利用双权重稀疏打包应对联邦学习中的资源受限与数据异质性

Qiantao Yang[1], **Liquan Chen**[1, 2*], **Mingfu Xue**[3], **Songze Li**[1]
[1] 东南大学网络空间科学与工程学院，中国
[2] 紫金山实验室，中国
[3] 华东师范大学通信与电子工程学院，中国
{yangqiantao、lqchen、songzeli}@seu.edu.cn, mfxue@cee.ecnu.edu.cn

arXiv:2601.01840v1 [cs.LG] 5 Jan 2026

摘要

联邦学习已受到研究者的广泛关注，但边缘客户端之间的数据异质性仍是一项关键挑战，常常会降低模型性能。现有方法通过拆分模型和知识蒸馏来增强模型对数据异质性的兼容性。然而，这些方法忽视了客户端有限的通信带宽和计算能力，未能在应对数据异质性与适应有限客户端资源之间取得有效平衡。为解决这一局限性，我们提出了一种基于余弦稀疏化参数打包和双重加权聚合的个性化联邦学习方法（FedCSPACK），能够有效利用有限的客户端资源，并减小数据异质性对模型性能的影响。在 FedCSPACK 中，客户端对模型参数进行打包，并基于余弦相似度选择贡献最大的参数包进行共享，从而有效降低带宽需求。随后，客户端以共享参数包为基础生成掩码矩阵，以提高服务器上稀疏更新的对齐与聚合效率。此外，掩码中嵌入了方向权重和分布距离权重，以实现加权引导的聚合机制，增强全局模型的鲁棒性和泛化性能。在四个数据集上使用十种最先进方法开展的大量实验表明，FedCSPACK 在保持较高模型准确率的同时，能够有效提升通信和计算效率。

代码— https://github.com/NigeloYang/FedCSPACK

引言

联邦学习（FL）因能够在不交换原始数据的情况下，通过客户端之间共享模型参数进行模型协同训练，同时利用客户端的大量数据并保护隐私，而受到广泛关注（McMahan 等人，2017年；Alam 等人，2022年）。然而，不同地理位置的边缘客户端设备所生成和收集的数据存在显著异质性，导致其数据具有非独立同分布（Non-IID）特征。其计算、存储、带宽及其他资源也各不相同，这造成了系统资源的异质性。在此背景下，单一全局模型难以涵盖所有客户端的本地数据集，而采用单一全局模型的传统方法往往会导致客户端之间的本地数据分布出现显著偏差，进而造成收敛缓慢、推理性能不佳，甚至无法在资源受限环境中部署（Li 等人，2021年、2022年；Oh、Kim 和 Yun，2022年）。为解决联邦学习中的异质性问题，一些研究提出了个性化联邦学习（PFL）（Huang 等人，2021年；Tan 等人，2022年a），根据各客户端的数据分布和系统资源为其量身定制本地模型，从而增强模型解决异质性问题的能力。

已有研究表明，模型性能退化的本质是由数据异质性导致的客户端漂移。当本地模型与全局模型之间存在显著的聚合差异时，聚合后的全局模型无法充分适应各客户端的异质数据分布。为减少客户端漂移，Karimireddy 等，2020 使用本地模型在旧轮次与全局轮次之间的差异，对当前轮次的本地更新进行补偿。Li 等人 2020、Acar 等，2021 和 Zhang 等，2022 在本地损失函数中加入正则化，以促使本地模型接近全局模型。此外，还有一些模型聚合策略（Luo 和 Wu，2022；Zhang 等人 2023b）以及知识蒸馏方法（Tan 等，2022b；Xu、Tong 和 Huang 2023）。尽管这些方法提升了模型在面对数据异质性时的性能，却忽视了客户端资源有限这一问题。

在联邦学习中，客户端通常由处理器、内存和带宽各异的实体设备组成，因此设备之间存在资源差异（系统资源异质性）。由于联邦学习依赖频繁的模型交互，资源受限客户端往往难以跟上复杂全局模型的协同训练。这会导致通信瓶颈、计算延迟和参与不平衡（Caldas 等，2018a；Diao、Ding 和 Tarokh，2021）。一些研究通过共享稀疏模型参数、优化模型架构和拆分模型，来减轻资源约束，并提升交互效率及模型两端的整体性能（Jiang 等，2024；Horv´ath 等，2021；Isik 等，2023；Zhang 等人 2023a；Wu 等，2023）。然

*通讯作者
版权所有 © 2026年，人工智能促进协会（www.aaai.org）。保留所有权利。

而，这些研究仅考虑了有限系统资源的不一致性，没有考虑客户端的数据异质性。在现实世界中，数据异质性与客户端资源约束并非彼此孤立，而是相互交织、协同作用的核心挑战。数据异质性会影响模型收敛速度和性能，而系统资源则限制通信和计算效率。有效平衡这两个因素，既能减轻资源受限客户端的负担，又能提升模型性能，从而推动联邦学习在现实世界中的可行部署。

针对现有工作的局限性，我们提出了一种面向个性化联邦学习（PFL）的余弦稀疏化参数打包与双权重聚合方法，能够有效平衡数据异质性与系统资源有限的问题，旨在减少通信瓶颈，同时保持模型性能。具体而言，客户端将模型参数划分为多个扁平参数包，并通过余弦相似度自适应地选择相似度最高的 K 个参数包进行共享。这降低了传输过程中的通信开销，并有效帮助资源受限客户端克服通信瓶颈。在客户端计算共享参数包时，会生成一个用于锚定参数包的掩码矩阵，并为该掩码赋予双重权重值，包括方向权重和分布距离权重，从而帮助服务器有效对齐参数包。在聚合阶段，服务器利用掩码的双重权重值完成加权聚合，加快全局模型吸收新知识，并减轻客户端数据异质性对训练过程的负面影响。为验证 FedCSPACK 的有效性，我们在各种异质场景下进行了广泛实验。结果表明，FedCSPACK 能够在确保模型性能和客户端通信效率的同时，有效平衡数据异质性与资源约束。我们的主要贡献如下：

- 据我们所知，FedCSPACK 是首个在参数包级别实现的个性化联邦学习方法，通过参数打包和双权重聚合，在受限的客户端资源与数据异质性之间实现了平衡。
- 对本地模型参数进行打包可以有效降低通信开销。基于参数包生成个性化掩码，并利用 KL 散度和余弦相似度计算参数包的双权重，从而提高服务器对分散参数包的聚合效率和知识吸收能力，减轻数据异质性对模型性能的影响。

•我们将 FedCSPACK 与 10 种最先进水平（SOTA）的方法进行比较，并在具有不同数据异质性的真实世界数据集上进行了广泛评估。结果表明，FedCSPACK 在保持计算效率的同时，将训练速度提升了 2-5× ，模型准确率提升了 3.34%。

相关工作

在各种异质场景中，联邦学习的效果可能不如各客户端在本地进行的独立训练（Tan 等人，2022c；Lee 等人 2022）。为应对这些挑战，研究者提出了个性化联邦学习（Huang 等人，2021；Yi 等人 2023；Zhang 等人 2023a），使每个客户端都能训练出适合自身的个性化联邦学习模型。

在个性化联邦学习中，FedProx（Li 等人 2020）通过在本地目标函数中引入 L2，约束本地模型与全局模型之间的偏移，但 L2 的大小会影响模型拟合程度与收敛速度之间的平衡。FedNova（Wang 等人 2020）根据各参与方的本地步数对全局模型进行归一化。FedGH（Yi 等人 2023）将计算得到的本地平均表示和类别标签上传至服务器，由服务器训练全局预测头并将其广播给客户端，以替代本地预测头。FedDBE（Zhang 等人 2023a）通过促进服务器与客户端之间的双向知识迁移，缩小表示空间中的域差异。FedAS（Yang、Huang 和 Ye 2024）将全局参数与本地信息相结合，增强全局参数的本地化。然而，这些方案没有深入考虑模型参数之间的划分，容易导致模型性能瓶颈。FedNTD（Lee 等人 2022）使用全局模型和历史模型作为教师模型，促进全面的知识蒸馏，并完成将全局泛化知识和历史个性化知识迁移至本地模型，从而减轻灾难性遗忘。FedPAC（Xu、Tong 和 Huang 2023）通过特征对齐和分类器协作，整体提升模型的准确率和鲁棒性，但对于资源受限客户端而言，模型的实现与优化可能较为困难。Li 等人（Li、He 和 Song 2021 ）通过比较全局模型在两轮训练中的损失，纠正用户模型的优化方向。FedALA（Zhang 等人 2023b）将每轮的全局模型与上一轮的本地模型进行融合，以保留本地个性化信息。

此外，还有基于 Top-k 的模型稀疏方法，例如 FedSPU（Niu、Dong 和 Qin 2025）。尽管这些方法有效解决了联邦学习中的异质性问题或资源约束，但它们只分别考虑了这两个因素之一，忽视了协调数据异质性与有限系统资源所面临的挑战。FedCSPACK 旨在提升资源受限设备上联邦学习的计算效率，并缓解通信瓶颈。通过双权重聚合机制，FedCSPACK 降低了数据异质性对客户端和服务器模型性能的影响，从而实现高效且稳健的个性化模型训练。

问题定义

在联邦学习中，假设有 N 个客户端参与训练，其本地数据集为 D_1，D_2，……，D_N，本地模型为 w_1，w_2，……，w_N，协作训练的目标是

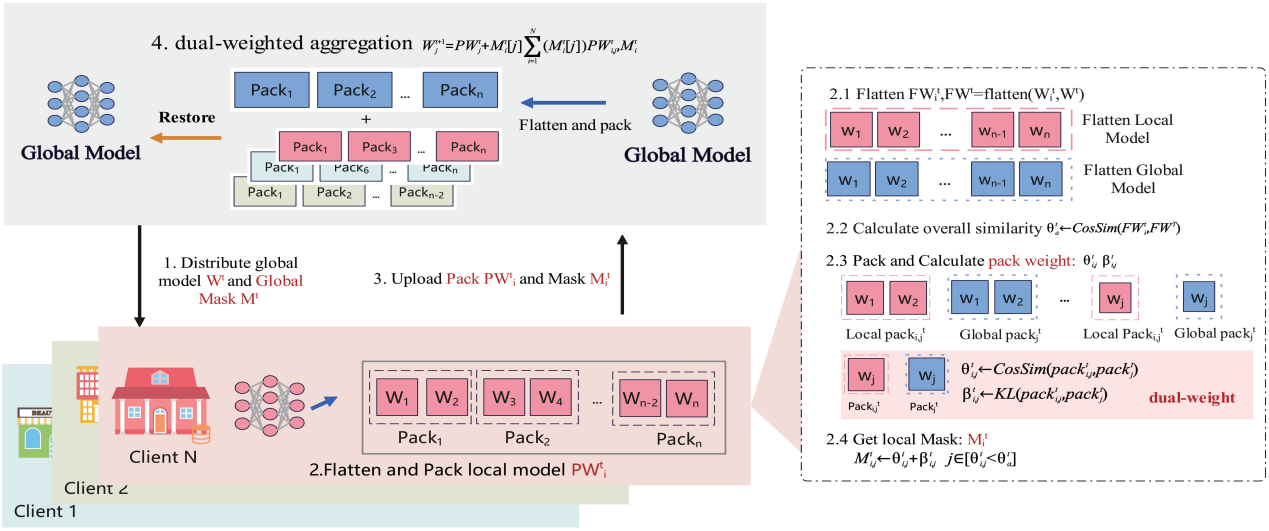


图 1：**FedCSPACK 概述**。客户端展示了本地参数包和掩码的生成与选择的详细过程。服务器端展示了双重加权聚合的详细过程。

最小化全局模型 $F(w)$ 在所有客户端数据上的平均损失:

$$minF(w) = \sum_{i=1}^{N} \frac{|D_i|}{\sum |D_i|} L_i(F(w), D_i) \quad (1)$$

其中，全局模型的参数 w 是一个 d维向量。$L_i(F(w)\ D_k)$ 是客户端 i上本地数据集 D_i 的全局模型 $F(w)$的损失函数。$|D_i|$是客户端 i的数据集大小， $sum\,|D_i|$ 是所有客户端数据集大小之和。

由于客户端数据呈非独立同分布，训练得到的最佳全局模型并不能保证在所有客户端上都具有最佳泛化性能。个性化联邦学习的目标是确定一组最优本地模型 $w_1^a st$, $w_2^a st$, ..., $w_N^a st$，使多个模型能够共存，并确保每个客户端 i 都能根据其本地目标学习最优的个性化模型:

$$w_1^*, ..., w_N^* = \underset{w_1,...,w_N}{argmin} \sum_{i=1}^{N} \frac{|D_i|}{\sum |D_i|} L_i(F_i(w_i), D_i) \quad (2)$$

其中， w_i^* 是客户端（$i \in N$）的最优模型，$L_i(F_i(w_i)$ $D_i)$是客户端 i的损失函数。考虑到实际场景中数据异质性与系统资源受限并存，FedCSPACK 可以通过参数打包和双重加权聚合有效地获得最佳本地模型，并最大限度地降低整体体验损失，从而在有限的客户端资源与数据异质性之间实现平衡。

FedCSPACK 设计

概述

为应对上述挑战，我们提出了 FedCSPACK，通过打包模型参数并有选择地共享部分参数包，高效利用有限资源，同时通过双权重掩码聚合机制增强模型对异质数据的兼容性。FedCSPACK 的框架如图 1 所示。在第 t 轮通信中，FedCSPACK 的步骤如下:

- **步骤 1.** 服务器完成全局模型 W^t和全局掩码 M^t的聚合过程，然后将它们广播给所有客户端。
- 步骤 2-3. 客户端 i在 M^t和 W^t的指导下，使用本地数据 D_i更新本地模型 W_i^t。在共享 W_i^t之前，客户端 i会将 W_i^t, W^t转换为一维向量 FW_i^t, FW^t，并使用余弦相似度计算两个模型之间的整体相似度 θ_a^t。然后，根据 $PACK$的大小，将 FW_i^t和 FW^t打包成参数包 $PW_{i,j}^t$和PW_j^t，并计算相似度阈值 $\theta_{i,j}^t$ 以及每个 $PW_{i,j}^t$和 PW_j^t的离散值 $\beta_{i,j}^t$。同时，使用 Top-k 根据 $\theta_{i,j}^t < \theta_a^t$选出 K 个 $PW_{i,k}^t, k < j$，作为客户端 i 共享的参数包 $PW_{i,k}^t$。最后，为了有效对齐客户端 i 的 $PW_{i,k}^t$，使掩码 $M_{i,j}^t = \theta_{i,j}^t + \beta_{i,j}^t$的有效权重值。
- **步骤4**。服务器收集客户端 i共享的本地模型 W_i^t和本地掩码 M_i^t，整合 M_i^t，并根据其有效权重值完成新一轮全局模型 W^{t+1} 的对齐加权聚合任务。

针对客户端 $i \in [N]$，重复执行上述步骤1-4，直到达到指定训练轮次或模型 W_i 收敛。FedC-SPACK 的实现细节见算法1。

基于Top-k的余弦参数打包

客户端与服务器之间频繁的模型交互，对有限资源客户端构成了重大挑战，因为

算法1： FedCSPACK

输入： 本地数据集 D_i，客户端数量 N，全局轮次 T，本地轮次 E，学习率 η

输出： 全局模型 W^T

服务器执行：
初始化全局模型 w^t
对于 全局轮次 $t \in [T]$ **执行**
  随机采样的客户端子集 S^t
  对于 每个客户端 $i \in S_t$ 并行**执行**
    $PW_{i,\ k}^t,\ M_i^t \leftarrow$**客户端(**$i,\ W^t,\ M^t$**)**
  更新 $W^{t+1},\ M^{t+1} \leftarrow$ 公式 9
  发送 W^{t+1}, M^{t+1} 发送给所有客户端
返回 W^T
本地执行：**客户端(**$i,\ W^t,\ M^t$**)**
接收全局模型：W^t
更新：$W_i^t \leftarrow update(W^t,\ M^t)$
对于本地轮次 $le \in [E]$**执行**
  $W_i^t \leftarrow W_i^t - \eta \nabla f_i(W_i^t)$
$\Delta W_i^t \leftarrow W^t - W_i^t$
展平：$FW_i^t,\ FW^t \leftarrow flattened(W_i^t,\ W^t)$
获取掩码和参数包：$M_i^t,\ PW_i^t \quad _{,k} \leftarrow$ 公式5和8
返回 $PW^t \ _{i,k}, M_i$ 至服务器

大规模模型参数的频繁交互会降低交互效率，并给有限资源客户端带来沉重的通信负担。Top-k（Aji 和 Heafield，2017年）是一种常用的通信优化方法。模型参数在迭代训练过程中会动态变化，而传统的 Top-K 稀疏方法难以在每次迭代中选出合适的参数子集。为缓解通信瓶颈并准确选择参数子集，我们提出一种由余弦相似度引导的分组级 Top-K 参数打包与共享机制，因为余弦相似度越低，推理损失越高，而提升这些相似度可能带来更好的模型性能。

客户端 i 接收全局模型 W^t 和全局掩码M^t后，将本地模型 W_i^t和 W^t展平，得到$FW_i^t,\ FW^t$，然后根据 $PACK$ 将 $FW_i^t,\ FW^t$打包成 $PW_i^t,\ PW^t$，并完成本地模型更新 $PW_i^t[j] \leftarrow PW^t[j],\ if M^t[j] = Valid$。随后，恢复 $PW_i^t[j]$ 并开始进行 t轮本地训练。在共享本地模型之前，客户端 i 将 W_i^t和 W^t展平，得到$FW_i^t,\ FW^t$，并使用余弦相似度（公式 3）计算 FW_i^t与 FW^t之间的总体相似度阈值 θ_a^t。

$$\theta_a^t = CosSim(FW_i^t, FW^t) = \frac{FW_i^t \cdot FW^t}{\parallel FW_i^t \parallel \parallel FW^t \parallel} \quad (3)$$

然后，客户端 i 将根据 $PACK$ 的大小对 FW_i^t 和 FW^t 进行打包，确保每个包 $PW_{i,\ j}^t \leftarrow {}^{FW_i^t}_{PACK}$，$j \in \left|\frac{FW_i^t}{PACK}\right|$包含 $PACK$ 个参数（全局模型 $PW_j^t \leftarrow {}^{FW^t}_{PACK}$），并根据公式 4 计算参数包相似度阈值 $\theta_{i,\ j}^t$。

$$\theta_{i,j}^t = \frac{PW_{i,j}^t \cdot PW_j^t}{\parallel PW_{i,j}^t \parallel \parallel PW_j^t \parallel},\ j \in \left|PW_i^t\right| \quad (4)$$

接下来，客户端 i使用公式 5 选择 K 个参数包 $(PW_{i,\ k}^t,\ k \in j)$，即 $\theta_{i,\ j}^t < \theta_a^t$，作为客户端 i的共享参数包；其余参数包 $PW_{i,\ j\setminus k}^t$将成为该客户端的独有特征，从而避免其余客户端的异质性造成的干扰，并最大限度地发挥每个通信轮次的效用。

$$PW_{i,k}^t = TopK(PW_{i,j}^t, \theta_{i,j}^t, \theta_a^t),\ j \in \left|PW_i^t\right| \quad (5)$$

尽管 $PW_{i,\ k}^t$提高了通信效率，但对 $PW_{i,\ k}^t$进行错位或错误的聚合可能导致全局模型更新产生偏差，从而降低整体模型性能。为确保服务器能够高效索引 $PW_{i,\ k}^t$的位置并有效完成全局模型聚合，我们设计了客户端掩码 $M_{i,\ j}^t$ 来定位 $PW_{i,\ j}^t$的位置，并根据公式 6，将 $\theta_{i,\ k}^t \leftarrow \theta_{i,\ j}^t < \theta_a^t$视为 $PW_{i,\ k}^t$的有效索引位置及其所占聚合权重的比例。

$$M_{i,j}^t = \begin{cases} \theta_{i,k}^t & \theta_{i,k}^t \in \theta_{i,j}^t \\ 0 & \theta_{i,j} \setminus \theta_{i,k}^t \end{cases} \quad (6)$$

通过这种数据包级模型参数选择与共享机制，我们的方法确保个性化模型在资源受限的客户端上进行高效通信的同时，仍能保持模型性能。

掩码双重权重聚合

由于数据异质性，服务器聚合得到的全局模型往往难以适应不同客户端之间变化的数据分布。尽管余弦相似度可以维持参数空间中本地更新与全局方向的一致性，但它只能反映方向对齐，无法衡量参数值之间的距离差异和幅度偏移。这种对幅度的不敏感可能导致相应掩码 $M_{i,\ j}^t$权重过高，过度影响聚合结果，并损害模型稳定性。为此，我们提出一种双权重参数包聚合机制。在基于余弦相似度的方向权重基础上，我们进一步引入 Kullback-Leibler（KL）散度，以量化本地模型参数更新与全局模型参数更新之间的距离偏移。具体而言，在计算 $PW_{i,\ j}^t$ 与 PW_j^t 的数据包相似度阈值 $\theta_{i,\ j}^t$ 时，客户端 i 将使用 KL 散度（公式 7）计算 $PW_{i,\ j}^t$ 与 PW_j^t 之间的距离差异阈值$\beta_{i,\ j}^t$，并将 β_j 作为 $M_{i,\ j}^t$ 的补充权重。

$$\beta_{i,j}^t = \sum PW_{i,j}^t \log \frac{PW_{i,j}^t}{PW_j^t},\ j \in \left|PW_i^t\right| \quad (7)$$

接着，我们根据公式 6 对公式 8 中的掩码权重 $M^t_{i,j}$ 进行新的分配，以确保客户能够更全面地提取每个模型参数包的权重比例。该策略在保留方向对齐优势的同时，有效减轻了分布差异的影响，提升了全局模型吸收各参数包的通用能力。

$$M^t_{i,j} = \begin{cases} \theta^t_{i,k} + \beta^t_{i,k} & \theta^t_{i,k} \in \theta_{i,j} \\ 0 & \theta_{i,j} \setminus \theta^t_{i,k} \end{cases} \tag{8}$$

服务器收到共享参数包$PW^t_{i、k}、\ i \in S^t$和掩码 $M^t_i、\ i \in S^t$后，首先更新全局掩码 $M^{t+1} = sum(M^t_i)$，以获得每个参数包位置的总权重。随后，将全局模型 W^t展平，并按照 PACK 大小完成参数打包（$PW^t_{i、k}$）。在公式 9 的指导下，服务器对客户端的参数包进行加权聚合，并将新的结果 $PW^t_{i、j}$与 PW^t融合，得到新的全局模型 $PW^{t+1}_j \leftarrow PW^t_j + PW^t_{i、j}$。

$$PW^t_{i,j} = \begin{cases} \sum_{i=1}^{S_t} \frac{M^t_{i,j}}{M^{t+1}_j} PW^t_{i,k} & M^t_{i,j} \neq 0 \\ 0 & else \end{cases} \tag{9}$$

随着训练的推进，全局模型可以通过包级别的掩码双重加权聚合，高效、快速地吸收客户端模型知识，同时保持其泛化能力。当客户端合并新的全局模型时，由于客户端保留的本地参数包包含本地个性化知识，全局模型在跨客户端异构数据下的稳定性和泛化性能得以提升。

实验

实验设置

数据集设置：我们使用四个公开可用的图像分类数据集作为实验数据：Fashion-MNIST（FMNIST）、EMNIST、CIFAR-10 和 CIFAR-100。为验证 **FedCSPACK** 的有效性，我们使用卷积神经网络和 ResNet-18 作为测试模型。为模拟这种异质性，我们采用狄利克雷采样（方向(α)）和病态采样，这两种方法在联邦学习中得到广泛应用（Caldas 等，2018b；Acar 等，2021；Li 等人，2022）。α 值越小，非独立同分布特征越明显。狄利克雷采样如图2所示。病态采样和实验设置的相关信息见附录 A。所有实验均重复三次，并取平均值作为实验结果。

基线：我们将 FedCSPACK 与十种最先进水平的方法进行比较。FedAvg（McMahan 等人 2017）、FedProx（Li 等人 2020）和 FedNova（Wang 等人 2020）属于传统方法；聚合方案中的 FedALA （ Zhang 等人 2023b ）可帮助能力较弱的客户端选择全局模型与本地模型之间合适的融合比例。模型拆分（MOON（Li, He, and Song 2021）、Fed-DBE（Zhang 等人 2023a）和 FedAS（Yang, Huang, and Ye2024））以及知识蒸馏（FedNTD（Lee 等人

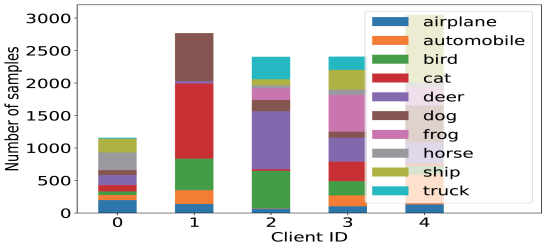


图2：在 CIFAR-10 上使用狄利克雷采样（Dir(0.6)）模拟 5 个客户端的数据异质性。

2022）和 FedPAC（Xu、Tong 和 Huang 2023））能够有效提升准确率和通信效率。FedSPU（Niu, Dong, andQin 2025）通过稀疏化模型参数来提高通信效率。

性能比较

全局模型的整体性能。表1列出了 FedCSPACK 与其他联邦学习最先进水平方法在原始数据集上的结果，涵盖了从浅层到深层的基础模型。在 FMNIST 上，随着数据分布从 Dir(0.3) 转变为 Dir(1.0)，模型准确率总体有所提升。FedCSPACK 在所有数据分布下均表现出色，尤其是在 Dir(1.0) 时，准确率达到 90.73%，显著高于其他方法。在较为复杂的 CIFAR-10 和 CIFAR-100 上，FedCSPACK 在 Dir(1.0) 时分别达到 78.71% 和 43.20%。在 EMNIST 上，模型准确率总体较高，这可能是因为数据量增加有助于模型更好地学习特征。随着数据分布从 Dir(1.0) 转变为 Dir(0.3)，模型准确率有所下降。这可能是由于数据分布同质化导致某些类别的样本减少，进而影响其他方案模型的泛化能力。不过，FedCSPACK 仍能通过双权重参数包聚合方法确保模型保持较高的准确率，尤其是在 EMNIST 上 Dir(0.6) 时达到 86.26%。病态结果见附录 B。总体而言，FedCSPACK 可通过双权重聚合维持较高的模型准确率，有效适应从简单到复杂、从小规模到大规模的数据集。

全局模型在客户端上的泛化能力。如图3所示，Fed-CSPACK 训练的全局模型在适应客户端的异质数据方面表现出良好的模型性能。在泛化性能最佳的客户端5上，FedCSPACK 的模型准确率达到 0.66，而其他方法的最高准确率为 0.62，提升了 6.5%。在泛化性能最差的客户端 3 上，FedC-SPACK 的泛化性能依然稳定，比最先进水平中表现最好的方法（FedDBE 和 FedSPU）高出 20%。尽管 FedCSPACK 在客户端 2 和客户端7上的泛化性能低于 FedPAC 和 FedAS，但在其他客户端上的表现优于 FedCSPACK。这表明，Fed-CSPACK 采用 Top-k余弦参数包和双重加权聚合方案，能够有效确保全局模型的泛

数据集	FMNIST			EMNIST			CIFAR-10			CIFAR-100		
狄利克雷(α)	0.3	0.6	1.0	0.3	0.6	1.0	0.3	0.6	1.0	0.3	0.6	1.0
FedAvg	84.39	86.96	87.65	83.12	84.03	83.98	69.71	74.71	75.01	39.15	40.44	40.34
FedProx	84.39	86.96	87.65	84.55	84.12	83.99	69.58	**74.59**	74.54	38.48	39.58	39.68
FedNova	84.40	**87.50**	**88.68**	**84.97**	84.18	84.03	**70.93**	74.37	75.53	38.65	40.30	39.96
MOON	85.44	86.40	87.90	83.17	84.08	83.94	70.03	74.63	**76.00**	38.43	39.85	39.90
FedDBE	85.06	86.28	88.24	83.14	84.08	83.83	69.66	74.76	74.80	38.33	40.03	39.92
FedAS	42.47	57.27	71.96	70.46	77.32	77.63	43.97	70.00	69.46	20.51	24.12	31.42
FedNTD	84.35	86.67	88.01	83.56	**84.49**	**84.32**	70.32	74.68	75.52	**39.44**	**40.67**	**40.88**
FedPAC	41.53	65.73	69.34	58.11	59.02	66.98	57.23	68.78	58.60	35.39	38.63	39.54
FedALA	84.39	86.96	87.65	83.29	83.98	83.82	69.49	74.64	74.96	38.93	40.03	40.24
FedSPU	**85.29**	88.76	88.25	83.56	84.22	83.91	67.38	74.52	74.62	37.81	38.25	39.35
FedCSPACK（我们的方法）	**88.13**	**89.50**	**90.73**	**85.55**	**86.26**	**86.19**	**73.23**	**77.15**	**78.71**	**41.60**	**42.96**	**43.20**

表1：在四个数据集上采用三种狄利克雷采样时的测试 Top-1准确率（%），其中 $\alpha \in [0.3, 0.6, 1.0]$ 的最佳结果为 FedCSPACK，粗体值表示 SOTA 中的最佳方法。

数据集 / 模型	指标	FedAvg	传统方法	Split	蒸馏	FedALA	FedSPU	**FedCSPACK**
FMNIST/卷积神经网络	通信流量	0.30	0.30	0.30	0.32	0.30	0.07	**0.02**
	Time	1.13	1.45	2.34	1.80	1.18	1.41	**1.21**
EMNIST/卷积神经网络	通信流量	18.18	18.18	18.18	18.29	18.18	4.29	**0.73**
	Time	11.16	14.37	23.11	17.92	11.67	13.92	**11.32**
CIFAR-10/卷积神经网络	通信流量	2.59	2.59	2.59	2.61	2.59	0.61	**0.20**
	Time	1.40	1.83	2.89	2.23	1.46	1.75	**1.42**
CIFAR-100/ResNet18	通信流量	251.00	251.00	251.00	251.80	251.00	49.22	**9.24**
	Time	0.81	1.05	1.67	1.30	0.84	1.01	**0.88**

表2：在四个数据集上，训练轮数为T=100 且 Dir(0.3)时，对比 FedCSPACK 与最先进水平方法的总训练时间（小时）和传输参数数量（GB）。传统方法：FedProx 和 FedNova。Split：MOON、FedDBE 和 FedAS。蒸馏：FedNTD 和 FedPAC。参数打包提升了计算和通信性能。

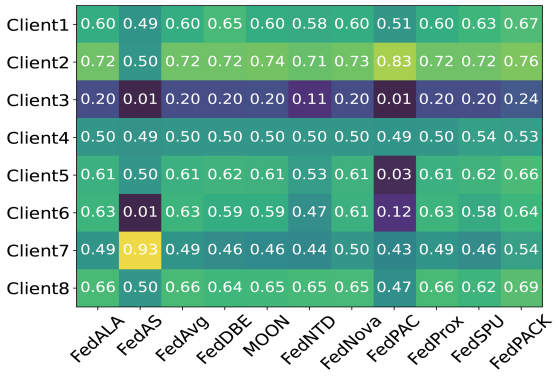


图3：全局模型在 CIFAR-10 和 Dir(0.3) 下客户端上的泛化能力。

化性能并使其更加稳定。更多结果见附录 C。

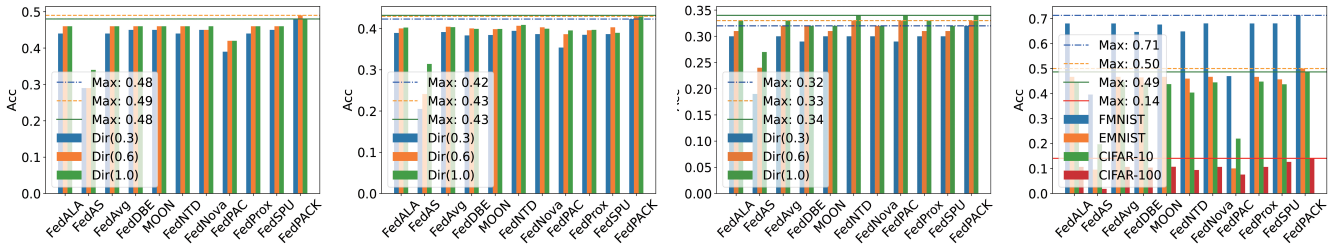
资源消耗

表 2 展示了对比结果，证明计算性能有所提升，资源利用达到最优。**通信流量**。FedCSPACK 显著降低了其在各种数据集上的传输开销。例如，在 EMNIST 上，通信流量从 18.18GB 降至 0.73GB，压缩率达到 96.0%，为所有对比方法中最低。在 CIFAR-100 上，FedCSPACK 将 ResNet18 的通信流量从 251.00GB 压缩至 9.24GB，降幅为27× ；比 FedSPU 低 5× 。这一结果表明，FedCSPACK 的数据包级稀疏通信机制有效减少了上传参数数量，展现出卓越的压缩能力，尤其适用于复杂模型，显著优于最先进的方法。

训练时间。在 EMNIST 和 CIFAR-10 上，与基线相比，FedC-SPACK 的训练时间几乎没有增加，保持了高度稳定。在 CIFAR-100 上，尽管训练时间从 0.81 小时略微增加至 0.88 小时，FedC-SPACK 的训练时间仍分别比 Split、蒸馏和 FedSPU 低 89.8%、47.7% 和 38.11%，展现出卓越的计算效率。这些结果表明，虽然 Top-K参数包选择机制引入了少量额外计算，但由此导致的训练时间增幅微乎其微，在某些场景下甚至有所改善。

资源受限客户端数量

图4展示了客户端参与率对模型性能和稳定性的影响。在图4a至图4c中，随着客户端参与度的下降，模型性能显著波



(a) 客户端参与率=1.0,Dir(0.5)　(b) 客户端参与率=0.6,Dir(0.5)　(c) 客户端参与率=0.3,Dir(0.5)　(d) 客户端参与率=0.3,病态

图4：资源受限情况下，客户端参与率（CPR）对 CIFAR-100 的影响，其中 CPR 表示每轮客户端数量与客户端总数之比。水平线表示 FedCSPACK 的准确率。

动。在低参与度场景下（图4c），由于可用数据分布不完整，数据异质性对现有最先进的方法的影响进一步加剧，导致性能显著下降。然而，FedCSPACK仍保持了良好的稳定性和鲁棒性，准确率分别达到0.32、0.33和0.34。在图4d中，FedCSPACK在四种不同的病态数据划分场景下展现出更优的模型性能。这一结果进一步证明了FedCSPACK在多样化数据分布下的有效性和泛化能力。更多客户端参与率结果见附录D。

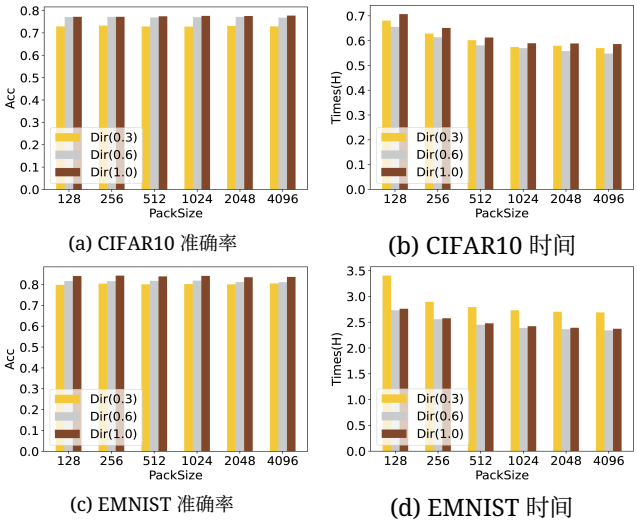
消融实验

权重的影响。我们评估了关键权重对 FedCSPACK 中模型性能的训练影响，包括余弦相似度（CS）权重和 KL 散度权重，以展示它们各自的作用。如表 3 所示，与仅使用 CS 相比，采用双重加权聚合的 FedCSPACK 的平均准确率分别提高了 0.01、0.06、0.04 和 0.05；与仅使用 KL 相比，准确率分别提高了 0.03、0.06、0.03、0.07 和 0.1。结果表明，结合参数方向和参数距离的双重加权聚合能够更全面地评估参数包的贡献，并有效提高模型的训练效率。

权重	轮次				
	10	30	50	70	100
CS	0.33	0.62	0.65	0.70	0.74
KL	0.30	0.57	0.68	0.67	0.69
Ours	0.33	0.63	0.71	0.74	0.79

表 3：在 CIFAR-10 和 Dir(0.5) 上，不同训练阶段的关键权重对模型准确率的影响。

参数包的影响。图 5 进一步验证了$PACK$ 对模型性能的影响。在图 5a 和 5c 中，随着$PACK$ 增加，在不同数据异质性条件下，整体模型性能基本保持不变。这表明，参数包选择技术能够选出贡献最大的 K 个本地模型



(a) CIFAR10 准确率　(b) CIFAR10 时间

(c) EMNIST 准确率　(d) EMNIST 时间

图 5：$PACK$ 大小对模型性能和计算时间的影响。

参数包。双权重聚合机制进一步增强了全局模型整合本地模型知识的能力，从而确保模型稳定性。在图 5b 和 5d 中，随着 $PACK$ 规模增大，模型的总体耗时逐渐减少；在数据量较大的异质场景中，FedCSPACK 的改进更为明显。更多细节见附录 E。

结论

在本研究中，我们提出了一种新颖的参数包级个性化联邦学习方法（FedCSPACK），以应对资源约束与数据异质性并存的问题。其核心思路是对客户端模型参数进行打包，并利用余弦相似度和 Top-k 计算共享参数包，从而降低客户端对有限资源的需求。借助掩码矩阵和双重权重，FedCSPACK 能够确保服务器高效聚合客户端模型的稀疏参数包，保证全局模型的鲁棒性和泛化能力。在多个数据集上进行的大量实验表明，FedCSPACK 达到了最先进水平。

致谢

本研究得到国家自然科学基金（U22B2026、62572121、62372231）、江苏省前沿技术研发计划（BF2025067）、航空科学基金（2022Z071052008）、中央高校基本科研业务费专项资金（2242025K30025）以及东南大学大数据计算中心的支持。

参考文献

Acar, D. A. E.; Zhao, Y.; Navarro, R. M.; Mattina, M.; Whatmough, P. N.; and Saligrama, V. 2021年. 基于动态正则化的联邦学习。载于第*9*届国际学习表征会议（*ICLR 2021*），线上会议，奥地利，*2021*年*5*月*3*日至*7*日。OpenReview.net。

Aji, A. F.; and Heafield, K. 2017年. 用于分布式梯度下降的稀疏通信。载于 Palmer, M.; Hwa, R.; and Riedel, S., 编，*2017*年自然语言处理实证方法会议论文集（*EMNLP 2017*），丹麦哥本哈根，*2017*年*9*月*9*日至*11*日，440–445。计算语言学协会。

Alam, S.; Liu, L.; Yan, M.; and Zhang, M. 2022年. Fe-dRolex：通过滚动子模型提取实现模型异构联邦学习。载于 Koyejo, S.; Mohamed, S.; Agarwal, A.; Belgrave, D.; Cho, K.; and Oh, A., 编，神经信息处理系统进展第*35*卷：*2022*年神经信息处理系统年会，*NeurIPS 2022*，新奥尔良，路易斯安那州，美国，*2022*年*11*月*28*日至*12*月*9*日。

Caldas, S.; Koneˇcn´y, J.; McMahan,H. B.; 和 Talwalkar, A. 2018a. 通过降低客户端资源需求，扩大联邦学习的覆盖范围。*CoRR*，abs/1812.07210。

Caldas, S.; Wu, P.; Li, T.; Koneˇcn´y, J.; McMahan, H. B.; Smith, V.; 和 Talwalkar, A. 2018b. LEAF：联邦设置基准。*CoRR*，abs/1812.01097。

Diao, E.; Ding, J.; 和 Tarokh, V. 2021. HeteroFL：面向异构客户端的计算与通信高效联邦学习。载于第九届国际学习表征会议（*ICLR 2021*），线上活动，奥地利，*2021*年*5*月*3*日至*7*日。OpenReview.net。

Horv´ath, S.; Laskaridis, S.; Almeida, M.; Leontiadis, I.; Venieris, S. I.; 和 Lane, N. D. 2021年。FjORD：采用有序丢弃，在异构目标下实现公平、准确的联邦学习。载于 Ranzato, M.; Beygelzimer, A.; Dauphin, Y. N.; Liang, P.; 和 Vaughan, J. W. 编，神经信息处理系统进展*34*：神经信息处理系统年会 *2021*，*NeurIPS 2021*，*2021*年*12*月*6*日至*14*日，线上活动，12876–12889。

Huang, Y.; Chu, L.; Zhou, Z.; Wang, L.; Liu, J.; Pei, J.; and Zhang, Y. 2021年. 非独立同分布数据上的个性化跨孤岛联邦学习. 收录于 第三十五届 *AAAI* 人工智能会议，*AAAI 2021*，第三十三届人工智能创新应用会议，*IAAI 2021*，*2021*年线上活动，教育进展研讨会第十一届，人工智能，*EAAI 2021*，线上活动，*2021*年*2*月*2*日至*9*日，7865–7873。AAAI出版社。Isik, B.; Pase, F.; G¨und¨uz, D.; Weissman, T.; and Zorzi, M. 2023年。用于通信高效联邦学习的稀疏随机网络。收录于 第十一届国际学习表征会议，*ICLR 2023*，基加利，卢旺达，*2023*年*5*月*1*日至*5*日。OpenReview.net。Jiang, Z.; Xu, Y.; Xu, H.; Wang, Z.; Liu, J.; Qian, C.; and Qiao, C. 2024年。采用自适应模型剪枝的计算与通信高效联邦学习。*IEEE* 移动计算汇刊，23(3): 2003–2021。

Karimireddy, S. P.; Kale, S.; Mohri, M.; Reddi, S. J.; Stich, S.U.; 和 Suresh, A. T. 2020年。SCAFFOLD：用于联邦学习的随机控制平均算法。载于第*37*届国际机器学习会议，*ICML 2020*，*2020*年*7*月*13*日至*18*日，线上活动，机器学习研究论文集第119卷，5132–5143。PMLR。

Lee, G.; Jeong, M.; Shin, Y.; Bae, S.; and Yun, S. 2022年。通过非真实蒸馏在联邦学习中保留全局知识。载于 Koyejo, S.; Mohamed, S.; Agarwal, A.; Belgrave, D.; Cho, K.; and Oh, A. 编，神经信息处理系统进展第*35*卷：神经信息处理系统*2022*年度会议，*NeurIPS 2022*，新奥尔良，路易斯安那州，美国，*2022*年*11*月*28*日至*12*月*9*日。

Li, Q.; Diao, Y.; Chen, Q.; and He, B. 2022年。非独立同分布数据孤岛上的联邦学习：一项实验研究。载于第*38*届 *IEEE* 国际数据工程会议，*ICDE2022*，吉隆坡，马来西亚，*2022*年*5*月*9*日至*12*日，965–978。IEEE。

Li, Q.; He, B.; and Song, D. 2021. 模型对比联邦学习。载于 *IEEE* 计算机视觉与模式识别会议，*CVPR 2021*，线上，*2021*年*6*月*19*日至*25*日，10713–10722。计算机视觉基金会 / IEEE。

Li, T.; Hu, S.; Beirami, A.; and Smith, V. 2021. Ditto：通过个性化实现公平且鲁棒的联邦学习。载于 Meila, M.; and Zhang, T., 编，第*38*届国际机器学习会议论文集，*ICML 2021*，*2021*年*7*月*18*日至*24*日，线上活动，第139卷，机器学习研究论文集，6357–6368。PMLR。

Li, T.; Sahu, A. K.; Zaheer, M.; Sanjabi, M.; Talwalkar, A.; and Smith, V. 2020. 异构网络中的联邦优化。载于 Dhillon, I. S.; Papailiopoulos, D. S.; and Sze, V., 编，*2020*年机器学习与系统会议论文集，*MLSys 2020*，美国德克萨斯州奥斯汀，*2020*年*3*月*2*日至*4*日。ml-sys.org。

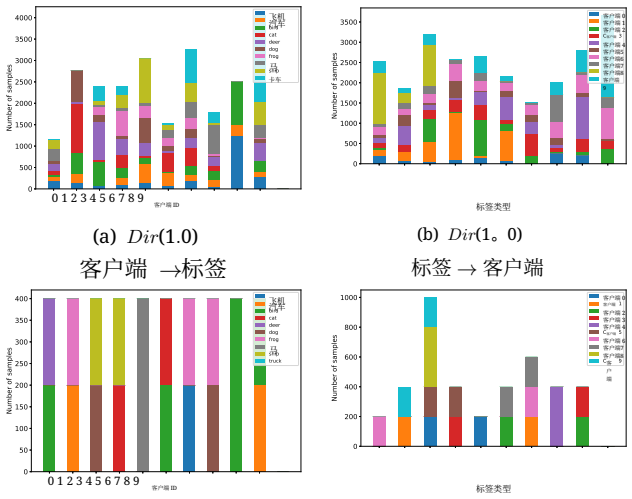
Luo, J.；和 Wu, S. 2022年。适应于适应：跨孤岛联邦学习的个性化学习。载于 Raedt, L. D. 编，第三十一届国际人工智能联合会议论文集，*IJCAI 2022*，奥地利维也纳，*2022*年*7*月*23*日至*29*日，2166–2173。ij-cai.org。

McMahan, B.; Moore, E.; Ramage, D.; Hampson, S.; and y Arcas, B. A. 2017年。基于去中心化数据的深度网络通信高效学习。载于 Singh, A.; and Zhu, X. J., 编，第*20*届国际

人工智能与统计学会议，*AIS-TATS 2017*，*2017*年*4*月*20*日至*22*日，劳德代尔堡，佛罗里达州，美国，机器学习研究论文集第54卷，1273–1282页。PMLR。Niu, Z.; Dong, H.; and Qin, A. K. 2025年。FedSPU：通过随机参数更新，为资源受限设备实现个性化联邦学习。载于 Walsh, T.; Shah, J.; and Kolter, Z., 编，*AAAI-25*，由人工智能促进协会赞助，*2025*年*2*月*25*日至*3*月*4*日，费城，宾夕法尼亚州，美国，19721–19729页。AAAI出版社。Oh, J.; Kim, S.; and Yun, S. 2022年。FedBABU：迈向联邦图像分类的增强表征。载于第十届国际学习表征会议，*ICLR 2022*，线上活动，*2022*年*4*月*25*日至*29*日。OpenReview.net。Tan, A. Z.; Yu, H.; Cui, L.; and Yang, Q. 2022a。迈向个性化联邦学习。*IEEE*神经网络与学习系统汇刊，1–17页。Tan, Y.; Long, G.; Liu, L.; Zhou, T.; Lu, Q.; Jiang, J.; and Zhang, C. 2022b。FedProto：面向异构客户端的联邦原型学习。载于第三十六届*AAAI*人工智能会议（*AAAI 2022*）、第三十四届创新人工智能应用会议（*IAAI 2022*）、第十二届人工智能教育进展研讨会（*EAAI 2022*）线上活动，*2022*年*2*月*22*日至*3*月*1*日，8432–8440页。AAAI出版社。Tan, Y.; Long, G.; Ma, J.; Liu, L.; Zhou, T.; and Jiang, J. 2022c。基于预训练模型的联邦学习：一种对比学习方法。载于 Koyejo, S.; Mohamed, S.; Agarwal, A.; Belgrave, D.; Cho, K.; and Oh, A., 编，神经信息处理系统进展第*35*卷：*2022*年神经信息处理系统年会，*NeurIPS 2022*，美国路易斯安那州新奥尔良，*2022*年*11*月*28*日至*12*月*9*日。Wang, J.; Liu, Q.; Liang, H.; Joshi, G.; and Poor, H. V. 2020年。解决异构联邦优化中的目标不一致问题。载于 Larochelle, H.; Ranzato, M.; Hadsell, R.; Balcan, M.; and Lin, H., 编，神经信息处理系统进展第*33*卷：*2020*年神经信息处理系统年会，*NeurIPS 2020*，*2020*年*12*月*6*日至*12*日，线上。Wu, X.; Liu, X.; Niu, J.; Zhu, G.; and Tang, S. 2023年。大胆而审慎：通过审慎而积极的协作，释放个性化联邦学习的潜力。载于*IEEE/CVF*国际计算机视觉会议，*ICCV 2023*，法国巴黎，*2023*年*10*月*1*日至*6*日，19318–19327页。IEEE。Xu, J.; Tong, X.; and Huang, S. 2023年。通过特征对齐与分类器协作实现个性化联邦学习。载于第十一届国际学习表征会议，*ICLR 2023*，卢旺达基加利，*2023*年*5*月*1*日至*5*日。OpenReview.net。Yang, X.; Huang, W.; and Ye, M. 2024年。FedAS：弥合个性化联邦学习中的不一致性。载于*IEEE/CVF*计算机视觉与模式会议识别，*CVPR 2024*，美国华盛顿州西雅图，*2024*年*6*月*16*日至*22*日，11986–11995。IEEE。Yi, L.; Wang, G.; Liu, X.; Shi, Z.; and Yu, H. 2023。FedGH：采用广义全局头部的异构联邦学习。收录于 El-Saddik, A.; Mei, T.; Cucchiara, R.; Bertini, M.; Vallejo, D. P. T.; Atrey, P. K.; and Hossain, M. S. 编，第*31*届*ACM*国际多媒体会议论文集，*MM 2023*，加拿大安大略省渥太华，*2023*年*10*月*29*日至*11*月*3*日，8686–8696。ACM。Zhang, J.; Hua, Y.; Cao, J.; Wang, H.; Song, T.; Xue, Z.; Ma, R.; and Guan, H. 2023a。消除表示空间中的领域偏差，以实现联邦学习。收录于 Oh, A.; Naumann, T.; Globerson, A.; Saenko, K.; Hardt, M.; and Levine, S. 编，神经信息处理系统进展*36*：*2023*年神经信息处理系统年度会议，*NeurIPS 2023*，美国路易斯安那州新奥尔良，*2023*年*12*月*10*日至*16*日。Zhang, J.; Hua, Y.; Wang, H.; Song, T.; Xue, Z.; Ma, R.; and Guan, H. 2023b。FedALA：面向个性化联邦学习的自适应本地聚合。收录于 Williams, B.; Chen, Y.; and Neville, J. 编，第三十七届人工智能促进协会人工智能会议，*AAAI 2023*；第三十五届人工智能创新应用会议，*IAAI 2023*；第*13*届人工智能教育进展研讨会，*EAAI 2023*，美国华盛顿特区，*2023*年*2*月*7*日至*14*日，11237–11244。AAAI出版社。Zhang, J.; Li, Z.; Li, B.; Xu, J.; Wu, S.; Ding, S.; and Wu, C. 2022。通过逻辑值校准应对标签分布偏斜的联邦学习。收录于 Chaudhuri, K.; Jegelka, S.; Song, L.; Szepesv´ari, C.; Niu, G.; and Sabato, S. 编，国际机器学习会议，*ICML 2022*，*2022*年*7*月*17*日至*23*日，美国马里兰州巴尔的摩，机器学习研究论文集第162卷，机器学习研究论文集，26311–26329。PMLR。

附录 A

数据集设置：在现实世界的联邦学习中，由于偏好、生活方式以及时空背景各不相同，用户数据的分布自然存在差异。这种差异使每个客户端都面临非独立同分布（Non-IID）问题，导致模型准确率下降、收敛速度变慢、模型可用性降低。假设一项联邦学习任务针对特征 x 和标签 y 开展。用户需要从其私有数据中提取样本数据以完成本地训练，即 $(x,\ y) \sim p(x,\ y)$，这可以划分为 $p(x,\ y) = p(x|y)p(y)$。由于估计 $p(x|y)$ 的成本较高，而估计标签先验 $p(y)$ 则极为轻量，因此我们将标签分布不平衡作为非独立同分布划分的高效切入点——其核心思想是让不同客户端的标签分布存在显著差异。基于此，我们采用联邦学习领域中最经典的两种划分策略：**狄利克雷采样**和**病态采样**。



(a) $Dir(1.0)$ 客户端 →标签　(b) $Dir(1。0)$ 标签 → 客户端

(c) $Pathological$ 客户端 →标签　(d) $Pathological$ 标签 → 客户端

图 A1：使用狄利克雷分布和 CIFAR-10 上的数据分片模拟 10 个客户端的数据异质性。

在**狄利克雷采样**中，分布模式由浓度参数 α 调节。图 A1a 直观地展示了这一效果：α 越小，标签分布越尖锐，客户端之间的异质性越强；反之，分布则趋于收敛。与狄利克雷采样的“概率异质性”不同，**病理采样** 采用“物理异质性”：首先，将 CIFAR-10 的 10 个类别平均划分为若干数据块，然后将这些数据块随机且不重叠地分配给 10 个客户端。图 A1c 显示，每个客户端最终只有少数几个完整类别的数据块。这种“类别隔离”更贴近医疗机构和移动终端等真实场景，因为这些场景通常按业务或区域自然地收集数据，从而为强非独立同分布提供了另一种验证基准。总的来说，这两种方法共同涵盖了从轻度到严重的非独立同分布场景，为联邦学习算法的鲁棒性评估提供了标准测试基准。

参数设置与系统实现。：最大全局迭代次数设为 T = 100，客户端总数为 (M =100)。每轮活跃客户端数设为 10，每个客户端进行三/五个本地训练轮次。FMNIST、EMNIST、CIFAR-10 和 CIFAR-100 的学习率分别设为 $0.001-0.05$ 和 $0.05-0.1$。FMNIST 和 EMNIST 的批量大小设为 16/32/64/128，CIFAR-10/CIFAR-100 的批量大小设为 16/32。为验证 FedCSPACK 的有效性，我们采用了三种卷积神经网络模型（每种均包含 3 个卷积层和 3 个线性层）以及一个 ResNet18 模型作为基础模型，所有方法均使用随机梯度下降作为本地优化器。实验在原生 Linux 环境下使用 PyTorch 2.0.0 和 NVIDIA 3090 Ti GPU 实现，每种方法运行三次并报告平均结果。

附录 B

全局模型的整体性能。

我们在三种异构数据环境中对所有方法进行了实验评估，具体结果如表 A1 所示。如表 A1 所示，FedCSPACK 在所有评估数据集上均优于所有对比方法。在 FMNIST、EMNIST、CIFAR-10 和 CIFAR-100 上，其性能指标均达到最高，并大幅领先于其他方案。在 FMNIST 上，FedCSPACK 达到 70.28%，比排名第二的 FedNova（68.06%）高出约 2.2%。在 EMNIST 上，其得分为 49.00%，比次优方法 FedAvg 和 FedALA（均为 46.67%）高出约 2.3%。

在整体性能较低的 CIFAR-10 子集上，FedCSPACK 仍以 13.08% 的准确率保持明显领先，较 MOON（10.71%）高出约 2.4%。尽管其他方法（例如 FedAS、FedPAC）在特定数据集上表现严重不佳（例如，FedAS 在 EMNIST 上的得分仅为 9.55%），FedCSPACK 仍能在不同数据集上持续取得领先结果。相比之下，FedAvg 和 FedProx 等传统方法在不同数据集上的性能停滞不前（例如，在 FMNIST 上为 68.03%），而 FedCSPACK 实现了系统性突破——其优势在 CIFAR-10 和 EMNIST 等更复杂的数据集上尤为显著。总之，通过双权重聚合，FedCSPACK 保持了较高的模型准确率，展现出对从简单到复杂、从小规模到大规模数据集的有效适应能力。

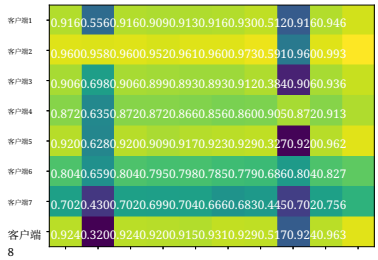
附录 C

全局模型在客户端上的泛化能力。

图 A2 显示，在八个本地客户端中，FedCSPACK 训练得到的全局模型能够有效适应本地客户端的异质数据。如图 A2 所示，FedCSPACK 的全局模型始终能够取得显著更高的本地测试准确率

方法	FedALA	FedAS	FedAvg	FedDBE	MOON	FedNTD	FedNova	FedPAC	FedProx	FedSPU	FedCSPACK
FMNIST	68.03	39.60	68.03	64.64	67.60	64.85	68.06	46.97	68.03	68.18	70.28
EMNIST	46.67	9.55	46.67	46.44	46.67	45.97	46.68	10.05	46.67	44.18	49.00
CIFAR-10	44.78	19.72	44.78	44.02	43.77	40.39	44.47	21.97	44.78	43.21	47.65
CIFAR-100	10.65	1.87	10.65	10.39	10.71	9.41	10.66	7.59	10.65	11.13	13.08

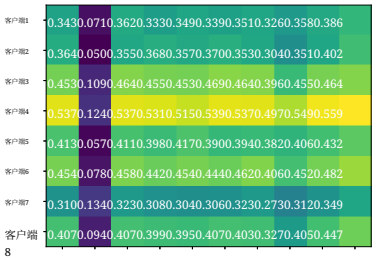
表 A1：四个数据集上病态采样的测试准确率（%）。背景色表示我们的方法。



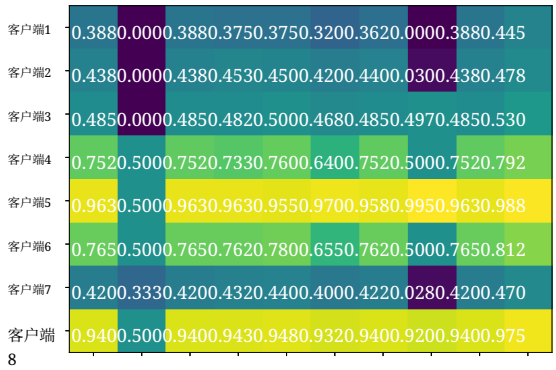
(a) Dir(0.3)，FMNIST



(b) Dir(0.3)，EMNIST



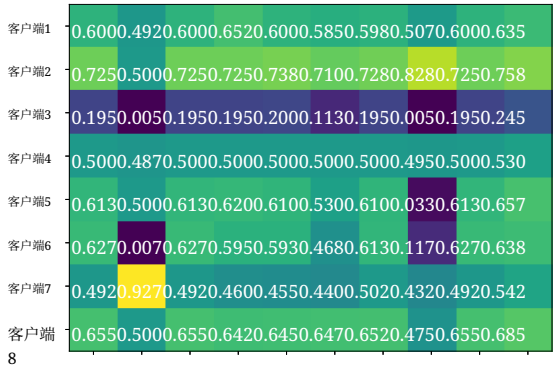
(c) Dir(0.3)，CIFAR-100



(d) 病态，FMNIST



(e) 病态，EMNIST



(f) 病态，CIFAR-10



(g) 病态，CIFAR-100

图 A2：全局模型在客户端上的泛化能力（$Dir(0.3)$ 和病态采样）。使用卷积神经网络模型进行测试

在狄利克雷和数据分片划分场景下，均优于其他方案。当狄利克雷参数 α 从 0.3 变为 1.0 时，所有方法的客户端泛化能力均有所提升，但在强非独立同分布条件下，FedCSPACK 的提升最为显著。这表明其能够出色地抵御严重偏斜的标签分布。当 $alpha = 0.3$ 时，FMNIST、EMNIST、CIFAR-10 和 CIFAR-100 的局部最差值约为 32%、35.7%、1.3% 和 5%，FedCSPACK 的提升率则分别为 96.3%、84.1%、76.8% 和 40.2%。

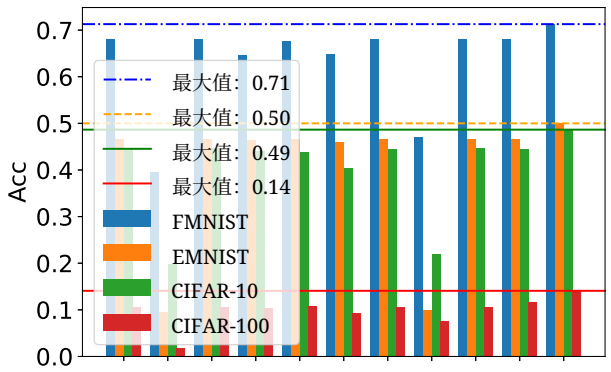
对于图 A2d、图 A2e、图 A2f 和图 A2g，FedCSPACK 在所有客户端上的平均准确率分别保持在 68.63%、67.2%、58.63% 和 21.31%，这表明余弦 Top-k 参数打包技术能够有效增强客户端吸收知识的能力，并提升全局模型的泛化能力。总而言之，无论是在狄利克雷划分还是数据分片划分下，FedCSPACK 在全局到局部的泛化方面均表现出显著优势。在 $\alpha < 0.6$ 或高度异质的数据分片场景下，其优势尤为突出，证明稀疏参数打包和双权重聚合能够在保留全局模型对数据异质性的适应能力的同时，有效训练个性化模型。

附录 D

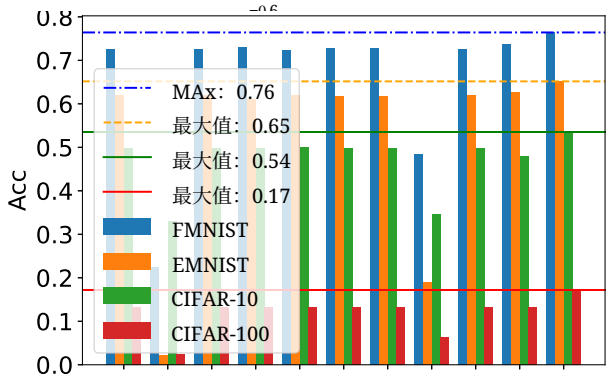
资源受限客户端数量。
考虑到客户端可能因资源约束而断开连接，图 A3 和 A4 展示了客户端参与成功率对模型性能和稳定性的影响。在图 A3 中，当客户端参与率为 =0.6 时，大多数方法（如 FedAvg、Fed-PAC 和 FedProx）在 CIFAR-10 和 CIFAR-100 上的准确率都大幅下降，甚至低于 0.1。相比之下，FedCSPACK 在不同数据集上均表现出更高的准确率：FMNIST 上为 0.71，EMNIST 上为 0.50，CIFAR-10 上为 0.49，CIFAR-100 上为 0.14。当客户端参与率为 = 1.0 时，整体准确率普遍有所提升，FedCSPACK 的表现也进一步提高，在 FMNIST 上达到 0.76，在 EMNIST 上达到 0.65，在 CIFAR-10 上达到 0.54，在 CIFAR-100 上达到 0.17。这些结果表明，客户端参与度较低时，数据异质性对模型聚合过程的影响更为显著。传统方法依赖所有客户端的更新或随机抽样的客户端更新，因此容易受到“少数派主导”问题的影响。相反，FedCSPACK 通过采用双权重掩码机制和分组级稀疏通信，有效地筛选

高质量的参数更新能够抑制噪声干扰，即使客户端参与率较低，也能保持性能稳定。这表明 FedCSPACK 在应对客户端断连和数据异质性挑战方面具有卓越能力。

在图 A4 中，FedCSPACK 在三个典型数据集（FMNIST、EMNIST 和 CIFAR-10）上表现出色，其直方图长度的波动显著小于其他一些模型，表明即使客户端参与资源有限，也能保持良好的稳定性。即使在异质性极高（Dir(0.3)）且客户端参与率极低的极端条件下



(a) 客户端参与率



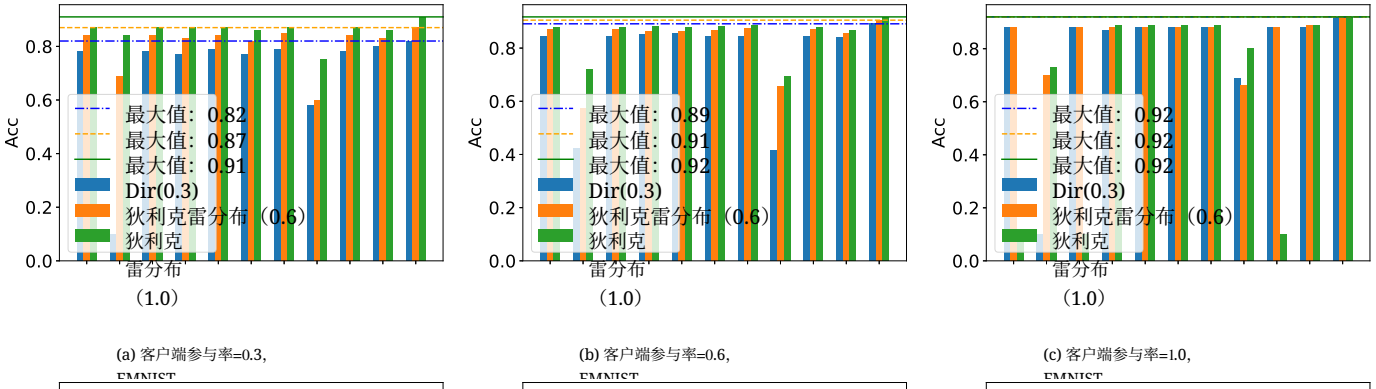
(b) 客户端参与率 =1.0

图 A3：有限资源下客户端参与率（CPR）对病态数据的影响，其中 CPR 表示每轮客户端数量与客户端总数的比值。水平线表示 FedCSPACK 的准确率。使用卷积神经网络模型进行测试

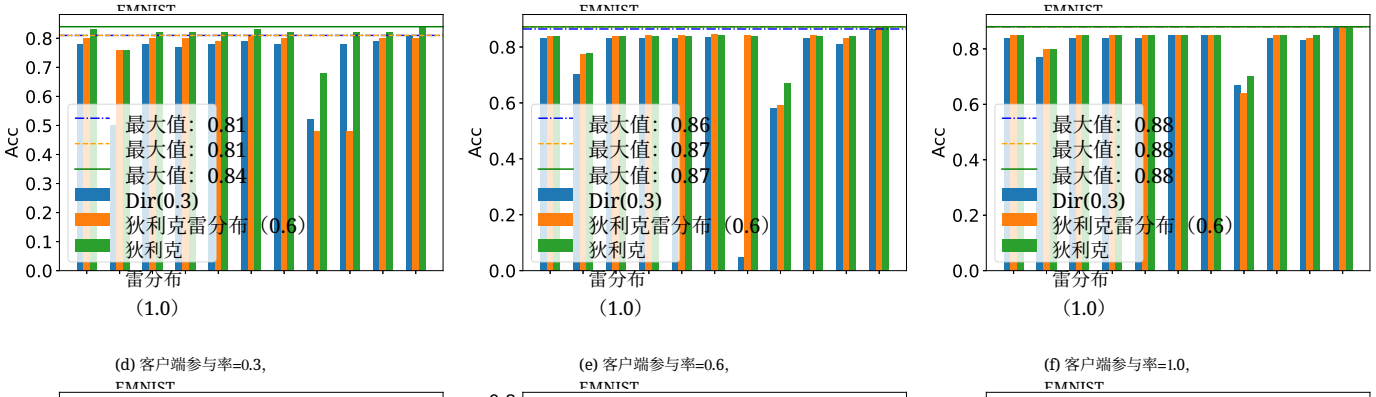
（客户端参与率=0.3），FedCSPACK 仍保持最高准确率，显著优于最先进方法。其卓越性能得益于所提出的双权重掩码机制和分组级稀疏通信策略，这两者有效实现了高质量更新，并在资源有限和数据高度异质这双重挑战下维持了模型的一致性与稳定性。这表明，FedCSPACK 不仅适用于标准联邦学习场景，也具有实际部署价值。

附录 E

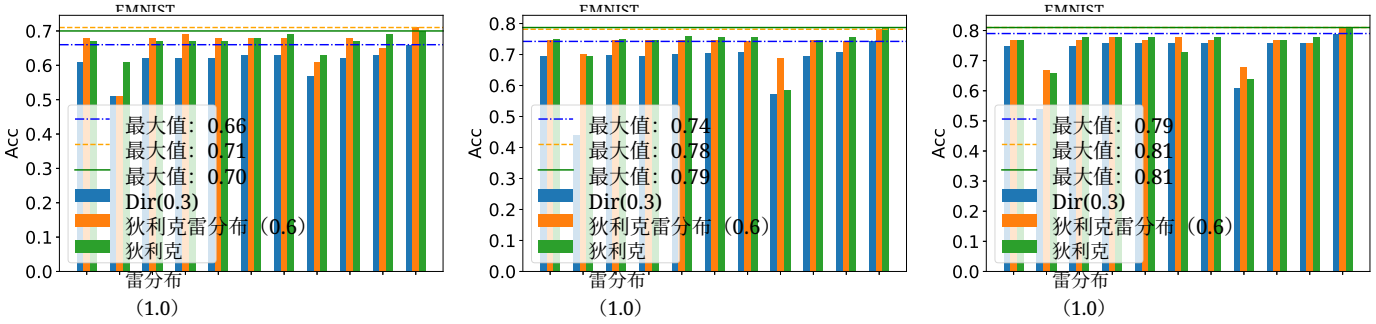
参数包 $PACK$ 大小。
我们进一步在图 A5 中验证参数包对模型性能和计算时间的影响。如图 A5a 和图 A5c 所示，在三种狄利克雷数据异质性下，尽管模型准确率随着增加而提高



(a) 客户端参与率=0.3, EMNIST　(b) 客户端参与率=0.6, EMNIST　(c) 客户端参与率=1.0, EMNIST

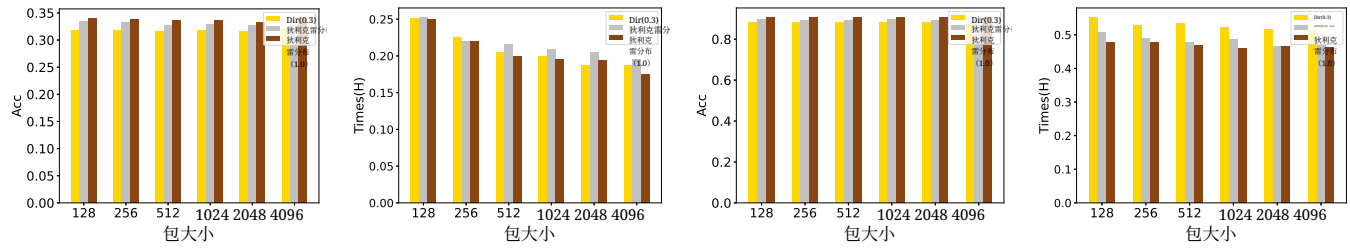


(d) 客户端参与率=0.3, EMNIST　(e) 客户端参与率=0.6, EMNIST　(f) 客户端参与率=1.0, EMNIST



(g) 客户端参与率=0.3, CIFAR-10　(h) 客户端参与率=0.6, CIFAR-10　(i) 客户端参与率=1.0, CIFAR-10

图 A4：有限资源客户端参与率（CPR）对 FMNIST、EMNIST 和 CIFAR-10 的影响，其中 CPR 表示每轮客户端数量与客户端总数之比。水平线表示 FedCSPACK 的准确率。使用卷积神经网络模型进行测试



(a) CIFAR-100，准确率　(b) CIFAR-100，时间　(c) FMNIST，准确率　(d) FMNIST，时间

图 A5： $PACK$ 大小对模型性能和计算时间的影响。使用卷积神经网络模型进行测试

对于 $PACK$ 大小，其影响程度有限。尤其是当$PACK$ 大小超过 512 时，准确率的提升几乎停滞不前。这表明，无论数据集复杂还是简单，其对$PACK$ 大小的敏感度都不高。在图 A5b 和图 A5d 中，随着$PACK$ 大小增加，CIFAR-100 的计算时间有所下降，而 FM-NIST 的计算时间降幅相对较小。这表明，在不同的$PACK$ 大小下，FedCSPACK 都能有效提高复杂数据集的处理效率。

总而言之， $PACK$ 大小对模型准确率的影响有限，尤其是当 $PACK$ 大小超过 512 时，准确率几乎不再提高。然而， $PACK$ 大小会显著影响计算时间。特别是当 $PACK$ 大小超过 512 时，计算时间将大幅减少。因此，在实际应用中，应根据可用计算资源和数据集的复杂度选择最合适的 $PACK$大小，以实现模型性能与计算资源使用之间的最佳平衡。