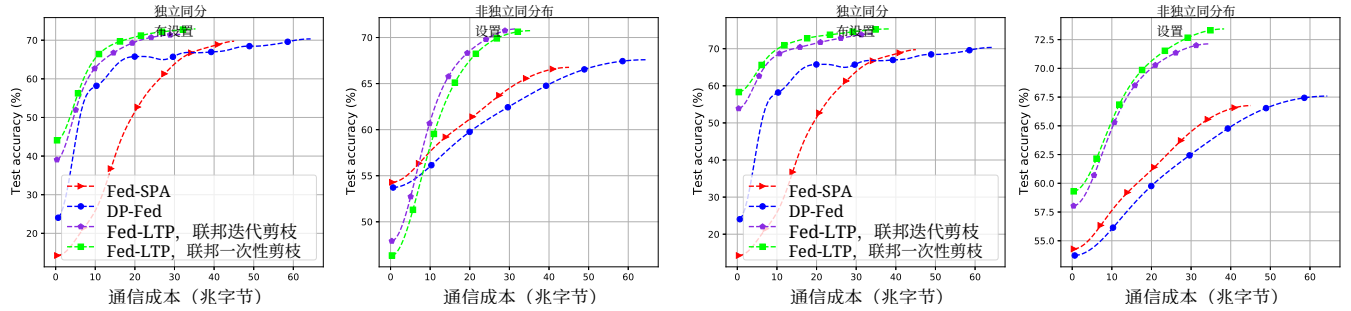
每项实验运行 3 次，并报告基于验证数据集的测试准确率，以及每项实验中所有轮次的上行和下行通信成本累计总和。接下来，我们将从多个方面评估 Fed-LTP。

1) MT上的高效计算和内存占用。假设模型参数以32位浮点数表示，存储在MT上的模型压缩率为



(a) FEMNIST (b) CIFAR-10

图3. Th 在独立同分布和非独立同分布设置下，两种数据集上不同算法的效用-隐私权衡 .



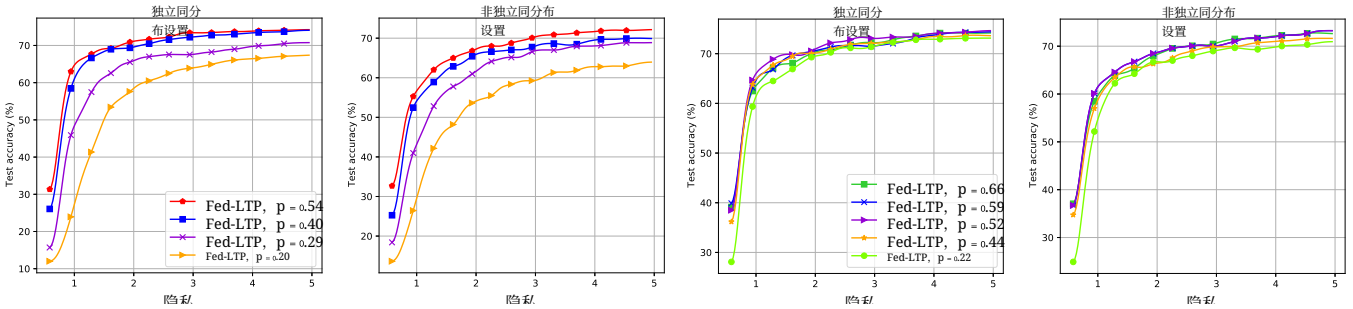
(a) FEMNIST (b) CIFAR-10

图 4. 在独立同分布和非独立同分布设置下，两种数据集上不同算法的测试准确率与累计通信成本（兆字节）。

R_{loc}，这也是决定MT上计算开销和内存占用的关键参数。由于MT所需的计算和内存开销与训练模型的大小成正比，因此缩小模型大小可以缓解MT的资源限制。具体而言，Fed-LTP中的本地模型采用稀疏到稀疏技术进行训练，而基线方法Fed-SPA和DP-Fed分别采用稠密到稀疏和稠密到稠密训练。实际上，Fed-SPA仅降低上传通信成本，并未缩小模型大小；而DP-Fed训练网络时不进行任何模型压缩或剪枝。因此，对于所有基线方法，$p \leq R_{loc} = 1.00$；而对于Fed-LTP，$p = R_{loc} < 1.00$。此外，考虑到MT之间的资源异构性，我们提出了联邦迭代剪枝策略。如表 2 所示，选定的MT之间的 R_{loc} 各不相同，且与基线方法相比更小；而对于联邦一次性剪枝策略，R_{loc} 的值与 p的值相同。因此，在选择相同的WT作为候选全局模型时，联邦迭代剪枝的资源开销低于联邦一次性剪枝。

2) 更优的效用-隐私权衡。 我们比较了所有算法在相同隐私预算下的最佳测试准确率，并将其称为效用-隐私权衡。如图 3 所示，在独立同分布和非独立同分布设置下，对于两个数据集 FEMNIST 和 CIFAR-10，采用两种不同剪枝策略的 Fed-LTP 均比基线方法实现了更优的效用-隐私权衡。具体而言，在隐私损失相同的情况下，Fed-LTP 的测试准确率高于基线方法。例如，在图 3(a) 中，当 $\epsilon = 4.0$ 时，Fed-LTP 将测试准确率提高了约 21% 和9% ，相比于在 FEMNIST 的独立同分布和非独立同分布设置下的基线方法，分别取得了更好的结果。同时，Fed-LTP 的模型训练收敛速度高于基线方法。因此，Fed-LTP 实现了更好的效用-隐私权衡，这意味着模型性能更佳，隐私保证更严格。

3）高效通信。 对于每种算法及其最优保留率，图 4 展示了其测试准确率与上行和下行通信成本累计总和之间的关系。在 FEMNIST 和 CIFAR-10 上，算法设置如下：Fed-LTP（联邦迭代剪枝，分别为 $p = 0.40$ 和 $p = 0.39$；联邦一次性剪枝，分别为 $p = 0.40$ 和 $p = 0.30$）、Fed-SPA（$p = 0.6$）以及 DP-Fed（$p = 1$）。显然，对于不同的设置和数据集，Fed-LTP 的累计总通信成本低于基线方法，同时具有更快的收敛速度和更高的测试准确率。如图 4(a) 和图 4(b) 所示，在通信成本相同的情况下，由于采用了彩票假设并在服务器端进一步对全局模型进行剪枝，采用两种剪枝策略的 Fed-LTP 始终比基线方法达到更高的准确率。因此，显然，采用两种不同剪枝策略的 Fed-LTP 比基线方法具有更高的通信效率，并且联邦一次性剪枝策略比联邦迭代剪枝策略具有更高的通信效率。例如，要在图 4(a) 的 FEMNIST 上达到 70% 的目标准确率，在独立同分布设置下，这两种剪枝策略产生的通信成本分别约为 22 兆字节和 13 兆字节；在非独立同分布设置下，产生的通信成本分别约为 25 兆字节和 18 兆字节。这一观察结果表明，客户端异质性和下行成本分别是影响模型性能和



(a) 联邦迭代剪枝 (b) 联邦一次性剪枝

图 5. 剪枝的影响：在独立同分布和非独立同分布设置下，使用联邦迭代剪枝和联邦一次性剪枝策略时，全局模型的测试准确率随不同保留率 p 下累积隐私损失 ϵ 的变化。

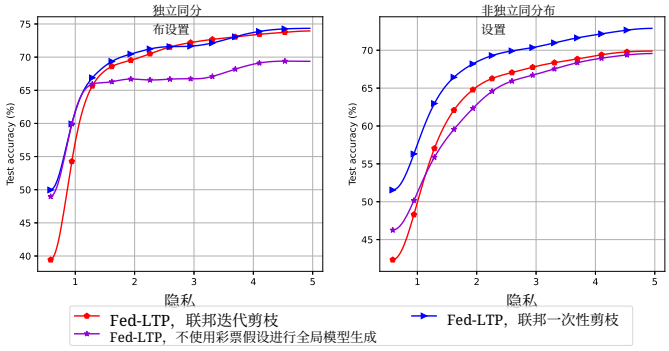


图6. 采用 LTH 的全局模型生成效果：在不同场景下，采用或不采用 LTH 模块进行全局模型生成时，全局模型的测试准确率及累计隐私损失 ϵ。所有场景均采用联邦迭代剪枝和联邦一次性剪枝策略，在独立同分布和非独立同分布设置下进行训练。

通信效率的两个重要因素。

6.3 剪枝方案和保留率的讨论

本节研究了在 FEMNIST 的独立同分布和非独立同分布数据上，采用不同保留率和剪枝方案时剪枝对 Fed-LTP 的影响。

剪枝的影响（最终保留率 p）。图 5 显示，随着保留率 p 降低，测试准确率也呈下降趋势。可以从参数共享的角度分析这一现象：当保留率 p 较低时，服务器接收到的模型信息损失显而易见，从而导致训练过程中出现显著误差。例如，在图 5(a) 中，在独立同分布设置下采用联邦迭代剪枝策略时，保留率 p 设为 0.54、0.40、0.29和 0.20 后，测试准确率分别降至约72%、70%、66% 和58%。此外，与联邦一次性剪枝相比，采用联邦迭代剪枝时测试准确率的下降尤为明显（例如，独立同分布和非独立同分布设置下分别为 $p = 0.20$ 和 $p = 0.22$ ）。

剪枝方案的影响。从图 3 和图 5 可以看出，联邦迭代剪枝与联邦一次性剪枝策略的性能差异。显然，联邦迭代剪枝在实现与联邦一次性剪枝相近性能的同时，降低了移动终端的资源开销和通信成本。这意味着，通过降低本地模型的复杂度，可以更好地平衡性能、计算开销和通信成本。然而，尤其是在图 5(a) 所示的非独立同分布设置下，随着（平均）最终保留率 p 降低，由于模型异质性和数据异质性的双重影响，其性能不如联邦一次性剪枝。我们目前针对数据异质性影响开展的实验研究，以及对移动终端间客户端异质性影响的研究，都是未来值得探讨的方向。

6.4 消融研究

本节展示 Fed-LTP 的消融研究。其目的是通过将其他组件或超参数固定为默认值，考察 Fed-LTP 中某个特定组件或超参数的作用和影响。消融实验在 FEMNIST 上进行。

采用 LTH 进行全局模型生成的效果。LTH 用于生成全局模型的统一稀疏结构，同时确保服务器端的模型性能更佳。为验证采用 LTH 进行全局模型生成的效果，我们对采用两种剪枝策略的 Fed-LTP 进行实验，并以不进行全局模型生成的 Fed-LTP 作为对照，此时使用未剪枝网络。与集中式学习场景中的 LTH 情况类似 [27]–[29], ，图6显示，包含 LTH 模块的全局模型能够提升两种剪枝策略下的性能：采用联邦迭代剪枝的 Fed-LTP 在独立同分布和非独立同分布设置下的测试准确率分别提高约5% 和1%。因此，采用联邦一次性剪枝时，性能分别提升约5% 和4%。

服务器端 WT 广播机制的效果。为验证服务器端 WT 广播机制的效果，我们针对三种方案开展了实验：1）客户端 WT选择，即移动终端在客户端使用本地私有数据生成并保存 WT；2）客户端迁移 WT 选择，即移动终端只需选择并训练由服务器端创建的预训练模型；3）服务器端 WT 广播机制。图8中的结果表明，与服务器端 WT 广播机制相比，客户端WT选择和客户端迁移 WT 选择方案的性能较差且波动更大，这是因为不同移动终端生成的 WT 结构差异较大，并且移动终端之间的数据分布有偏。

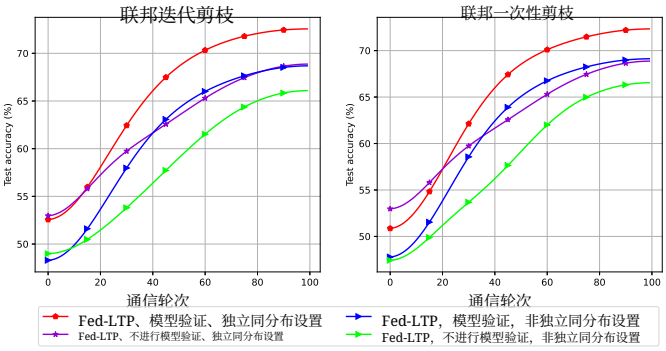


图7. 拉普拉斯机制下模型验证的影响：在不同场景下，使用或不使用拉普拉斯机制模型验证模块时，全局模型在累积隐私损失ϵ下的测试准确率。该模块分别在独立同分布和非独立同分布设置下，采用联邦迭代剪枝和联邦单次剪枝进行训练。

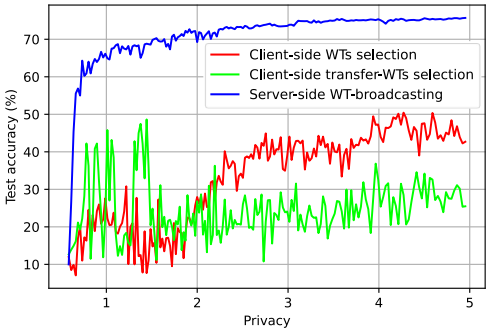


图8. 服务器端 WT 广播机制的影响：独立同分布设置下不同方案中全局模型的测试准确率与累计隐私损失 ϵ 。

采用拉普拉斯机制进行模型验证的效果。如果服务器端不使用全局模型验证，那么最终训练过程中的全局模型可视为最终模型。因此，训练模型的性能通常会随着轮数增加而下降，在差分隐私设置下尤其如此。这可能导致最终模型的性能不如使用全局模型验证时的性能。如图7所示，在两种设置和不同剪枝方案下，采用模型验证的 Fed-LTP 测试准确率都更高。

隐私损失计算方法（零集中差分隐私）的影响。如第 V 节所述的隐私分析，采用零集中差分隐私计算的隐私损失可以在通信轮次中提高隐私保护水平。如图 9 所示，与未采用零集中差分隐私隐私损失的 Fed-LTP 相比，两种设置下全局模型的收敛速度和最佳准确率总体上都较差。因此，在两种设置及不同剪枝方案下，零集中差分隐私均有助于模型效用和隐私保证。

7 结论

本文设计了一种联邦学习（Fed-LTP）中的隐私保护算法，可在提升模型效用的同时，妥善平衡计算、边缘设备的内存效率和通信效率。该算法包含一个用于探索稀疏网络结构的预训练模型，以及一种差分隐私全局模型验证机制。

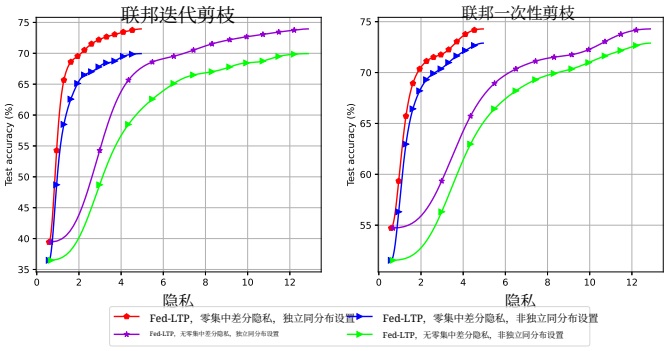


图 9. 隐私损失计算方法（零集中差分隐私）的影响：在有无零集中差分隐私模块的不同场景下，全局模型在累计隐私损失 ϵ 时的测试准确率。该模型分别在独立同分布和非独立同分布设置下，采用联邦迭代剪枝和联邦一次性剪枝策略进行训练。

为确保所选模型的质量并避免过拟合。同时，我们结合模型训练和验证的隐私成本，提出了隐私分析，并采用稀疏到稀疏训练，以节省边缘设备有限的资源。此外，所提出的加噪方法和轻量级模型能够更好地平衡隐私预算与模型性能。最后，我们进行了大量实验，以验证所提算法相较于最先进方法的有效性和优越性。未来，我们将进一步研究剪枝方法的迭代特性对采用非独立同分布数据集的客户端/移动终端的影响。此外，在非独立同分布数据集上验证泛化保证的能力也有待进一步探索。

参考文献

[1] T. Li、A. K. Sahu、A. Talwalkar 和 V. Smith，《联邦学习：挑战、方法与未来方向》，*IEEE Signal Processing Magazine*，第 50–60 页，2020 年。[2] B. McMahan、E. Moore、D. Ramage、S. Hampson 和 B. A. y Arcas，《基于去中心化数据的深度网络通信高效学习》，载于人工智能与统计国际会议（*AISTATS*）论文集，第 54 卷，2017 年 4 月，第 1273–1282 页。[3] P. Kairouz、H. B. McMahan、B. Avent、A. Bellet、M. Bennis、A. N. Bhagoji、K. Bonawitz、Z. Charles、G. Cormode、R. Cummings等，《联邦学习的进展与开放问题》，机器学习基础与趋势®，第 1–210 页，2021 年。[4] F. Sattler、S. Wiedemann、K.-R. M¨uller 和 W. Samek，《基于非独立同分布数据的鲁棒且通信高效的联邦学习》，*IEEE*神经网络与学习系统汇刊，2019 年。[5] J. Hamer、M. Mohri 和 A. T. Suresh，《Fedboost：一种通信高效的联邦学习算法》，载于国际机器学习会议。PMLR，2020 年。[6] R. Dai、L. Shen、F. He、X. Tian 和 D. Tao，《Dispfl：通过去中心化稀疏训练实现通信高效的个性化联邦学习》，*arXiv*预印本 *arXiv:2206.00187*，2022 年。[7] Z. Xu、Z. Yang、J. Xiong、J. Yang 和 X. Chen，《Elfish：面向异构边缘设备的资源感知联邦学习》，*CoRR*，2019 年。[8] S. Wang、T. Tuor、T. Salonidis、K. K. Leung、C. Makaya、T. He 和 K. Chan，《资源受限边缘计算系统中的自适应联邦学习》，*IEEE*通信领域精选期刊，2019 年。[9] Y. Jiang、S. Wang、V. Valls、B. J. Ko、W.-H. Lee、K. K. Leung 和 L. Tassiulas，《模型剪枝使边缘设备上的联邦学习更加高效》，*IEEE*神经网络与学习系统汇刊，2022 年提前访问。

[10] T. Huang、S. Liu、L. Shen、F. He、W. Lin 和 D. Tao，“通过稀疏本地模型实现个性化联邦学习”，arXiv预印本 arXiv:2201.11380，2022。[11] M. Fredrikson、S. Jha 和 T. Ristenpart，“利用置信度信息的模型反演攻击及基本对策”，载于 ACM SIGSAC计算机与通信安全会议（CCS）论文集，2015 年，第 1322–1333 页。[12] R. Shokri、M. Stronati、C. Song 和 V. Shmatikov，“针对机器学习模型的成员推断攻击”，载于 IEEE安全与隐私研讨会（SP）论文集，2017 年，第 3–18 页。[13] L. Melis 、 C. Song 、 E. De Cristofaro 和 V. Shmatikov，“ 利用协同学习中意外的特征泄露”，载于 IEEE安全与隐私研讨会（SP）论文集，2019 年，第 691–706 页。[14] M. Nasr、R. Shokri 和 A. Houmansadr，“深度学习的全面隐私分析：针对集中式和联邦学习的被动与主动白盒推断攻击”，载于 IEEE安全与隐私研讨会（SP）论文集，2019 年，第 739–753 页。[15] L. Zhang、L. Shen、L. Ding、D. Tao 和 L.-Y. Duan，“通过无数据知识蒸馏微调全局模型，实现非独立同分布联邦学习”，载于 IEEE/CVF计算机视觉与模式识别会议论文集，2022 年，第 10 174–10 183 页。[16] C. Dwork 和 A. Roth，“差分隐私的算法基础”，《理论计算机科学基础与趋势》，第 211–407 页，2014 年 8 月。[17] N. Agarwal、A. T. Suresh、F. X. Yu、S. Kumar 和 B. McMahan，“cpSGD：通信高效且满足差分隐私的分布式随机梯度下降”，载于神经信息处理系统年会（NeurIPS）论文集，2018 年 12 月，第 7575–7586 页。[18] R. Hu、Y. Gong 和 Y. Guo，“采用稀疏化增强隐私和自适应优化的联邦学习”，载于第三十届国际人工智能联合会议（IJCAI）论文集，2021 年 8 月，第 1463–1469 页。[19] L. Sun 和 L. Lyu，“采用无噪声差分隐私的联邦模型蒸馏”，载于第三十届国际人工智能联合会议（IJCAI）论文集，2021 年 8 月，第 1563–1570 页。[20] L. Sun、J. Qian 和 X. Chen，“LDP-FL：采用本地差分隐私的实用联邦学习私有聚合”，载于第三十届国际人工智能联合会议（IJCAI）论文集，2021 年 8 月，第 1571–1578 页。[21] H. B. McMahan、D. Ramage、K. Talwar 和 L. Zhang，“学习满足差分隐私的循环语言模型”，载于国际学习表征会议（ICLR）论文集，2018 年 4 月。[22] R. C. Geyer、T. Klein 和 M. Nabi，“差分隐私联邦学习：客户端层面的视角”，CoRR，2017 年 8 月。[23] P. Kairouz、Z. Liu 和 T. Steinke，“用于安全聚合联邦学习的分布式离散高斯机制”，载于国际机器学习会议（ICML）论文集，2021 年 7 月，第 5201–5212 页。[24] R. Hu、Y. Gong 和 Y. Guo，“采用稀疏化模型扰动的联邦学习：在客户端级差分隐私下提高准确率”，CoRR，2022 年。[25] A. Cheng、P. Wang、X. S. Zhang 和 J. Cheng，“采用局部正则化和稀疏化的差分隐私联邦学习”，CoRR，2022 年。[26] Y. Shi、Y. Liu、K. Wei、L. Shen、X. Wang 和 D. Tao，“使差分隐私联邦学习的损失景观更加平坦”，arXiv预印本 arXiv:2303.11242，2023。[27] J. Frankle 和 M. Carbin，“彩票假设：寻找稀疏可训练神经网络”，载于国际学习表征会议（ICLR）论文集，2019 年。[28] J. Frankle、G. K. Dziugaite、D. M. Roy 和 M. Carbin，“稳定彩票假设”，arXiv预印本 arXiv:1903 01611，2019 年。. .[29] J. Frankle、G. K. Dziugaite、D. Roy 和 M. Carbin，“线性模式连通性与彩票假设”，载于国际机器学习会议（ICML）论文集，2020 年，第 3259–3269 页。[30] A. Li、J. Sun、B. Wang、L. Duan、S. Li、Y. Chen 和 H. Li，“LotteryFL：基于彩票假设、面向非 IID 数据集的个性化通信高效联邦学习”，载于 IEEE/ACM边缘计算研讨会（SEC）论文集，美国加利福尼亚州圣何塞，2021 年 12 月，第 68–79 页。[31] E. Diao、J. Ding 和 V. Tarokh，“HeteroFL：面向异构客户端的计算与通信高效联邦学习”，arXiv预印本 arXiv:2010.01264，2020 年。[32] S. Itahara、T. Nishio、M. Morikura 和 K. Yamamoto，“基于彩票假设的模型压缩无监督预训练

在联邦学习中，”载于*IEEE*车辆技术会议论文集（*VTC2020-Fall*），2020年，第1–5页。[33] S. Seo、S.-W. Ko、J. Park、S.-L. Kim和M. Bennis，“通信高效且个性化的联邦彩票票据学习”，载于*IEEE*无线通信信号处理进展国际研讨会论文集（*SPAWC*），2021年，第581–585页。[34] L. Gondara、K. Wang和R. S. Carvalho，“差分隐私彩票票据机制”，*arXiv*预印本 *arXiv:2002 11613*，2020年。。[35] M. Abadi、A. Chu、I. J. Goodfellow、H. B. McMahan、I. Mironov、K. Talwar和L. Zhang，“基于差分隐私的深度学习”，*CoRR*，2016年。[36] I. Mironov，“Rényi差分隐私”，载于*IEEE*计算机安全基础研讨会论文集（*CSF*），2017年，第263–275页。[37] M. Bun和T. Steinke，“集中式差分隐私：简化、扩展与下界”，载于密码学理论国际会议论文集（*TCC*），计算机科学讲义丛书，第9985卷，中国北京，2016年10月至11月，第635–658页。[38] E. Jang、S. Gu和B. Poole，“基于Gumbel-Softmax的类别重参数化”，*arXiv*预印本 *arXiv:1611.01144*，2016年。[39] Z. Luo、D. J. Wu、E. Adeli和L. Fei-Fei，“通过稀疏网络微调实现可扩展的差分隐私”，载于*IEEE/CVF*计算机视觉与模式识别会议论文集（*CVPR*），2021年，第5057–5066页。[40] J. Wang和Z.-H. Zhou，“利用少量公共数据进行差分隐私学习”，载于*AAAI*人工智能会议论文集（*AAAI*），2020年。[41] N. Papernot、S. Song、I. Mironov、A. Raghunathan、K. Talwar和Ú. Erlingsson，“基于PATE的可扩展隐私学习”，*arXiv*预印本 *arXiv:1802.08908*，2018年。[42] T. Li、M. Zaheer、S. Reddi和V. Smith，“利用辅助信息进行私有自适应优化”，载于国际机器学习会议。PMLR，2022年，第13 086–13 105页。[43] A. Yousefpour、I. Shilov、A. Sablayrolles、D. Testuggine、K. Prasad、M. Malek、J. Nguyen、S. Gosh、A. Bharadwaj、J. Zhao、G. Cormode和I. Mironov，“Opacus：PyTorch中用户友好的差分隐私库”，*CoRR*，2021年。[44] Y. LeCun、L. Bottou、Y. Bengio和P. Haffner，“应用于文档识别的基于梯度的学习”，*IEEE*论文集，第86卷，第11期，第2278–2324页，1998年。[45] S. Caldas、S. M. K. Duddu、P. Wu、T. Li、J. Konečnỳ、H. B. McMahan、V. Smith和A. Talwalkar，“Leaf：联邦学习场景的基准测试”，*arXiv*预印本 *arXiv:1812.01097*，2018年。[46] A. Krizhevsky、G. Hinton 等，“从微小图像中学习多层特征”，2009年。[47] H. Xiao、K. Rasul和R. Vollgraf，“Fashion-MNIST：用于机器学习算法基准测试的新型图像数据集”，*arXiv*预印本 *arXiv:1708.07747*，2017年。[48] T.-M. H. Hsu、H. Qi和M. Brown，“衡量非独立同分布数据分布对联邦视觉分类的影响”，*arXiv*预印本 *arXiv:1909.06335*，2019年。

附录

.1 符号和高参数

在表 1 中，我们介绍了本文所用的概念和主要参数。

.2 实现细节

模型。 对于 MNIST 私有数据集，CNN 模型由两个带有 ReLU 激活函数的 5×5 卷积层组成（第一层包含 10 个滤波器，第二层包含 20 个滤波器，每层之后均接有 2×2 最大池化层）、一个包含 320 个单元并使用 ReLU 激活函数的全连接层，以及最后的 Softmax 输出层，该模型称为模型*1*。对于 Fashion-MNIST 私有数据集，采用一个 CNN 模型，该模型除卷积层大小为 3×3 且使用 ReLU 激活函数（第一层包含 32 个滤波器，第二层包含 64 个滤波器），以及全连接层包含 512 个单元外，与模型*1* 完全相同，该模型称为模型*2*。此外，模型*1*的通信轮次 $T = 50$和批量大小 $B = 10$，以及模型*2*的 $T = 100$ 和 $B = 15$。

数据集。MNIST、FEMNIST 和 Fashion-MNIST 数据集各包含 60K 个训练样本和 10K 个测试样本。具体而言，MNIST 和 Fashion-MNIST 数据集分别由 10 类 28×28 手写数字图像和灰度图像组成。FEMNIST 数据集是根据数字/字符的书写者对扩展 MNIST 中的数据进行划分而构建的，包含 62 个类别。CIFAR-10 数据集由 10 类 32×32 图像组成。该数据集包含 50K 个训练样本和 10K 个测试样本。

.3 Fashion-MNIST 私有数据实验结果详情

我们比较了所有算法在相同隐私预算下所能达到的最佳测试准确率，即效用-隐私权衡，如图 3 所示。此外，我们在表 3 中列出了最终结果，总结了 DP-Fed、Fed-SPA 和 Fed-LTP 在 Fashion-MNIST 私有数据集的独立同分布和非独立同分布设置下经过 T 轮后的最终结果。基线方法的开销按照与 Fed-SPA [18] 相同的方式计算。Fed-LTP 中每个客户端使用 $p \times d \times 32 \times T \times q$ 位，其中 d 是需要更新到服务器的未剪枝模型参数数量。

显然，存在三个主要评价指标，而我们的算法在这三个指标之间实现了良好的平衡。为了便于讨论和比较分析，以下各节将在固定某一指标的情况下，分别对其他指标进行讨论和分析。我们可以看到，由于联邦客户端之间存在数据异质性，非独立同分布设置下的性能通常不如独立同分布设置。此外，在两种剪枝策略下，与采用不同最终保留率的两种基线方法 DP-Fed 和 Fed-SPA 相比，Fed-LTP 在准确率与通信成本之间也实现了更好的权衡。具体而言，当 p 较大时，Fed-SPA具有更高的测试准确率，但同时也会产生更高的通信成本和隐私成本。与此同时，在测试准确率和通信成本方面，DP-Fed的整体表现通常不如Fed-LTP。

.4 MNIST 私有数据实验结果详述

结果与 Fashion-MNIST 私有数据集上的结果相同。具体而言，我们从多个方面列出了评估结果。

.4.1 边缘设备的高效计算与内存占用。

边缘设备所需的计算和内存占用开销与训练模型的大小成正比。在 MNIST 私有数据集上，训练模型的大小与 Fashion-MNIST 私有数据集上的模型大小相同。因此，计算和内存占用开销的结果也与 Fashion-MNIST 私有数据集上的结果相同，如表 2 所示。

.4.2 更优的效用-隐私权衡。

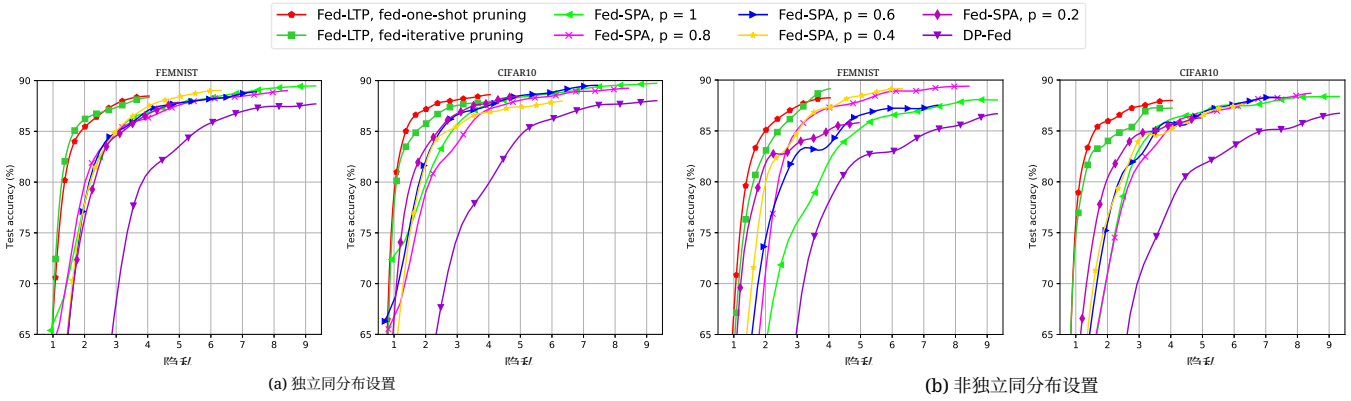
图 10 展示了所有算法在相同隐私预算 ϵ 下所能达到的最佳测试准确率。显然，在较小的隐私预算下，Fed-LTP 能够获得更高的测试准确率。这意味着我们的算法实现了更优的效用-隐私权衡，这与 Fashion-MNIST 私有数据集上的结果相同。

.4.3 高效通信。

在图 11 中，我们展示了所有算法在相同通信成本下的最佳测试准确率。显然，Fed-LTP 以更低的成本取得了更高的测试准确率。这意味着我们的算法实现了更优的效用-通信权衡，这与在 Fashion-MNIST 私有数据集上的结果一致。

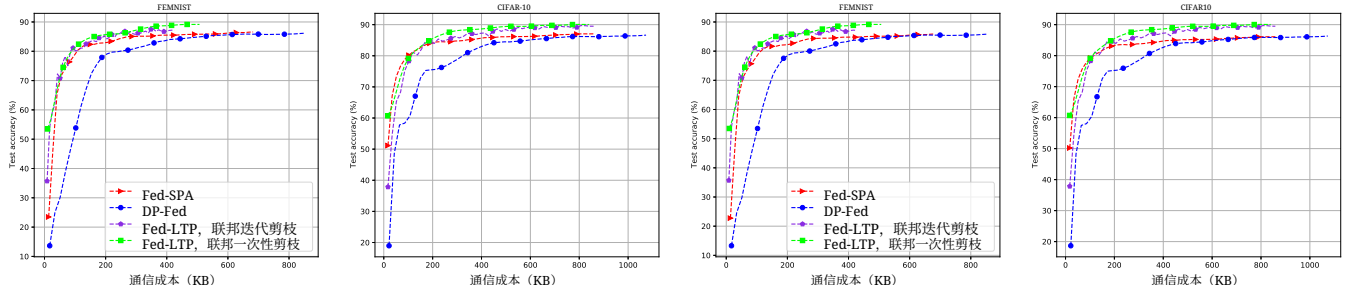
表 3 在 Fashion-MNIST 私有数据集下，Fed-LTP 和基线方法在两个公开数据集的独立同分布和非独立同分布设置下的实验结果。请注意，由于采用相同的模型结构，在独立同分布和非独立同分布设置下，我们具有相同的通信成本和隐私预算 ϵ。同时，表 3 中的“准确率”和“通信量”分别表示全局模型的最佳测试准确率，以及所有通信轮次中上行和下行通信成本的累计总和。Fed-SPA 的压缩率为 p，也可视为最终保留率。对于 DP-Fed，$p = 1$。0 表示未采用加速技术。所有算法的隐私损失 ϵ 均在通信轮次 [43] 中累积。对于 Fed-LTP，ϵ 根据公式 (20) 进行累积，因此与最终保留率 p 无关。

方法（最终保留率：p）	FEMNIST				CIFAR10			
	Acc		通信量（MB）	ϵ	Acc		通信量（MB）	ϵ
	IID	非独立同分布			IID	非独立同分布		
Fed-LTP（联邦迭代剪枝）	72.25	**71.23**	**31.66**	5.35	75.16	72.57	**34.38**	5.35
Fed-LTP（联邦一次性剪枝）	73.65	70.86	35.30	5.35	**75.87**	**73.47**	38.34	5.35
Fed-SPA，$p = 1$	**74.40**	68.40	64.36	9.71	74.40	68.40	64.36	9.71
Fed-SPA，$p = 0.8$	72.21	69.09	57.92	8.78	72.21	69.09	57.92	8.78
Fed-SPA，$p = 0.6$	70.45	65.40	51.49	7.77	70.45	65.40	51.49	7.77
Fed-SPA，$p = 0.4$	69.21	69.23	45.05	6.60	69.21	69.23	45.05	6.60
Fed-SPA，$p = 0.2$	65.26	68.96	38.62	**5.16**	65.26	68.96	38.62	**5.16**
DP-Fed	70.41	68.89	64.36	9.71	70.41	68.89	64.36	9.71



(a) 独立同分布设置　　(b) 非独立同分布设置

图 10。效 用-隐私权衡：不同算法在独立同分布和非独立同分布设置下，分别在 FEMNIST 和 CIFAR10 上的表现。



（a）独立同分布设置　　（b）非独立同分布设置

图 11。不同算法的测试准确率 在两个数据集的独立同分布和非独立同分布设置下，随累计通信成本（兆字节）变化的表现。 gs.