DOI: 10.13718/j.cnki.xdzk.2026.04.011
李杰，刘芮渝，乔德文，等. 基于数据集蒸馏的安全高效一次性交互联邦学习 [J]. 西南大学学报(自然科学版)，2026，48(4)：156-166.

基于数据集蒸馏的安全高效一次性交互联邦学习

李杰[1]，　刘芮渝[2]，　乔德文[3]，　乐俊青[2]，　向涛[2]

1. 中治赛迪信息技术(重庆)有限公司 智慧城市事业部，重庆 404100；2. 重庆大学 计算机学院，重庆 400044；
3. 陆军军医大学(第三军医大学)新桥医院 生物医学信息研究中心，重庆 400037

摘要：联邦学习通过在各客户端本地训练并仅上传模型参数，在无需共享原始数据的情况下实现协同学习，但频繁的参数交互仍带来较高的通信开销。针对上述问题，结合数据集蒸馏与差分隐私技术，提出了一种安全高效的一次性交互联邦学习方案。该方案中，服务器仅向用户提供学习任务的模型结构信息，无需下发待训练模型参数；用户基于本地数据生成低维合成数据并一次性上传，以替代传统的多轮模型参数交互，从而显著降低通信成本。针对合成数据可能泄露原始数据信息的问题，在合成数据生成阶段引入差分隐私机制，并设计自适应噪声注入优化策略，在保障隐私的同时有效缓解模型性能退化。服务器端通过聚合用户上传的合成数据并进行集中训练，获得高性能全局模型。隐私分析表明，所提方案满足本地差分隐私约束，能够有效抵御成员推理攻击。实验结果表明，与现有隐私保护联邦学习方法相比，该方案在显著降低通信开销的同时，在非独立同分布数据场景下取得了更优的模型准确度。

关　键　词：数据集蒸馏；一次性交互学习；联邦学习；差分隐私
中图分类号：TP309；TP18　　**文献标识码**：A
文 章 编 号：1673-9868(2026)04-0156-11
开放科学(资源服务)标识码(OSID)：

Secure and Efficient One-Shot Federated Learning Based on Dataset Distillation

LI Jie[1]，　LIU Minyu[2]，　QIAO Dewen[3]，
LE Junqing[2]，　XIANG Tao[2]

1. *Smart City Business Division，CISDI Information Technology(Chongqing) Co.，Ltd.，Chongqing 404100，China*；
2. *College of Computer Science，Chongqing University，Chongqing 400044，China*；
3. *Bio-Med Informatics Research Centre，Army Medical University(Third Military Medical University)，Xinqiao Hospital，Chongqing 400037，China*

Abstract：Federated learning allows clients to train models locally and share only model parameters instead of raw data，enabling collaborative learning without exposing original data. However，frequent parameter

收稿日期：2026-02-27
基金项目：国家自然科学基金项目(62502525)；陆军军医大学第二附属医院青年博士人才孵化计划项目(2025YQB020)。
作者简介：李杰，高级工程师，主要从事智慧园区、物联网、大模型研究。
通信作者：乐俊青，博士，副研究员。

exchanges still lead to high communication overhead. To address the above issues, a secure and efficient one-shot federated learning scheme was proposed based on dataset distillation and differential privacy techniques. In this scheme, the server only provided users with a description of the model structure for the learning task, without distributing the model parameters for training. Then, users generated low-dimensional synthetic data based on their local data and uploaded them to the server in a single interaction. This method replaced the traditional multi-round model parameter interactions, and significantly reduced the communication costs. To address the potential privacy leakage issue of synthetic data, a differential privacy mechanism was introduced during the synthetic data generation stage and an adaptive noise injection optimization strategy was designed to effectively mitigate model performance degradation while ensuring privacy protection. Subsequently, the server aggregated the synthetic data uploaded by users and performed centralized training to obtain a high-performance global model. Privacy analysis showed that the proposed scheme satisfied local differential privacy constraints and effectively resisted membership inference attacks. The experimental results showed that, compared with existing privacy-preserving federated learning methods, this scheme achieved better model accuracy in scenarios with non-independent and identically distributed data while significantly reducing communication overhead.

Key words: dataset distillation; one-shot learning; federated learning; differential privacy

随着智能终端的普及，数据驱动的机器学习模型在智能感知与决策支持等场景中发挥着重要作用。然而，数据通常分散存储在用户或设备侧，且包含敏感信息，直接共享这些数据易引发隐私泄露风险，形成数据孤岛问题，制约智能模型的发展[1]。联邦学习(Federated Learning, FL)[2-3]通过在本地训练模型、仅上传参数，实现了在不集中原始数据前提下的协同建模，在一定程度上缓解了数据孤岛与隐私风险。然而，传统联邦学习依赖多轮模型参数交互，深度模型参数规模庞大，带来显著通信开销，限制了其在资源受限设备上的部署。同时，上传的模型参数仍可能泄露敏感信息[4-5]，攻击者可通过成员推理攻击[6-7]或模型反演攻击[8]恢复训练数据特征，对隐私构成威胁[9]。

为增强隐私保护，已有研究引入多方安全计算、同态加密、可信执行环境及差分隐私(Differential Privacy, DP)[10]等技术。其中，DP 因具备严格的理论隐私保证和较高效率而被广泛采用。但现有方法多在梯度或参数层面注入噪声[11-13]，在高维空间中往往需要较大噪声强度，导致模型性能下降。

为解决分布式学习场景中存在的通信低效和隐私泄露问题，引入数据集蒸馏[14]的思想，提出基于合成数据的联邦学习方案。在该方案中，通过数据集蒸馏技术生成小规模的合成数据，并在联邦学习中替代高维模型参数进行服务器与用户之间的交互，可以显著降低联邦学习的通信开销。此外，不同于传统方法仅对模型参数进行一次性聚合，合成数据可在服务器端被重复用于多轮次的模型训练与更新，使得全局模型能够充分收敛并得到更优的性能。这一特性也为构建一次性交互联邦学习策略提供了可能，即每个用户仅需一次性上传其本地生成的合成数据即可实现协同训练。

然而，合成数据仍有可能泄露原始数据的分布信息，难以抵御成员推理攻击。为保障合成数据隐私安全，在基于分布匹配(Distribution Matching, DM)[15]数据集蒸馏框架中引入差分隐私技术，并设计了一种新的自适应噪声注入策略，使合成数据在满足本地差分隐私要求的同时，在用于模型训练时仍能得到较高准确度。主要贡献如下：1) 提出了一种基于数据集蒸馏的一次性交互联邦学习框架，用户仅需一次性上传小规模的合成数据，服务器端即可基于该数据进行多轮次模型训练，以得到高准确度的全局模型。2) 不同于传统联邦学习依赖高维模型权重或梯度的多轮次交互进行模型更新，该方法在整个训练过程中仅涉及一次小规模的低维合成数据交互，从而有效避免了模型参数传输带来的高通信开销。3) 将差分隐私处理嵌入合成数据的生成阶段，有效降低全局敏感度，并提出自适应噪声注入策略，在满足本地差分隐私保护的同时保障较高的模型准确度。4) 从理论上证明了所提方案能满足本地差分隐私要求，并在满足 Non-IID 分布的基准数据集上进行了性能对比实验。实验结果表明，该方法在模型准确度和通信效率方面均优于其他隐

私保护联邦学习方法。

1 相关工作

1.1 联邦学习方案

联邦学习可以在本地原始数据不出域的情况下，仅与服务器交换模型参数信息，实现分布式协同训练，已被广泛应用于隐私敏感的分布式学习场景。目前，学术界已提出了大量关于联邦学习的研究成果。其中，FedAvg算法[16]是最经典的一种联邦架构，其通过增加本地训练轮次来减少通信频率，在一定程度上提升了通信效率。然而，该类方法仍依赖多轮模型参数交互，在大规模系统或资源受限环境中通信开销依然较高。

为进一步降低通信成本，研究者提出了一次性交互联邦学习范式，通常结合数据集蒸馏技术，使用户端仅需上传一次小规模合成数据即可完成全局模型训练或模型更新过程的重构。例如，FedSynth[17]利用合成数据进行交互，随后基于合成数据恢复用户端模型更新并进行聚合；Zhou等[18]提出了DOSFL方案，该方案基于本地真实数据集蒸馏得到合成数据，合成数据随后在服务器端聚合，并用于更新全局模型；Song等[19]设计了FedD3方案，该方案通过在用户端采用Coreset-based或KIP-based的数据集蒸馏方法，将原始数据压缩为少量合成样本，并一次性上传给服务器，实现一次性交互联邦学习。尽管现有的一次性交互联邦学习在通信方面保持高效，但其对用户端模型的初始化一致性较为敏感，且蒸馏计算开销相对较高。

1.2 隐私保护策略

在联邦学习的训练过程中，无论通过模型参数还是合成数据进行服务器与用户端之间的交互，都可能遭受诚实但好奇(honest-but-curious)服务器或恶意窃听者发起的成员推理攻击[6-7]和模型反演攻击[8]。为防御这些恶意攻击，现有研究引入了安全多方计算、同态加密和差分隐私等防御技术。其中，最具代表性的工作包括Bonawitz等[20]提出的安全聚合机制，以及文献[21-22]中设计的基于同态加密的联邦学习方案，但这些方法通常存在较高的通信与计算开销。

相比之下，差分隐私因其严格的隐私理论保证和较低的计算复杂度，已被广泛应用于联邦学习的隐私保护。文献[23-24]提出在模型参数更新阶段注入DP噪声，以抵御隐私泄露，但该方法仍面临较高的通信开销，且添加的噪声会影响模型性能。为兼顾通信效率与隐私保护，近几年已有研究开始将DP引入数据集蒸馏过程，并通过合成数据交互来降低通信开销。例如，Chen等[25]在合成数据的表示层添加DP噪声；PPFL-DC则利用DP-SGD方式更新合成数据所需的梯度[26]；NDPDC基于DM方法在固定裁剪阈值下对合成数据进行DP保护[27]；FedDM则在合成数据梯度中引入DP噪声，并通过多轮训练保障模型的较高准确度[28]。

然而，上述方案通常采用固定差分噪声，或者对高维模型权重添加噪声，这在一定程度上降低了合成数据的可用性，并导致模型训练准确度下降。

2 关键技术

2.1 数据集蒸馏技术(DM)

基于分布匹配的数据集蒸馏(DM)[15]方法的核心思想是通过对齐真实数据与合成数据在多种嵌入空间中的特征分布，以此生成能够有效替代原始数据的小规模合成数据集。在DM中，原始数据集 T 与待学习的合成数据 S 会分别经过一系列随机增强 $A(\cdot,\omega)$，并在不同的嵌入空间 $\psi_\vartheta(\cdot)$ 中进行特征提取。最小化两组特征分布之间的差异是该方案的优化目标，其优化目标的表示如下：

$$\min_S E_{\vartheta\sim P_\vartheta}\left\|\frac{1}{|T|}\sum_{i=1}^{|T|}\psi_\vartheta(A(x_i,\omega))-\frac{1}{|S|}\sum_{j=1}^{|S|}\psi_\vartheta(A(s_j,\omega))\right\|_2^2 \tag{1}$$

x_i 是数据集 T 中的一个样本。通过在不同嵌入空间对原始数据与合成数据的分布进行对齐，DM方法能够逼近高维的特征分布，以提升合成数据的泛化能力，且在合成数据生成过程中无需进行传统数据集蒸馏(如DD[14])的双层优化，因此该方法是一种高效的数据集蒸馏方法。

2.2　差分隐私技术

定义 1(差分隐私[10])　设 D 为数据集，机制 $M: D \to R$，其中 D 为定义域、R 为值域。若对于任意邻接数据集 d，$d' \in D$(即二者有且仅有一个样本不同)以及任意输出集合 $S \subseteq R$，机制 M 都满足

$$Pr[M(d) \in S] \leqslant e^{\varepsilon} Pr[M(d') \in S] + \delta \tag{2}$$

则称机制 M 满足(ε，δ)-差分隐私，其中 $\varepsilon > 0$ 为隐私预算。ε 越小，表示隐私保护越强。

敏感度与噪声机制：为了近似实现差分隐私，通常采用向查询函数 f 的输出注入差分噪声，以保证差分隐私保护。函数 f 的 l_2 敏感度定义如下，

$$\Delta f = \max_{d,\, d'} \| f(d) - f(d') \|_2 \tag{3}$$

在该方法中，采用高斯噪声机制 M 来实现差分隐私保护，其添加噪声的形式可表示为

$$M(d) = f(d) + N(0, \sigma^2 \boldsymbol{I}) \tag{4}$$

其中，$N(0, \sigma^2 \boldsymbol{I})$ 表示均值为 0 且协方差矩阵为 $\sigma^2 \boldsymbol{I}$ 的正态分布，$\boldsymbol{I}$ 为单位矩阵。

此外，差分隐私机制还具有顺序组合和并行组合两种重要性质，具体表述如下。

定理 1(顺序组合[29])　假设机制 M_1，M_2，…，M_m 分别满足ε_1，ε_2，…，ε_m-差分隐私，则由这些机制组成的联合机制 $M(d) = (M_1(d), M_2(d), \cdots, M_m(d))$ 满足 $\sum_i^m \varepsilon_i$-差分隐私。

定理 2(并行组合[29])　假设机制 M_1，M_2，…，M_m 分别作用于互不重叠的数据集 D_1，D_2，…，D_m，且分别满足(ε_1，δ_1)-差分隐私，(ε_2，δ_2)-差分隐私，…，(ε_m，δ_m)-差分隐私，则由这些机制组成的联合机制 $M(d) = (M_1(d), M_2(d), \cdots, M_m(d))$ 满足($\max_i \varepsilon_i$，$\max_i \delta_i$)-差分隐私。

3　一次性交互联邦学习方案设计

本节首先介绍基于数据集蒸馏的一次性交互联邦学习的主要流程，整体系统模型如图 1 所示。随后，对方案中 ANIS 策略的具体实现进行详细说明。

该方法通过在用户端生成并上传少量高质量的合成数据，替代传统联邦学习中多轮模型参数或梯度交互，从而显著降低通信开销并缓解隐私泄露风险。该方案包含 2 个核心组件：1) 高效的一次性交互联邦训练架构；2) 用于合成数据生成的自适应噪声注入策略。

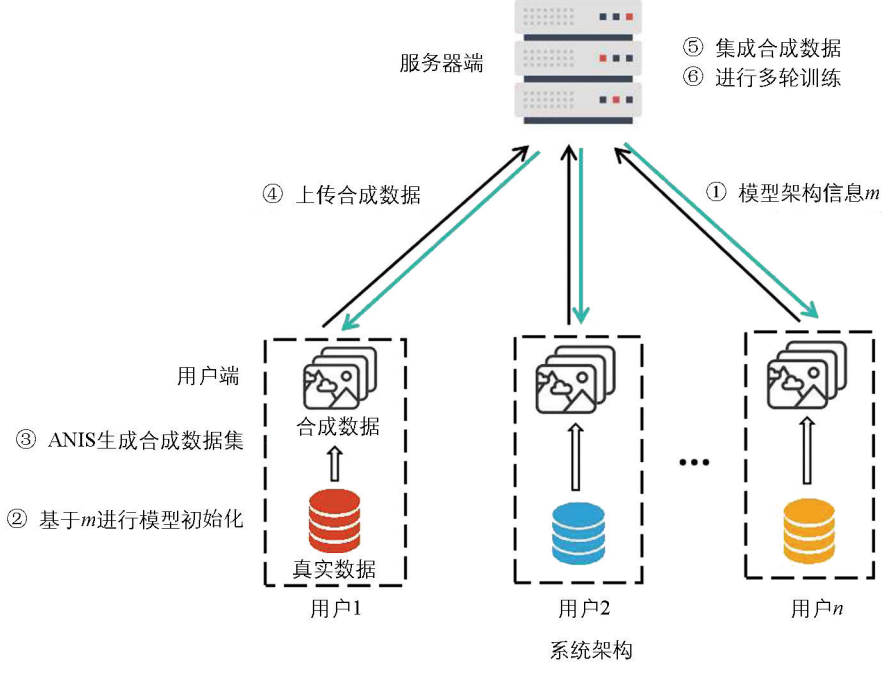


图 1　一次性交互联邦学习架构

3.1 高效的一次性交互联邦训练架构

3.1.1 用户端：隐私保护的合成数据生成

在用户端阶段，服务器首先向所有参与用户下发学习任务的模型架构信息 m，如算法1的步骤1所示。随后，每个用户端 i 在本地执行 R 轮合成数据训练过程，如算法1的步骤2－9所示。具体而言，在每一轮本地训练中，用户端基于模型架构 m 对模型参数进行随机初始化，以避免模型初始化对合成数据生成产生偏置。在算法1步骤5中，从用户 i 的真实数据集 D_i 中按类别随机采样小批量数据 $B_i=\{B_i^1, B_i^2, \cdots, B_i^c\}$。为了在保证合成数据有效性的同时防止隐私泄露，在用户端合成数据更新过程中引入自适应噪声注入策略，对合成数据 S_i 的更新过程进行差分隐私保护，如算法1步骤6所示。该策略基于差分隐私技术，在合成数据优化过程中动态注入差分噪声，从而限制单个样本对合成结果的影响，增强隐私安全性，具体实现过程见算法2。用户端在合成数据训练过程中对不同类别数据分别进行生成训练，每个类别的数据合成都经过 R 轮迭代。最后，将不同类别的合成数据组合在一起，每个用户端得到其最终的合成数据集 $S_i=\{S_i^1, S_i^2, \cdots, S_i^c\}$，见算法1步骤8。

3.1.2 服务器：一次性交互与全局模型训练

完成本地合成数据生成后，每个用户端仅需一次性将其合成数据 S_i 上传至服务器，见算法1步骤10。在算法1步骤11中，服务器对来自所有用户端的合成数据进行聚合，形成统一的合成数据集 $S=\{S_1, S_2, \cdots, S_N\}$。随后，服务器端进行全局模型训练，如算法1步骤12－16所示。全局模型基于架构信息 m 进行随机初始化，并使用聚合后的合成数据集 S 进行集中训练。服务器执行最多 G 轮模型更新，或当模型满足收敛条件时提前终止训练，最终得到全局模型参数。

算法1：基于数据集蒸馏的一次性交互联邦学习

输入：$N=\{1, 2, \cdots, N\}$为本地用户集合，R 为本地训练轮次，G 为全局模型训练轮次，D_i 为用户 i 的真实数据集，S_i 为用户 i 的合成数据，σ 为噪声乘子，η_s 为合成数据更新学习率，C_i^r 为用户 i 在第 r 轮的裁剪阈值，m 为学习任务的模型架构信息，B_i^c 为用户 i 中类别为 c 的小批量数据。

输出：全局模型 θ

1. 服务器向各用户下发 m
2. for $i=1, 2, \cdots, N$ do:
3. for each round $r=1, 2, \cdots, R$ do:
4. 基于 m 随机生成模型参数 θ_i^r
5. 从 D_i 中随机选择小批量数据 $B_i=\{B_i^1, B_i^2, \cdots, B_i^c\}$
6. 基于 B_i^c 和方案更新 S_i^c
7. end for
8. 合并更新后的 S_i^c，得到 $S_i=\{S_i^1, S_i^2, \cdots, S_i^c\}$
9. end for
10. 将合成数据 S_i 上传至服务器
11. 服务器端合并用户合成数据，得到数据集 $S=\{S_1, S_2, \cdots, S_N\}$
12. for $j=1, 2, \cdots, G$ do:
13. 基于 m 随机生成全局模型参数 θ
14. 基于 S 进行训练，并更新 θ
15. if 模型收敛：break
16. end if
17. end for

3.2 自适应噪声注入策略(Adaptive Noise Injection Strategy, ANIS)

在合成数据生成过程中，如何兼顾数据的隐私保护与可用性至关重要。为此，提出一种用于合成数据

生成的自适应噪声注入策略(ANIS)。该策略利用差分隐私机制防御来自诚实但好奇服务器的成员推理攻击，并通过自适应噪声调节在隐私保护与数据可用性之间实现有效平衡。

ANIS 的实现过程如算法 2 所示。在每轮合成数据更新中，用户首先基于当前模型对合成数据进行前向计算，提取特征表示并计算任务损失(步骤 2—3)，进而获得用于更新合成数据的梯度。为限制单样本对更新的影响，ANIS 在步骤 4 中引入梯度裁剪机制，并根据前若干轮梯度的 l_2-范数平均值自适应地设定下一轮裁剪阈值，使阈值随训练过程平滑收敛。完成裁剪后，用户对小批量梯度进行聚合，并依据裁剪阈值与噪声乘子注入高斯噪声，实现差分隐私保护。其中，裁剪阈值为梯度范数设置的最大允许值，噪声乘子 σ 为控制加入噪声强度的参数；随后利用加噪梯度更新合成数据。最终，ANIS 输出经隐私保护的合成数据及对应裁剪参数，用于后续一次性交互联邦学习阶段。

算法 2：自适应噪声注入策略 ANIS

输入：B_i^c，S_i^c，当前轮数 r，θ_i^r，η_s，σ

输出：更新后合成数据 S_i^c 和裁剪阈值 C_i^{r+1}

初始化：用户 i 前几轮的合成数据更新梯度 g_i^{r-1}，g_i^{r-2}，C_i^r，σ

1. for each sample $(x_k, y_k)\in B_i^c$：
2. $\quad\psi_\vartheta=\theta_i^r$. embed //构建嵌入函数
3. $\quad$计算当前样本的任务损失 $\mathscr{L}$

$$\mathscr{L}=\left\|\psi_\vartheta(A(x_k,\omega))-\frac{1}{|S_i^c|}\sum_{s\in S_i^c}\psi_\vartheta(A(s,\omega))\right\|^2$$

4. $\quad$裁剪合成数据更新梯度：$g_i^{r,k}=\nabla\mathscr{L}/\max\left(1,\frac{\|\nabla\mathscr{L}\|_2}{C_i^r}\right)$
5. end for
6. 计算平均梯度 $g_i^r=\frac{1}{|B_i^c|}\sum_{k=1}^{|B_i^c|}g_i^{r,k}$
7. 对梯度注入差分噪声：$\overline{g_i^r}=g_i^r+N(0,\sigma^2C_i^{r2}\boldsymbol{I})$
8. 更新合成数据 $S_i^c=S_i^c-\eta_s\overline{g_i^r}$
9. 更新裁剪阈值 $C_i^{r+1}=(\|g_i^r\|_2+\|g_i^{r-1}\|_2+\|g_i^{r-2}\|_2)/3$
10. return S_i^c 和 C_i^{r+1}

4　隐私分析

本节基于 Rényi 差分隐私(Rényi Differential Privacy，RDP)框架，对所提出的一次性交互联邦学习方案进行隐私性分析。

在该一次性交互联邦学习框架中，用户真实数据始终保留在本地，服务器仅能访问经隐私保护的合成数据。因此，系统的隐私风险主要来源于合成数据生成过程。

1) 单轮合成数据更新的 RDP 保证

在第 r 轮合成数据更新中，用户以采样率 $q=\frac{|B_i|}{|D_i|}$ 从本地数据集中抽取小批量数据 B_i，并对逐样本梯度进行 l_2-范数裁剪，裁剪阈值为 C_i^r。随后，向裁剪并聚合后的梯度中注入均值为 0、标准差为 σC_i^r 的高斯噪声。根据 Rényi 差分隐私理论，对于任意 Rényi 阶数 $\alpha>1$，单轮合成数据更新机制满足 $(\alpha,\varepsilon_{\mathrm{RDP}}^{(r)}(\alpha))$—RDP，其中，

$$\varepsilon_{\mathrm{RDP}}^{(r)}(\alpha)\leqslant\frac{1}{\alpha-1}\log\left(1+q^2\frac{\alpha(\alpha-1)}{2}\min\{4(\mathrm{e}^{\frac{1}{\sigma^2}}-1),2\mathrm{e}^{\frac{1}{\sigma^2}}\}\right)\tag{5}$$

2) 多轮合成数据生成的RDP组合

在合成数据生成阶段，用户共执行R轮更新。由于RDP具有顺序组合性质，用户i在完成全部合成数据生成后满足$(\alpha,\varepsilon_{\mathrm{RDP}}^{(i)}(\alpha))$-RDP。其中，$\varepsilon_{\mathrm{RDP}}^{(i)}(\alpha)=\sum_{r=1}^{R}\varepsilon_{\mathrm{RDP}}^{(r)}(\alpha)$当各轮训练采用相同的采样率与噪声乘子时，上式可简化为$\varepsilon_{\mathrm{RDP}}^{(i)}(\alpha)=R\cdot\varepsilon_{\mathrm{RDP}}^{(\mathrm{single})}(\alpha)$。

3) 一次性交互联邦学习的隐私特性

与传统联邦学习在每一轮训练中都上传梯度或模型参数不同，该方法采用一次性通信策略，即每个用户仅在完成合成数据生成后向服务器上传一次合成数据。随后，服务器端的全局模型训练仅依赖于满足RDP约束的合成数据。根据差分隐私的后处理不变性，服务器端训练过程不会引入额外的隐私泄露。因此，整个一次性交互联邦学习框架的隐私损失不会随着服务器端全局训练轮数的增加而累积。

4) 从RDP到(ε,δ)-差分隐私的转换

根据RDP到(ε,δ)-差分隐私的标准转换定理，对于任意$\delta>0$，机制M_i同时满足(ε_i,δ)-差分隐私，其中，

$$\varepsilon_i=\min_{\alpha>1}\left(\varepsilon_{\mathrm{RDP}}^{(i)}(\alpha)+\frac{\log(1/\delta)}{\alpha-1}\right)\tag{6}$$

5) 隐私保证

由于不同用户的数据集是相互独立的，且各用户的合成数据均在本地独立生成，方案整体满足差分隐私的并行组合性质。因此，所提出的一次性交互联邦学习方案满足(ε,δ)-差分隐私，其中$(\varepsilon,\delta)=\max_i(\varepsilon_i,\delta)$。

5 实验仿真

5.1 实验设置

实验数据集：MNIST[30]、Fashion-MNIST[31]和CIFAR-10[32]。其中，MNIST和Fashion-MNIST各包含60 000张训练图像和10 000张测试图像，图像大小为28×28像素，共10类；CIFAR-10包含50 000张训练图像和10 000张测试图像，图像大小为32×32像素，具有3个颜色通道，同样为10类。实验主要在高度标签偏斜的Non-IID场景下进行评估，即每个用户仅拥有单一类别的训练样本。

训练模型：每类合成数据的数量为10，真实图像的批大小设为256，合成图像更新的学习率设置为$\eta_s=1.0$，每个客户端执行10 000次迭代。对于服务器端，全局模型采用SGD优化器进行训练，模型的学习率设置为$\eta_\theta=0.01$。其余对比方案均按照其对应文献中的参数设置进行实验。此外，在基于差分隐私的策略中设置噪声乘子$\sigma=1$。

5.2 模型性能

5.2.1 模型准确度对比

在本节中，在3个高度偏斜的Non-IID数据集上进行了实验评估。将所提出的联邦学习(FL)架构与传统FL方法FedAvg以及基于数据集蒸馏的一次性交互FL方法(如DOSFL、FedD3)进行对比评估，对比结果如表1所示。结果表明，所提出的方案在Non-IID数据集上的模型准确度均高于其他联邦学习方案。

为验证该方案在差分隐私保护下的模型准确度，与常见的差分隐私联邦学习DPFL、基于DC的差分隐私联邦学习PPFL-DC(优化更新策略：服务器端的合成数据先聚合再全局更新)、基于DM的差分隐私学习NDPDC以及基于DM的差分隐私联邦学习FedDM 4种代表性方案进行对比。根据表2结果可知，所提出的方案在3个Non-IID数据集上均取得了高于其他方法的模型准确度。其中，该方案的模型准确度显著高于DPFL和PPFL-DC方案，但仅略高于NDPDC和FedDM方案。其主要原因在于，NDPDC为集中式训练方式，在一定程度上避免了Non-IID导致的偏置；而FedDM方案对基于批量数据

训练得到的模型梯度进行裁剪，从严格意义上讲并不能满足标准差分隐私保护。本方案模型准确度略高，得益于自适应噪声注入策略，有效降低了差分噪声对模型准确度的负面影响。

表 1　不同联邦学习的准确度对比(无隐私保护)　%

数据集	方案			
	FedAvg[16]	DOSFL[18]	FedD3[19]	本方案
MNIST	95.67±0.32	76.54±0.15	91.86±0.57	96.18±0.08
Fashion	81.02±0.27	50.43±0.24	73.62±0.17	82.65±0.02
CIFAR-10	45.37±0.22	20.17±0.31	42.64±0.37	47.23±0.17

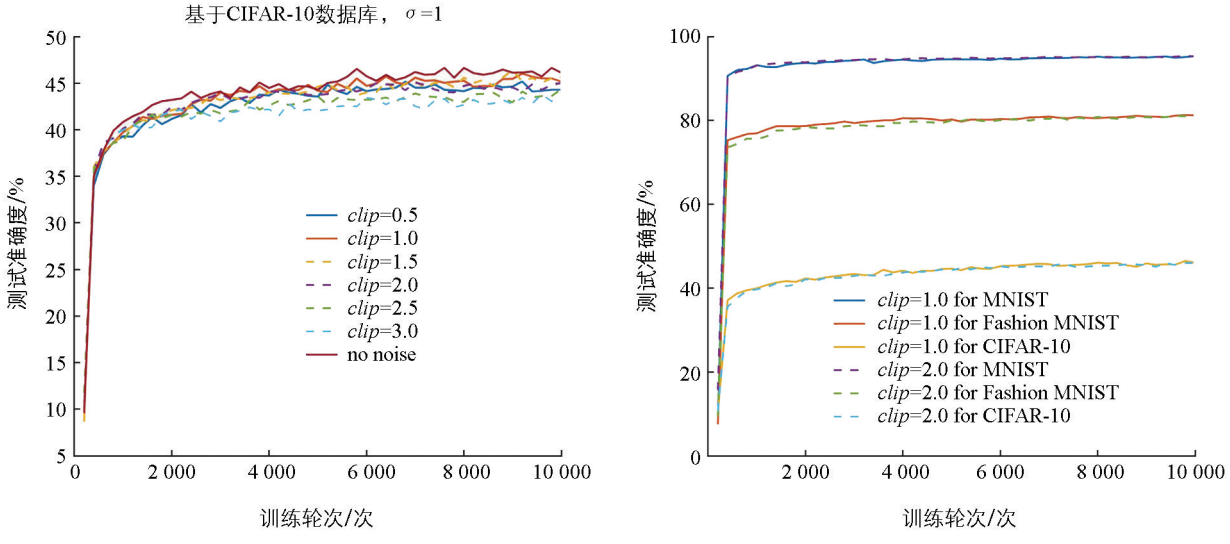
表 2　差分隐私保护下的模型准确度对比　%

数据集	方案				
	DPFL[12]	PPFL-DC[26]	NDPDC[27]	FedDM[28]	本方案
MNIST	80.08±0.23	90.43±0.15	95.35±0.16	95.47±0.19	95.83±0.06
Fashion	74.13±0.09	73.55±0.17	81.03±0.11	81.01±0.13	81.47±0.17
CIFAR-10	43.42±0.08	33.56±0.06	45.98±0.28	45.95±0.23	46.73±0.16

5.2.2　ANIS 的有效性验证

为验证 ANIS 策略的有效性，针对不同梯度裁剪阈值及裁剪策略下的模型准确度进行对比分析。图 2a 中，对比了一次性交互联邦学习在不同裁剪阈值下的模型准确度。从实验结果可以看出，当裁剪阈值等于 1 时，得到的模型准确度相较于其他裁剪阈值更优，并接近于不添加差分噪声的方案。因此，选择合适的梯度裁剪阈值对模型准确度至关重要。

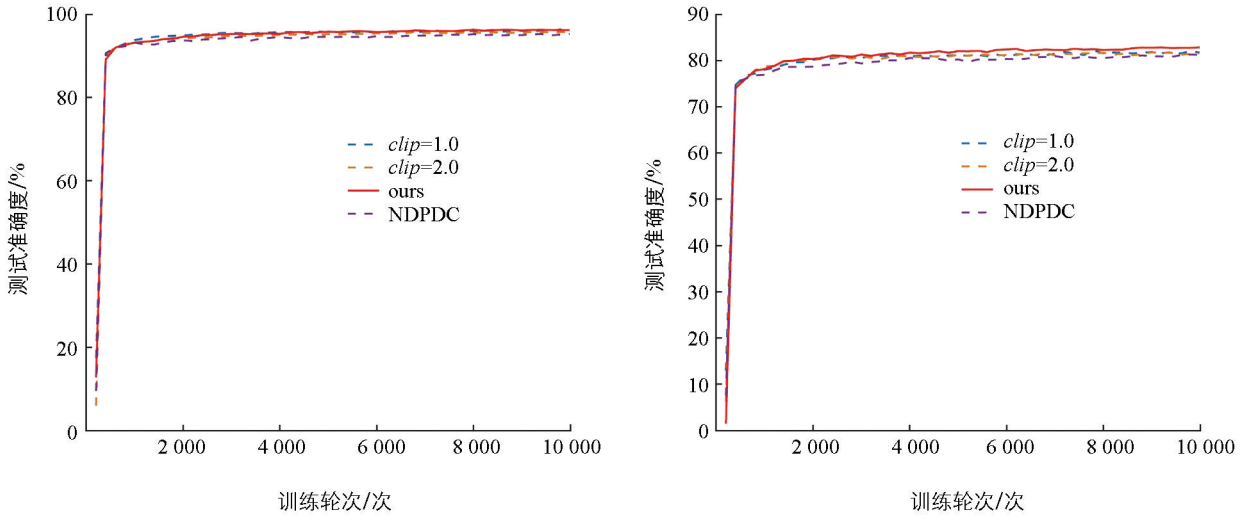
当裁剪阈值过大，所注入的差分噪声幅度增加，会影响模型准确度；若裁剪阈值过小，则梯度偏离原始值过大，同样会导致模型准确度下降。根据图 2b 在不同数据集上得到的实验结果可以看出，不同裁剪阈值可能得到接近的模型准确度，进一步表明裁剪阈值与模型准确度之间存在显著关系。



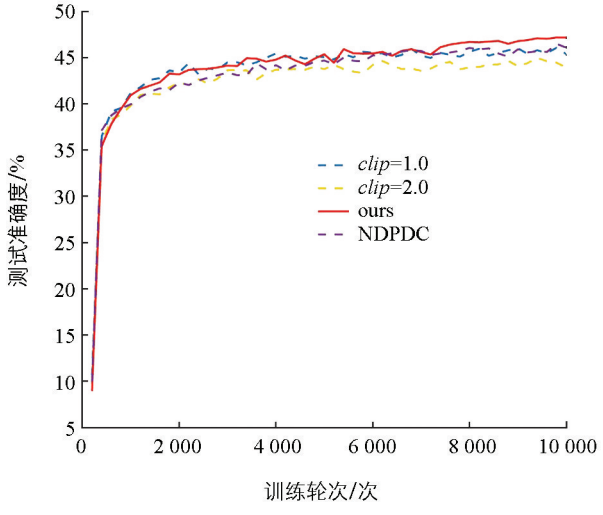
a. 不同裁剪阈值下的模型准确度　　b. 不同数据集中的裁剪阈值影响

图 2　梯度裁剪阈值的影响

此外，对不同阈值裁剪策略下的模型准确度进行了对比，包括对合成数据更新梯度添加固定裁剪阈值(即 $clip=1.0$ 和 $clip=2.0$)的方法、NDPDC 中对合成数据添加固定裁剪阈值的方法以及所提出的自适应噪声注入策略 ANIS。对比结果如图 3 所示，ANIS 在多个数据集上的模型准确度均优于其他固定阈值策略。由此表明所提出的自适应策略能够有效提升模型准确度。



a. MNIST数据集　　b. Fashion MNIST数据集



c. CIFAR-10数据集

图3　不同裁剪策略对模型准确度的影响

5.2.3　通信开销分析

对联邦学习架构下不同方案的通信开销进行了对比分析，包括上传通信开销与下载通信开销。其中，FedAvg、DPFL为传统联邦学习架构，PPFL-DC、FedDM、DOSFL、FedD3为基于数据集蒸馏的联邦学习方法，而NDPDC为集中式学习，因此不参与通信开销对比。为更直观地进行对比，设模型参数大小为W，合成数据大小为H，模型结构信息大小为G，DOSFL的本地迭代次数为K，FedAvg、DPFL、PPFL-DC、FedDM的全局训练轮数为T，其中$W \gg H \gg G$。

在上述设定下，FedAvg、DPFL的通信开销为$2WT$，PPFL-DC、FedDM的通信开销为$2HT$，DOSFL的通信开销为$KH+W$，FedD3的通信开销为$H+W$，所提出方案的通信开销为$G+H$。可以看出，所提出的一次性交互联邦学习方案的通信开销明显低于其他方案。

综上所述，所提出的一次性交互联邦学习方案在模型精度与通信开销方面均具有一定优势，是在Non-IID数据环境下表现较好的联邦学习方案。

6　总结

本文针对数据分散场景下隐私保护与通信效率难以兼顾的问题，提出了一种结合数据集蒸馏与差分隐

私的一次性交互联邦学习方法。不同于传统多轮参数交互方式，该方法以低维合成数据作为唯一交互载体，用户仅需一次上传经差分隐私保护的合成数据，服务器即可集中训练与复用，从而降低通信开销并减少模型参数泄露风险。在隐私方面，将差分隐私机制引入合成数据生成阶段，通过自适应噪声控制实现隐私预算与模型性能之间的平衡，并在理论上具备抵御成员推理攻击的能力。实验结果表明，在 Non-IID 数据环境下，该方法在准确率和通信效率方面均表现出较好的性能。总体而言，该工作为高隐私、低资源场景下的分布式建模提供了一种可行的技术路径。

参考文献：

[1] 刘立伟，傅超豪，孙泽堃，等. 数据要素流通全流程隐私关键技术：现状、挑战与展望 [J]. 软件学报，2026，37(1)：301-325.

[2] KONECNY J，MCMAHAN H B，RAMAGE D. Federated Optimization：Distributed Optimization Beyond the Datacenter [PP/OL]. ArXiv(2015-11-11) [2026-02-10]. http：//arxiv. org/abs/1511. 03575.

[3] KONECNY J，MCMAHAN H B，RAMAGE D，et al. Federated Optimization：Distributed Machine Learning for On-Device Intelligence [PP/OL]. ArXiv(2016-08-08) [2026-02-10]. https：//arxiv. org/abs/1610. 02527.

[4] WANG Z B，SONG M K，ZHANG Z F，et al. Beyond Inferring Class Representatives：User-Level Privacy Leakage from Federated Learning [C] // IEEE Conference on Computer Communications. Paris，France：IEEE，2019：2512-2520.

[5] LE J Q，ZHANG D，LEI X Y，et al. Privacy-Preserving Federated Learning with Malicious Clients and Honest-but-Curious Servers [J]. IEEE Transactions on Information Forensics and Security，2023，18：4329-4344.

[6] NASR M，SHOKRI R，HOUMANSADR A. Comprehensive Privacy Analysis of Deep Learning：Passive and Active White-Box Inference Attacks Against Centralized and Federated Learning [C] //2019 IEEE Symposium on Security and Privacy (SP). San Francisco，CA，USA：IEEE，2019：739-753.

[7] 王恺楠，张玉会，侯锐. 联邦学习中隐私攻击与防御综述 [J]. 信息安全学报，2025，10(2)：219-230.

[8] 郭施帆，缪祥华. 联邦学习中梯度反演攻击与防御研究综述 [J]. 信息安全与通信保密，2025(7)：55-65.

[9] ZHU L，LIU Z，HAN S. Deep Leakage from Gradients [J]. Advances in Neural Information Processing Systems，2019，32：8444-8454.

[10] DWORK C. DifferentialPrivacy [C] //Automata，Languages and Programming. Berlin，Heidelberg：Springer，2006：1-12.

[11] ABADI M，CHU A，GOODFELLOW I，et al. Deep Learning with Differential Privacy [C] //Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security. Vienna Austria. ACM，2016：308-318.

[12] WEI K，LI J，DING M，et al. Federated Learning with Differential Privacy：Algorithms and Performance Analysis [J]. IEEE Transactions on Information Forensics and Security，2020，15：3454-3469.

[13] 薛大暄，杜宜霏，陈红，等. 基于差分隐私的通信高效联邦推荐方法 [J/OL]. 软件学报，1-23. [2026-02-15]. https：//doi. org/10. 13328/j. cnki. jos. 007550.

[14] WANG T，ZHU J Y，TORRALBA A，et al. Dataset Distillation [PP/OL]. ArXiv(2020-02-24) [2026-02-10]. https：//arxiv. org/pdf/1811. 10959.

[15] ZHAO B，BILEN H. Dataset Condensation with Distribution Matching [C] //2023 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV). Waikoloa，HI，USA：IEEE，2023：6503-6512.

[16] MCMAHAN H B，MOORE E，RAMAGE D，et al. Communication-Efficient Learning of Deep Networks from Decentralized Data [PP/OL]. ArXiv(2016-02-17) [2026-02-10]. https：//arxiv. org/abs/1602. 05629v3.

[17] HU S Y，GOETZ J，MALIK K，et al. Fedsynth：Gradient Compression via Synthetic Data in Federated Learning [PP/OL]. ArXiv(2022-04-04) [2026-02-10]. https：//arxiv. org/abs/2204. 01273.

[18] ZHOU Y L，PU G，MA X Y，et al. Distilled One-Shot Federated Learning [PP/OL]. ArXiv(2020-09-17) [2026-02-10]. https：//arxiv. org/abs/2009. 07999.

[19] SONG R, LIU D, CHEN D Z, et al. Federated Learning via Decentralized Dataset Distillation in Resource-Constrained Edge Environments [C] //2023 International Joint Conference on Neural Networks (IJCNN). Gold Coast, Australia: IEEE, 2023: 1-10.

[20] BONAWITZ K, IVANOV V, KREUTER B, et al. Practical Secure Aggregation for Privacy-Preserving Machine Learning [C] //Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security. Dallas, Texas, USA: ACM, 2017: 1175-1191.

[21] ZHANG C L, LI S Y, XIA J Z, et al. BatchCrypt: Efficient Homomorphic Encryption for Cross-Silo Federated Learning [C] //USENIX Annual Technical Conference, 2020: 493-506.

[22] 李瑞芮, 郭瑞, 张应辉, 等. 基于多密钥同态加密的边缘联邦学习隐私保护方案 [J/OL]. 计算机科学, 2026, 1-16. https://link.cnki.net/urlid/50.1075.TP.20250922.1402.020.

[23] 王玉画, 张沁楠, 邱望洁, 等. 自适应拜占庭鲁棒的差分隐私联邦学习 [J]. 中国科学: 信息科学, 2025, 55(11): 2663-2682.

[24] 张淑芬, 汤本建, 田子坤, 等. 基于差分隐私的联邦学习研究综述 [J]. 计算机应用, 2025, 45(10): 3221-3230.

[25] CHEN D F, KERKOUCHE R, FRITZ M. Private Set Generation with Discriminative Information [C] //Proceedings of the 36th International Conference on Neural Information Processing Systems. New Orleans, LA, USA: ACM, 2022: 14678-14690.

[26] ZHANG D, LE J Q, MU N K, et al. Privacy-Preserving Federated Learning Based on Dataset Condensation [J]. IEEE Transactions on Consumer Electronics, 2025, 71(1): 748-760.

[27] ZHENG T H, LI B C. Differentially Private Dataset Condensation [C] //Proceedings 2024 Workshop on AI Systems with Confidential COmputing. San Diego, CA, USA: Internet Society, 2024: 1-10.

[28] XIONG Y H, WANG R C, CHENG M H, et al. FedDM: Iterative Distribution Matching for Communication-Efficient Federated Learning [C] //2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Vancouver, BC, Canada: IEEE, 2023: 16323-16332.

[29] DWORK C, LEI J. Differential Privacy and Robust Statistics [C] //Proceedings of the Forty-First Annual ACM Symposium on Theory of Computing. Bethesda, MD, USA: ACM, 2009: 371-380.

[30] LECUN Y, BOTTOU L, BENGIO Y, et al. Gradient-Based Learning Applied to Document Recognition [J]. Proceedings of the IEEE, 1998, 86(11): 2278-2324.

[31] XIAO H, RASUL K, VOLLGRAF R. Fashion-MNIST: A Novel Image Dataset for Benchmarking Machine Learning Algorithms [PP/OL]. ArXiv(2017-08-25) [2026-02-10]. https://arxiv.org/abs/1708.07747 .

[32] KRIZHEVSKY A. Learning Multiple Layers of Features from Tiny Images[PP/OL]. Computer Science (2009-04-08) [2026-02-10]. https://www.semanticscholar.org/paper/Learning-Multiple-Layers-of-Features-from-Tiny-Krizhevsky/5d90f06bb70a0a3dced62413346235c02b1aa086.

责任编辑 崔玉洁