CA-AFP：簇感知自适应联邦剪枝

Om Govind Jha
om22@iiserb.ac.in
IISER Bhopal 博帕尔，
印度

Harsh Shukla
harsh22@iiserb.ac.in
IISER Bhopal 博帕尔，
印度

Haroon R. Lone
haroon@iiserb.ac.in
IISER Bhopal 博帕尔，
印度

arXiv:2603.01739v1 [cs.LG] 2 Mar 2026

摘要

联邦学习（FL）在实际部署中面临重大挑战，原因在于客户端之间存在统计异质性，以及资源受限设备造成的系统异质性。基于聚类的方法虽可缓解统计异质性，剪枝技术也能提升内存和通信效率，但这些策略通常是分开研究的。

我们提出了 CA-AFP，这是一种通过执行簇特定模型剪枝来共同应对这两项挑战的统一框架。在 CA-AFP 中，首先将客户端分组为多个簇，并在训练过程中对每个簇的独立模型进行自适应剪枝。该框架引入了两项关键创新：（1）一种具备簇感知能力的重要性评分机制，结合权重幅值、簇内一致性和梯度一致性来确定待剪枝的参数；（2）一种迭代剪枝计划，逐步移除参数，同时通过权重再生实现模型自愈。

我们在两个广泛使用的人体活动识别基准数据集 UCI HAR 和 WISDM 上评估 CA-AFP，并采用基于自然用户的联邦划分。实验结果表明，CA-AFP 在预测准确率、客户端间公平性和通信效率之间实现了良好的平衡。与基于剪枝的基线方法相比，CA-AFP 在微调受限的情况下，能够持续提升准确率并降低客户端间的性能差异，同时所需通信量远低于基于密集聚类的方法。它对不同非独立同分布程度的数据也表现出鲁棒性。最后，消融研究分析了聚类、剪枝方案和评分机制的影响，为设计高效且自适应的联邦学习系统提供了实用见解。我们的实现已公开发布于 https://anonymous.4open.science/r/CAAFP-C36E/。

ACM 参考文献格式：
Om Govind Jha、Harsh Shukla 与 Haroon R. Lone。2026年。CA-AFP：簇感知自适应联邦剪枝。载于《会议全称会议录》（会议名称）。ACM，美国纽约州纽约市，12页。https://doi.org/10. 1145/XXXXXXX.XXXXXXX

1 引言

联邦学习（FL）已成为协作式机器学习的重要范式，使模型能够在去中心化设备上训练，同时保护用户数据隐私 [24]。这种范式尤其适用于涉及边缘设备所生成敏感数据的应用，例如智能手机、智能手表和物联网传感器；这类设备常见于人体活动识别（HAR）等领域 [14, 20, 28]。

在真实环境中部署联邦学习面临两项根本挑战：统计异质性 [18] 和系统异质性 [29]。当不同客户端之间的数据分布存在显著差异时，就会产生统计异质性，例如用户打字模式的差异，或人体活动识别应用中个人活动模式的差异。系统异质性则源于硬件和网络条件的差异，包括不同设备在计算速度、内存容量、带宽可用性和能源约束方面的差异 [3]。

统计异质性会导致非独立同分布数据，这对联邦优化尤为不利，从而减缓收敛，并产生无法很好地泛化到各个客户端的次优全局模型 [19]。为缓解这一问题，已提出个性化联邦学习（PFL）方法，旨在为单个客户端或客户端群体定制模型 [18]。在这一研究方向下，基于聚类的联邦学习已成为一种重要策略，它将具有相似数据分布的客户端分组，从而实现更相关、更有效的聚合。以 FedCHAR 为代表的方法表明，基于更新相似度对客户端进行聚类，可以在异质数据设置下显著提升个性化与鲁棒性 [21]。

在真实世界的联邦学习系统中，必须同时应对统计异质性和系统异质性。SAFL [17] 和 FLCAP [25] 等近期方法尝试通过客户端聚类应对统计异质性，并通过模型剪枝应对系统异质性。然而，这两种框架在制定剪枝决策时都存在固有局限。具体而言，在模型参数和簇分配仍在不断变化时就进行剪枝，导致稀疏化依赖于早期且可能不稳定的优化信号。SAFL主要利用批归一化统计量所提供的结构线索，而 FLCAP则主要依靠权重幅值来估计参数重要性。因此，这些方法无法捕捉同一簇内多个客户端的集体相关性。结果是，在单个客户端层面较弱、但在簇级别始终重要的参数可能会被过早移除。此外，缺乏可控且渐进的稀疏化机制，可能导致非独立同分布数据分布下的不稳定性和准确率下降。

为了解决这些局限，我们提出了*CA-AFP*（簇感知自适应联邦剪枝），该方法将簇形成与稀疏化解耦，并引入一种上下文感知的剪枝策略。CA-AFP 首先让客户端模型在初始密集训练阶段趋于稳定，随后形成簇分配并启动剪枝。这一设计确保重要性估计基于可靠且已收敛的学习行为。此外，CA-AFP 采用迭代式自适应掩码

会议名称，地点 2026年。ACM ISBN
978-1-4503-XXXX-X/18/06
https://doi.org/10.1145/XXXXXXX.XXXXXXX

调度机制，在训练过程中动态更新稀疏模式，使模型能够通过重新激活先前被剪枝、后来变得重要的权重来“修复”。

我们在两个标准的人体活动识别基准数据集 WISDM 和 UCI-HAR 上评估 CA-AFP。这两个数据集分别代表真实和受控的联邦设置，且具有显著的用户间异质性。我们将其与基于稠密聚类的 基线方法（FedCHAR [20], ClusterFL [27]）和基于剪枝的 基线方法（EfficientFL [29], FedSNIP [30]）进行比较，从而能够在公平的条件下比较各方法如何分别应对统计异质性和系统异质性。稠密基线着重使用全参数化模型缓解统计异质性，而剪枝基线则主要通过稀疏化应对系统异质性并提高通信效率。

实验结果表明，CA-AFP 的表现始终与基于剪枝的基线方法相当或更优，同时在相近的稀疏度下，其准确率接近稠密方法。在两个数据集上，与 EfficientFL 和 FedSNIP 相比，CA-AFP 的公平性更高，且通信成本显著低于稠密方法，即使微调次数有限也是如此。增加微调后，CA-AFP 进一步缩小了与稠密基线之间的性能差距。这些结果表明，CA-AFP 能有效保留在簇级别具有集体信息价值的参数，并能在设备约束下灵活权衡准确率与通信效率。为支持研究结果的可复现性，我们已在 https://anonymous.4open.science/r/CAAFP-C36E/README.md 发布 CA-AFP 的完整实现。

2 相关工作

基于簇的联邦学习：基于簇的联邦学习（FL）旨在应对统计异质性。例如，Fed- CHAR [20] 根据模型更新方向的余弦相似度进行层次聚类，通过簇内聚合提升鲁棒性——将恶意客户端隔离在外，同时增强参与者之间的公平性 [21]。类似地，ClusterFL [26]将联邦学习表述为一个聚类多任务优化问题，并通过 ADMM 联合学习簇分配和模型参数，从而实现簇内协作学习，同时剔除速度较慢或相关性较弱的客户端，以减少通信轮次并提高准确率 [27]。

近期的聚类联邦学习（CFL）框架通过显式建模梯度方差和部分参与，进一步细化客户端分组。早期的 CFL 方法假设簇内梯度方差较低，从而在非独立同分布数据下实现更快的收敛和更好的泛化能力。在此假设基础上，StoCFL [31]引入了一种随机聚类机制，利用基于梯度的数据分布表示之间的余弦相似度，并支持客户端的部分参与。通过双层优化策略，StoCFL 促进簇特定模型与全局模型之间的知识共享，在保持簇专门化的同时提高数据效率 [31]。作为补充，FedClust [12] 采用一次性、由权重驱动的层次聚类策略，利用部分模型权重（尤其是分类器层）与底层客户端数据分布之间的相关性，实现稳定的簇构建，并显著降低通信开销，同时动态支持新加入的客户端。

尽管这些聚类联邦学习方法能够有效缓解统计异质性，并改善非独立同分布数据下的收敛表现，但它们都依赖完全稠密的模型交换。它们主要关注统计、优化和安全相关挑战，却忽视了内存占用和通信效率。由于缺乏稀疏化、剪枝或压缩机制，这些方法难以应用于资源受限的联邦环境，例如带宽和内存限制严格的边缘设备。

基于剪枝的联邦学习：基于剪枝的联邦学习通过让客户端训练并传输从全局模型派生的子模型，降低联邦优化所需的通信量、计算和内存开销 [2, 9]。一种具有代表性的方法是联邦随机失活 [6], : 客户端在本地训练期间随机丢弃神经元或滤波器，仅通信剩余参数，从而降低资源需求，同时保持全局模型的收敛特性 [5]。为进一步优化训练初期阶段，FedSNIP[30] 采用基于连接敏感度的单次网络剪枝，在通信开始前构建稀疏结构。此外，EfficientFL [16] 通过在训练开始时识别“获胜子网络”，减轻边缘设备的负担，确保在整个联邦循环中仅更新和传输影响最大的参数。这些方法表明，在联邦设置中，模型的部分参与也能发挥作用，尤其适用于资源受限的边缘设备。

自适应联邦剪枝：自适应联邦剪枝通过在训练过程中动态调整模型稀疏度，以应对系统约束、优化进度和设备异质性，从而扩展了基于剪枝的联邦学习 [8]。不同于先训练密集模型、再进行一次性稀疏化的静态剪枝策略，自适应方法将剪枝直接融入联邦优化循环，使设备能够逐步学习稀疏模型，而无需承担密集训练的成本 [10]。近期的框架（如 FLCAP [13]）进一步采用聚类自适应剪枝来应对可扩展系统，而 SAFL [17]则采用结构感知剪枝，使子模型与特定客户端簇相匹配。

FedResCuE [33] 通过使客户端能够训练自蒸馏神经网络来应对系统异质性和不稳定的网络连接。这些网络可以在无需微调的情况下进行结构化剪枝，得到任意大小的模型。通过渐进式学习，预测知识被保留在嵌套子模型中，使资源预算各异的客户端能够使用不同大小的剪枝模型参与训练，同时仍为全局聚合过程贡献有意义的更新量。这种自适应剪枝范式使联邦学习系统能够灵活适应异质设备和部分模型传输。

联邦学习中的簇感知剪枝：聚类感知联邦学习旨在通过将数据分布相似的客户端分组，并将模型聚合限制在每个簇内，以缓解统计异质性 [23]。代表性工作

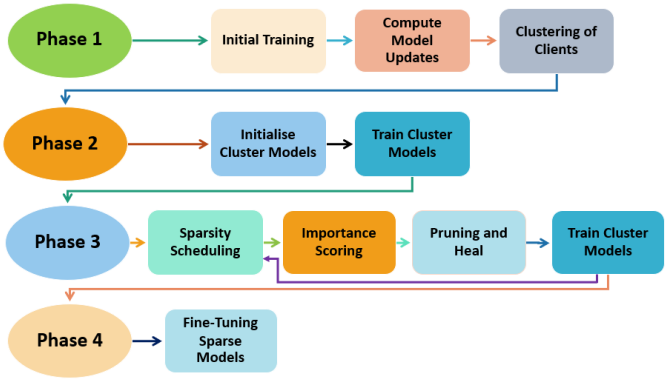

图 1：CA-AFP 的整体工作流程。

例如，ClusterFL将联邦学习表述为聚类多任务优化问题，其中服务器利用分布式优化技术联合学习簇分配与簇特定模型，从而在相似客户端之间实现协作学习，同时通过按簇的掉队者随机失活与节点选择降低通信开销 [26]。

这些基于聚类的方法表明，利用客户端之间的内在相似性，有助于提升联邦学习的收敛性、个性化能力和鲁棒性。然而，现有的聚类感知联邦学习方法主要基于完全稠密模型运行，并侧重于统计异质性、公平性和安全性等方面 [32]。这些方法未在簇内或簇间引入剪枝或稀疏化机制，因此部署在资源受限的边缘设备上时会产生大量内存和通信开销 [20, 26]。因此，尽管聚类能够减少异构客户端之间的有害干扰，但缺乏聚类感知剪枝限制了这些方法的可扩展性和效率。这一观察揭示了一个关键的研究空白：如何在联邦学习中联合利用客户端聚类与自适应剪枝，以同时实现统计效率和系统级资源效率。

3 方法论

CA-AFP 框架包含四个不同的阶段。算法 1（附录）提供了详细的伪代码，图 2 展示了工作流程概览。

3.1 阶段 1：初始训练与聚类

初始训练：训练按通信轮次进行。在每一轮 t，服务器都会向所有客户端广播全局模型 $w^{(t-1)}$。每个客户端 k 在其数据集 D_k 上进行 E 轮本地训练，以最小化标准交叉熵损失：

$$\mathcal{L}_k^{\mathrm{CE}}(w_k;\mathcal{D}_k) = -\frac{1}{|\mathcal{D}_k|}\sum_{(x,y)\in\mathcal{D}_k}\log p(y\mid x), \tag{1}$$

其中，$\mathrm{L}_k^{\mathrm{CE}}(w_k;\mathrm{D}_k)$表示客户端$k$上的交叉熵损失，$w_k$ 表示模型参数，D_k 是客户端 k的本地数据集，包含 $|\mathrm{D}_k|$个样本，(x,y)表示一个输入—标签对，而$p(y\mid x)$表示给定输入 x时标签 y的预测概率。

完成 E 个本地训练轮数后，客户端 k 会获得更新后的参数 $w_k^{(t)}$，并将其发送给服务器。服务器对客户端模型进行聚合，得到

$$w^{(t)} = \sum_{k=1}^{K}\frac{|\mathcal{D}_k|}{|\mathcal{D}|}w_k^{(t)}, \tag{2}$$

其中，$w^{(t)}$ 表示通信轮次 t时聚合后的全局模型参数，K 表示参与的客户端总数，$|\mathrm{D}| = \sum_{k=1}^{K}|\mathrm{D}_k|$ 表示所有客户端的训练样本总数。初始训练阶段进行 T_0 个通信轮次，其中 $T_0 \geq 0$。

聚类：经过 T_0 轮初始训练后，最终全局模型 $w_{(T_0)}$ 将用作客户端聚类的参考模型。如果 $T_0 = 0$，$w^{(T_0)}$对应一个随机初始化模型。

每个客户端 k使用 $w^{(T_0)}$初始化其模型，并在 D_k上进行一轮本地训练，生成更新后的参数 $w_{(T_0,\,1)k}$。用于聚类的本地更新定义为

$$\Delta w_k = w_k^{(T_0,1)} - w^{(T_0)}. \tag{3}$$

客户端 k 与 k' 之间的成对余弦距离计算如下：

$$d(k,k') = 1 - \frac{(\Delta w_k)^{\top}\Delta w_{k'}}{\|\Delta w_k\|_2\,\|\Delta w_{k'}\|_2}. \tag{4}$$

此处，Δw_k 和 $\Delta w_{k'}$ 分别表示客户端 k和 k'的单轮本地更新向量。运算符 $(\cdot)^{\top}$ 表示向量转置，$(\Delta w_k)^{\top}\Delta w_{k'}$ 表示它们的内积，$\|\cdot\|_2$ 表示欧几里得（L_2）范数。

该距离衡量客户端更新之间的角度不相似性，同时不受其幅值影响。随后，对所得距离矩阵应用采用平均链接的凝聚层次聚类，将每个客户端分配到一个簇中。聚类过程仅确定客户端分配，不会修改全局模型权重。

3.2 阶段 2：簇级稳定化

客户端聚类完成后，CA-AFP 使用阶段 1 的全局模型 $w_{(T_0)}$为每个簇初始化一个簇模型。随后，在各自对应的客户端组内，通过若干轮“密集联邦训练” [21] 对每个簇模型进行优化，此阶段不进行任何剪枝。

具体而言，对于每个簇 c，令 C_c 表示分配到簇 c的客户端集合。每轮 t结束后，服务器将当前簇模型 $w_c^{(t)}$广播给所有客户端 $k \in \mathrm{C}_c$。在每个簇内，每个客户端维护两个模型：个人参考模型$w_{p,\,k}$和簇模型 w_c。

每个客户端通过最小化正则化目标，在其数据集 D_k 上进行本地训练

$$\mathcal{L}_k^{c}(w_c;\mathcal{D}_k) = \mathcal{L}_k^{\mathrm{CE}}(w_c;\mathcal{D}_k) + \frac{\lambda}{2}\|w_c - w_{p,k}\|^2, \tag{5}$$

其中，$\mathrm{L}_k^{c}(w_c;\mathrm{D}_k)$ 表示簇 k 中客户端 c 的总训练损失，$\mathrm{L}_k^{\mathrm{CE}}(w_c;\mathrm{D}_k)$ 表示在客户端 k 的本地数据集上计算的标准交叉熵损失。参数 w_c 表示

簇模型参数，且$w_{p,\,k}$表示客户端 k 的个人参考模型，该模型由阶段 1 的全局模型 $w^{(T_0)}$ 初始化，并在阶段 2 和阶段 3 期间保持固定。

尽管所有参考模型共享相同的初始权重，但保留各自独立的副本有助于后续实现个性化。正则化项可防止簇模型偏离全局表示过多，同时允许其适应簇级数据。超参数 λ 控制该正则化的强度。

本地训练后，客户端将更新后的簇模型$w_{c\,k}^{(t)}$，返回服务器。随后，使用按数据量加权的聚合来更新簇模型：

$$w_c^{(t+1)} = \sum_{k\in C_c} \frac{|\mathcal{D}_k|}{|\mathcal{D}_c|} w_{c,k}^{(t)}, \tag{6}$$

其中，$|\mathrm{D}_c| = \sum_{k\in C_c} |\mathrm{D}_k|$表示簇 c中的样本总数。簇级稳定阶段持续进行 T_1 个通信轮次，其中 $T_1 \geq 0$。当 $T_1 = 0$时，不再进一步优化簇模型，而是直接将其传递到后续剪枝阶段。

3.3 阶段 3：带剪枝的簇训练

阶段 3 继续优化阶段 2 中由式 (5) 定义的同一正则化簇目标，同时将剪枝引入簇模型。

簇模型 w_c M_c 被施加二值剪枝掩码[22]。该步骤确保剪枝权重在全局簇状态中保持为零：

$$w_c^{(t+1)} = \left(\sum_{k\in C_c} \frac{|\mathcal{D}_k|}{|\mathcal{D}_c|} w_{c,k}^{(t)} \right) \odot M_c^{(t)} \tag{7}$$

为确定在二值掩码下哪些参数应被保留或剪枝，同时保留簇级知识，我们引入簇感知重要性评分。

重要性评分：令$w_{c,\,j}$表示簇模型 w_c的第 j个标量权重参数，并令W_c表示 w_c的全部标量权重参数集合。CA-AFP 整合簇级信号，通过重要性分数指导剪枝：

$$\mathrm{Score}_{w_{c,j}} = \alpha \cdot \mathrm{Mag}_{w_{c,j}} + \beta \cdot \mathrm{Coh}_{w_{c,j}} + \gamma \cdot \mathrm{Con}_{w_{c,j}}, \tag{8}$$

其中 α、 β和 γ为满足 $\alpha+\beta+\gamma = 1$的加权系数。

幅值分数。权重的归一化绝对幅值：

$$\mathrm{Mag}_{w_{c,j}} = \frac{|w_{c,j}|}{\max_{u\in W_c} |u|} \tag{9}$$

一致性得分。 衡量一个权重在簇内跨客户端的稳定性：

$$\mathrm{Coh}_{w_{c,j}} = \frac{1}{1 + \mathrm{Var}_{k\in C_c}(w_{c,j}^{(k)})} \tag{10}$$

其中 $w_c^{(k)},_j$表示客户端 k 的本地训练的簇模型中权重 $w_{c,j}$的值。较低的方差表明各客户端之间具有更高的结构一致性。

一致性得分。 衡量簇内客户端之间的梯度方向一致性：

$$\mathrm{Con}_{w_{c,j}} = \left| \frac{1}{|C_c|} \sum_{k\in C_c} \mathrm{sign}\left(\frac{\partial \mathcal{L}_k}{\partial w_{c,j}} \right) \right| \tag{11}$$

其中，L_k 表示客户端 k的局部目标函数。

此处，C_c 表示属于簇 c的客户端集合。重要性分数倾向于选择幅值较大、在各客户端间稳定且始终得到梯度方向支持的参数，从而在剪枝过程中同时保留结构重要性和学习动态。

随着轮数推进，CA-AFP会动态调整二值剪枝掩码，以利用这些轮次中趋于稳定的权重更新，并改进对模型权重进行剪枝或保留的决策。为此，它采用下文介绍的剪枝与修复机制。

剪枝与修复机制。在簇训练期间，该机制按照剪枝计划定期执行，并遵循迭代式的剪枝–修复循环。在每个剪枝步骤中，CA-AFP执行以下流程：

(1) **重要性重新评估：** 利用幅值、一致性和一致性信号重新计算重要性分数，以反映当前的训练动态。(2) **计划剪枝：** 设 $S^{(t)}$ 表示当前稀疏度，即在迭代 t时簇模型中被剪枝（值为零）的权重所占比例。根据当前稀疏度 $S^{(t)}$ 和目标稀疏度 S_{target}，移除排名最低的一部分权重。剪枝量会根据剩余剪枝步骤和用户定义的剪枝率自动调整，以平稳地达到目标稀疏度。(3) **梯度引导再生长：** 为保留模型在训练期间的适应能力，重新激活一部分梯度幅值最高的被剪枝连接，以进行结构探索并防止信息不可逆地丢失。(4) **修复阶段：** 随后按照既定计划对掩码稀疏模型进行若干轮训练，使剩余参数能够适应并进行补偿，然后再进入下一次剪枝步骤。

3.4 阶段4：客户端微调

剪枝过程完成后，每个客户端执行本地微调阶段，以恢复剪枝所引入的潜在精度损失。在此阶段，最终剪枝掩码 M_c 保持固定，不允许进一步的结构更新。

令 $\tilde{w}_k$ 表示客户端 k经本地微调后的稀疏模型，该模型由其在阶段 3 之后的最终簇模型初始化。每个客户端在稀疏性约束下优化 $\tilde{w}_k$：

$$\min_{\tilde{w}_k} \mathcal{L}_k^{\mathrm{CE}}(\tilde{w}_k; \mathcal{D}_k) \quad \text{s.t.} \quad \tilde{w}_k \odot M_c = \tilde{w}_k. \tag{12}$$

此约束确保仅更新剪枝掩码 M_c 中非零项对应的参数，而被剪枝的参数则固定为零。微调

表 1：WISDM 和 UCI-HAR 的异质性分析。

异质性类型	指标	WISDM	UCI HAR
数量偏斜	变异系数（$CV = \frac{\sigma}{\mu} \times 100\%$）	30.4%	10.2%
	样本数量（$\mu \pm \sigma$）	297 ± 90	343 ± 35
标签偏斜	缺失类别的客户端	17 / 36	0 / 30
	缺失率	47.2%	0.0%
特征偏斜	客户端间平均变异系数	6.5%	0.6%

会进行少量本地训练轮数，且不再进行通信或聚合，从而使每个客户端能够根据其本地数据分布适应稀疏模型，同时保持最终的稀疏度。

4 实验设置

4.1 数据集与数据分布

我们在两个广泛使用的人体活动识别（HAR）基准数据集上评估 CA-AFP：**WISDM**和**UCI-HAR**。选择这些数据集，是因为它们是联邦和个性化 HAR 研究中的标准基准， [1], 表现出显著的用户间异质性，且传感条件和数据采集环境各不相同，从而能够全面评估 CA-AFP。

WISDM 数据集：WISDM（无线传感器数据挖掘）活动识别数据集 [15] 包含从 36 名用户采集的 1,098,207 条原始智能手机加速度计记录，用户执行的活动包括六种：步行、慢跑、上楼、下楼、坐和站立。原始传感器流按长度为 200 个时间步、重叠率为 50% 的滑动窗口进行分段，生成 10,591 个形状为 (200,3) 的样本。每位用户都被视为一个独立的联邦客户端，从而形成自然的、基于用户的数据划分，并保留因运动模式、活动频率和设备使用方式差异而产生的真实世界异质性。

我们在表 1 中总结的这种自然划分，天然导致客户端之间存在非独立同分布的数据分布，包括特征偏斜（源于个体运动 dynamics）、标签偏斜（不同用户参与活动的程度不均）以及数量偏斜（每位用户的样本数量不同），且无需引入人为的数据操纵。因此，WISDM 可作为一个真实、含噪且不受约束的联邦学习基准。

UCI-HAR 数据集： UCI-HAR 数据集 [7] 包含 30 名受试者在受控实验室条件下完成六种活动时采集的惯性传感器数据，数据通过佩戴在腰部的智能手机收集。该数据集共包含 10,299 个带标签的样本，每个样本表示为形状为 (128,9) 的多元时间序列。与 WISDM 类似，我们采用自然的联邦设置，将每位受试者对应为一个客户端。与 WISDM 相比，UCI-HAR 的传感器噪声更低，且活动执行更加一致（表 1）。

WISDM 和 UCI-HAR 涵盖了多样的部署条件，包括真实环境与实验室环境、不同的窗口长度和传感器维度，以及程度各异的客户端间异质性。这种组合能够在真实且受控的联邦学习场景下，对 CA-AFP 进行均衡评估。

4.2 基线方法

我们将 CA-AFP 与四种代表性基线方法进行比较，这些方法通过客户端聚类或模型剪枝来应对联邦学习中的异质性。FedCHAR 和 ClusterFL 着重于通过聚类缓解统计异质性并提升客户端层面的性能，而 EfficientFL 和 FedSNIP 则采用剪枝来提高通信效率和设备兼容性。除了减少通信开销，剪枝还具有正则化作用，有助于缓解过拟合和客户端漂移 [11]。我们将不进行剪枝的方法称为稠密基线 （FedCHAR、ClusterFL），将进行剪枝的方法称为剪枝基线方法 （EfficientFL、FedSNIP）。

FedCHAR [21] 是一种基于聚类的个性化联邦学习方法，不采用剪枝，可作为面向准确率的上界。**ClusterFL** [27] 通过交替优化联合优化客户端模型与簇分配，以缓解统计异质性。**EfficientFL** [29]通过稀疏化和量化压缩客户端更新量，在保持收敛的同时降低通信成本。**FedSNIP** [4]在客户端应用一次性基于SNIP的剪枝，仅传输关键参数。

近期方法如 SAFL 和 FLCAP 同样结合了聚类与剪枝，但基于不同的架构与算法假设。SAFL 依赖基于批归一化的网络瘦身，这与我们不含 BN 的架构不兼容。FLCAP 从剪枝掩码动态而非模型更新相似度中提取聚类信号。为确保公平且受控的比较，我们将评估范围限定为与我们的模型及训练直接兼容的基线方法。

4.3 实验设置

我们描述实验配置以确保可复现性。更多细节见附录（表5–10）。

实现与硬件。 所有实验均在配备英特尔酷睿 i9-14900K 处理器的 Linux 工作站上进行。CA-AFP 及所有基线方法均使用 TensorFlow 2.14 实现，并采用固定随机种子进行权重初始化和客户端数据划分。除非另有说明，我们使用学习率为 10^{-3} 的 Adam。目标稀疏度固定为 70%，所有方法均训练 50 个全局轮次。表2中报告的结果为五个随机种子的平均值。

阶段敏感性分析。 为分离不同训练阶段的影响，我们在 WISDM 数据集上进行了敏感性分析。我们将初始训练阶段（P_1）和聚类后阶段（P_2）的持续时间设为 0 至 20 轮次，并将初始稀疏度（S_{start}）设为 0.0 至 0.9（表7）。剪枝阶段固定为 $P_3 = 40$ 轮次，微调设为 $P_4 = 3$ 个训练轮次。结果取三个随机种子的平均值。

重要性分数消融。 我们在 UCI-HAR 上评估重要性评分函数，分析幅值（α）、一致性（β）和一致性（γ）权重的七种组合。剪枝计划在 15 轮内从 30% 变化至 70%，以考察各组件的贡献。各项配置见表 8。所有实验均使用单个随机种子进行。

表2：在 WISDM 和 UCI-HAR 数据集上进行 3 和 25 个训练轮次微调（FT）的性能比较。

方法	稀疏度	微调训练轮次	WISDM 平均准确率 (μ) (↑)	WISDM 公平性 (σ) (↓)	WISDM 通信量 (MB)	UCI-HAR 平均准确率 (μ) (↑)	UCI-HAR 公平性 (σ) (↓)	UCI-HAR 通信量 (MB)
密集基线								
FedCHAR [21]	0%	3	96.64%	8.39%	483	96.87%	6.70%	347
	0%	25	97.16%	8.32%	483	97.48%	6.19%	347
ClusterFL [27]	0%	3	96.18%	7.02%	474	95.80%	6.87%	340
	0%	25	96.33%	6.50%	474	95.80%	6.87%	340
剪枝基线方法								
EfficientFL [29]	≈ 70%	3	92.30%	10.29%	142	95.02%	9.46%	102
	≈ 70%	25	95.68%	9.28%	142	97.49%	5.85%	102
FedSnip [4]	≈ 70%	3	94.29%	10.16%	142	95.56%	9.34%	102
	≈ 70%	25	95.24%	8.96%	142	95.82%	9.34%	102
Ours								
CA-AFP (0,0,50)	≈ 70%	3	95.56%	8.97%	176	95.99%	6.91%	122
	≈ 70%	25	95.62%	8.84%	176	97.13%	6.55%	122
CA-AFP (15,15,20)	≈ 70%	3	95.28%	9.50%	375	95.91%	6.49%	266
	≈ 70%	25	96.73%	8.35%	375	97.40%	6.40%	266

非独立同分布鲁棒性评估。 我们在三种统计异质性水平下评估 WISDM 上的所有方法（表10）。每个客户端最多仅限于 k 个活动类别：极端 （$k = 1$）、中度 （$k = 2$）和轻度 （$k = 3$）。从每个客户端的真实标签中随机抽取允许的类别，并丢弃其他类别的数据。所有实验均运行 50 个全局轮次，目标稀疏度为 $S = 0.7$，结果以五个随机种子的均值 ± 标准差报告。

微调设置。 微调配置汇总于表9中。

模型架构。 所有方法都使用相同的轻量级一维卷积神经网络进行人体活动识别。该网络由两个卷积块组成，每个卷积块包含一个具有 64 个滤波器、卷积核大小为 5 的一维卷积层，后接 ReLU 激活函数、最大池化和随机失活（比率为 0.3）。随后将得到的特征展平，并输入一个包含 32 个单元且采用 ReLU 激活函数的全连接层，之后接随机失活（比率为 0.2）和 Softmax 输出层。

4.4 准确率指标

为了与现有基线方法进行比较并分析消融实验，我们采用三个指标：准确率、公平性（标准差）和通信成本（以 MB 为单位）。平均准确率是所有客户端测试准确率的均值（μ）。我们用“公平性”一词指所有客户端平均准确率的标准差（σ）。公平性值越低，不同客户端之间的测试准确率差异越小，这表明模型在所有客户端上的表现较为均衡；反之，公平性值越高，不同客户端之间的测试准确率差异越大，这表明模型在所有客户端上的表现不够可靠，因为其性能波动较大。通信成本用于衡量整个训练过程中服务器与客户端之间传输的数据总量，以

表3：“重要性分数”最佳配置。

场景	数据集	α	β	γ	平均准确率 (%)	公平性 (%)
标准	UCI HAR	0.25	0.25	0.50	91.19	9.81
	WISDM	0.25	0.25	0.50	91.57	10.51
漂移	UCI HAR	0.25	0.25	0.50	66.54	32.70
	WISDM	0.00	0.00	1.00	57.63	33.12
噪声客户端	UCI HAR	0.50	0.25	0.25	74.42	19.03
	WISDM	0.25	0.25	0.50	74.36	20.72

兆字节（MB）为单位。该指标涵盖客户端到服务器以及服务器到客户端的传输，体现了稀疏度带来的效率提升。C_{total} 轮次内的总成本估算如下：

$$C_{total} = \sum_{r=1}^{R} \sum_{k \in \mathcal{S}_r} 2 \times |W| \times (1 - S_k^{(r)}) \times \frac{4}{1024^2} \tag{13}$$

其中，$\mathcal{S}_r$ 是第 r 轮次中选定客户端的集合，$|W|$ 是稠密模型中的参数总数（我们将其视为 32 位浮点数，即 4 字节），而 $S_k^{(r)}$ 是客户端 k 的模型在第 r 轮次中的稀疏度。

5 结果

表2展示了两种数据集上不同基线方法与 CA-AFP 的性能对比。从表中可以看出，准确率与通信成本之间存在权衡。FedCHAR 和 ClusterFL 等稠密方法准确率较高，但由于需要进行全模型通信，通信开销也最高。相比之下，基于剪枝的方法通过传输稀疏更新显著降低了通信成本。随着准确率提高，客户端之间的性能通常更加公平，但某些情况下并未呈现这种关系。CA-AFP 在.....之间实现了更好的平衡

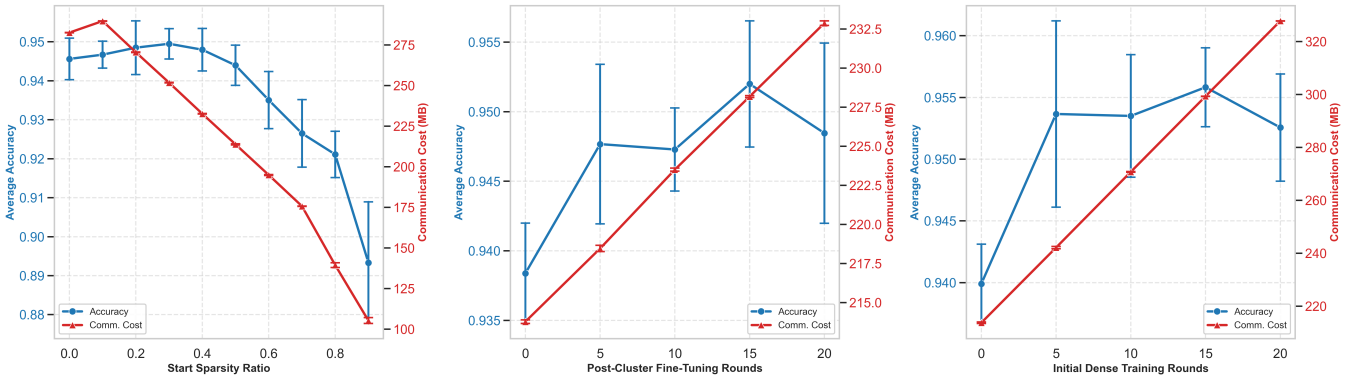

图 2：“初始稀疏度”和“轮数”对 WISDM 中准确率的影响。

准确率、效率和公平性。其表现始终优于或不逊于剪枝基线；在（15, 15, 20）轮数配置下，虽然通信成本更低，但其表现也能与稠密基线相媲美。采用（0, 0, 50）轮数配置的 CA-AFP 通信成本略高于其他剪枝基线，因为它在剪枝的同时还以预设速率重新生长权重。不过，对于给定的数据集和模型大小，额外成本保持不变。

综合来看，这些结果表明，CA-AFP 同时追求以下目标：高准确率、低通信成本以及提升的客户端间公平性。它还提供了灵活性，可通过调整不同阶段之间的轮次分配来利用通信预算，从而在模型部署时达到给定的目标稀疏度水平。

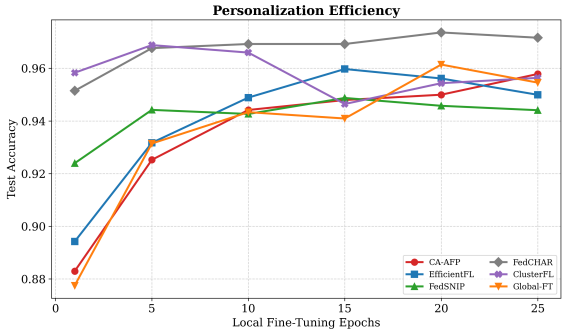

图 3：本地微调对 WISDM 准确率的影响。

5.1 消融与敏感性结果

5.1.1 重要性评分权重。 我们开展消融研究，以分析重要性评分权重（α，β，γ）在四种异质数据场景下的影响：标准非独立同分布、噪声客户端和漂移。

场景 1：标准（自然异质性）。此场景作为基线，采用 UCI HAR 数据集的自然划分方式。数据的分配方式是：每个联邦学习客户端严格对应原始研究中的一名人类受试者，从而在不进行合成修改的情况下保留固有的用户特定差异。

场景 2：噪声客户端（部分标签污染）。此场景模拟不可靠的数据标注。随机选取 40% 的客户端作为噪声参与者。对于这些客户端，将其 30% 的标签替换为从均匀离散分布 $U(0, 5)$ 中抽取的随机整数，而特征向量 X 保持不变。

场景 3：漂移（传感器故障）。此场景模拟传感器失灵或设备损坏。在此设置中，选取 40% 的客户端贡献纯噪声。对于这些参与者，将 100% 的标签替换为随机值，从而有效破坏特征与标签之间的映射关系（$X \rightarrow Y_{random}$）。这项压力测试用于评估重要性评分机制识别并抑制来自故障节点的更新量的能力。

各场景下表现最佳的配置汇总于表 3。总体而言，这些结果表明，没有任何单一重要性分数组件能在所有环境中占据主导地位。相反，混合公式使 CA-AFP 能够灵活适应不同类型的数据及其变化，调整剪枝行为。这验证了我们采用加权混合重要性分数，而非依赖任何单一剪枝准则的设计选择。

所有权重组合和场景的完整消融实验结果见附录中的表 4。我们将表 4 中准确率与公平性比值最高的组合称为“得分”；准确率越高、公平性值越低越好。

5.1.2 轮数和初始稀疏度的影响。在本研究中，我们通过实验分别考察并理解初始训练（阶段 1）的轮数、聚类后但剪枝前（阶段 2）的轮数，以及剪枝和修复（阶段 3）的初始稀疏度对通信成本和各客户端平均准确率的影响。根据结果（见图 2），可以发现，准确率的提高总是伴随着通信成本的增加。降低初始稀疏度，以及增加阶段 1 的初始训练轮数或阶段 2 的轮数，都能有效提高准确率，但同时也会增加通信成本。我们还发现，当初始稀疏度超过某个值（本例中为 0.5）后，准确率会急剧下降，而通信成本则几乎以恒定速率降低。因此，找到合适的初始稀疏度值，有助于以有限的成本获得更好的结果。

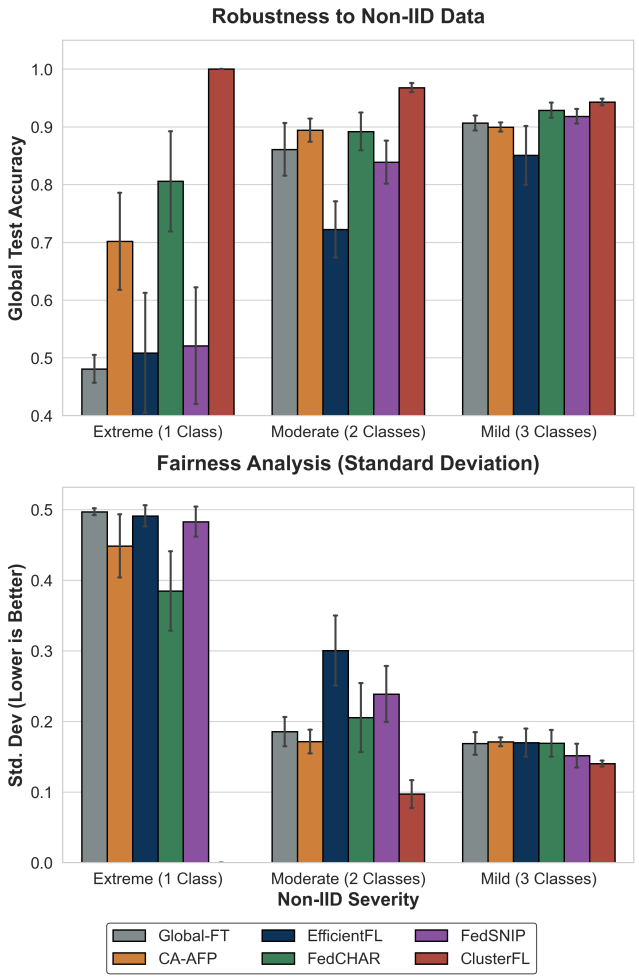

图 4：三种非独立同分布程度下的性能。

5.1.3 本地微调的影响。在本实验中，我们分析本地微调轮数（阶段 4）对最终“准确率”的影响。由于微调是在本地对稀疏模型进行的，因此不会产生额外的通信开销。

如图 3 所示，将本地微调轮数从 0 增加到 25，可使 CA-AFP 的准确率从约 88% 提升至 96%，同时通信成本保持在约 176 MB 不变。这表明 CA-AFP 保留的参数仍具有很强的适应能力，使客户端能够通过本地微调恢复性能，而无需增加通信开销。

图3中的*Global-FT* 基线对应于仅包含一个簇的 CA-AFP。在微调初期，其准确率始终低于聚类式 CA-AFP。这表明聚类有助于更早地学习个性化表示。此外，采用 10–25 轮本地微调能够有效缩小基于聚类的数据划分所造成的性能差距，从而实现稳定且鲁棒的收敛。表2中也观察到一致的趋势：在各种实验设置下，微调均能提升 CA-AFP 的性能。

5.1.4 非独立同分布水平下的鲁棒性。我们评估了 CA-AFP 和基线方法在三种统计异质性水平下的鲁棒性：轻度（每个客户端 3 个类别）、中度（每个客户端 2 个类别）和极端（每个客户端 1 个类别）。

如图 4 所示，在中度和极端非独立同分布设置下，基于剪枝的基线方法的准确率显著低于基于簇的稠密方法和 CA-AFP。尤其是在极端的每个客户端仅含 1 个类别的设置下，剪枝基线（FedSNIP 和 EfficientFL）的准确率降至约 50%，这主要是由于异构客户端之间存在严重的梯度冲突。

由于本实验未进行微调，这些剪枝方法无法适应客户端特定分布，因为它们依赖单个全局模型，且未进行聚类。相比之下，CA-AFP 通过将相互冲突的目标隔离到独立簇中缓解了这一问题，使准确率维持在约 80

在中度（2 类）非 IID 设置下也观察到类似趋势：与基于聚类的方法和 CA-AFP 相比，剪枝基线的表现仍然较差。此外，不采用聚类的 CA-AFP 变体（*Global-FT*）在极端异质性下表现不佳，但在中度异质性下准确率有所恢复，这凸显了聚类在处理严重客户端级数据多样性方面的关键作用。

6 局限性与未来工作

聚类策略。 在本研究中，我们采用固定的聚类策略，以平衡有效性和计算效率。尽管先前研究已探索了其他聚类机制 [12, 21], , 但系统分析不同聚类指标、调度方案和集成节点如何与自适应剪枝相互作用，仍是未来值得探索的方向。

超参数敏感性。 CA-AFP 依赖少量超参数，包括重要性评分权重和全局稀疏度目标（$S_{\text{target}} = 70\%$）。我们的消融研究表明，该方法在广泛的设置范围内都能保持稳定性能。尽管如此，开发能够根据底层数据分布调整这些参数的自适应机制，或可进一步减少调参工作，并提升其在大规模部署中的易用性。

7 结论

在本研究中，我们提出了 CA-AFP，这是一种统一框架，将基于聚类的个性化与自适应模型剪枝相结合，用于异构条件下的联邦学习。通过两项关键创新——结合权重幅值、簇内一致性和梯度一致性的簇感知重要性评分机制，以及迭代式自修复剪枝计划——CA-AFP 在预测准确率、客户端间公平性和通信效率之间实现了良好平衡。

大量实验结果表明，在极端非独立同分布设置下，CA-AFP 仍能保持鲁棒性，并通过可控的剪枝和微调方案，灵活适应不同的通信预算。CA-AFP 根据簇级共识而非孤立的客户端统计数据来优先选择参数，使联邦学习更适合部署在资源受限的边缘设备上。总体而言，我们的工作凸显了联合优化个性化与资源效率的潜力，有助于构建可扩展且可靠的联邦系统。

参考文献

[1] Aouedi, O., Sacco, A., Khan, L. U., Nguyen, D. C., and Guizani, M. 面向人体活动识别的联邦学习：概述、进展与挑战。《IEEE通信学会开放期刊》5（2024年），5687–5712。[2] Bai, G., Li, Y., Li, Z., Zhao, L., and Kim, K. Fedspallm：大语言模型的联邦剪枝。载于《2025年美洲国家计算语言学协会分会会议论文集：人类语言技术》（第1卷：长篇论文）（阿尔伯克基，新墨西哥州，2025年），计算语言学协会，第8361–8373页。[3] Bonawitz, K., Eichner, H., Grieskamp, W., Huba, D., Ingerman, A., Ivanov, V., Kiddon, C., Konečný, J., Mazzocchi, S., McMahan, H. B., Overveldt, T. V., Petrou, D., Ramage, D., and Roselander, J. 迈向大规模联邦学习：系统设计。《第2届SysML会议（MLSys）论文集》1（2019），374–388。[4] Bustincio, R. W. C., de Souza, A. M., da Costa, J. B. D., Gonzalez, L. F. G., and Bittencourt, L. F. 通过联邦学习中的单次模型剪枝降低通信开销。《电信年刊》80（2025）。[5] Caldas, S., Wu, P., Li, T., Konečný, J., McMahan, H. B., and Talwalkar, A. FedDropout：采用结构化丢弃的联邦学习。载于《IEEE国际分布式计算系统会议（ICDCS）论文集》（2019），第1–10页。[6] Cheng, G., Charles, Z., Garrett, Z., and Rush, K. 联邦随机失活真的有效吗？载于《IEEE/CVF计算机视觉与模式识别会议（CVPR）研讨会论文集》（2022），第3386–3394页。[7] Garcia-Gonzalez, D., Rivero, D., Fernandez-Blanco, E., and Luaces, M. R. 一个用于利用智能手机传感器进行真实场景人体活动识别的公有领域数据集。《传感器》20，8（2020），2200。[8] Huang, H., Yang, H., Chen, Y., Ye, J., and Wu, D. 去中心化联邦学习中的自适应模型剪枝。《IEEE网络科学与工程汇刊》PP，99（2025），1–17。提前在线发表。[9] Huang, H., Yang, H., Chen, Y., Ye, J., and Wu, D. FedRTS：通过组合汤普森采样实现联邦鲁棒剪枝。arXiv预印本 arXiv:2501.19122（2025）。已被NeurIPS 2025接收。[10] Huang, H., Zhuang, W., Chen, C., and Lyu, L. FedMEF：迈向节省内存的联邦动态剪枝。载于《IEEE/CVF计算机视觉与模式识别会议论文集》（2024），IEEE，第27538–27547页。[11] Huang, L., Yin, Y., Fu, Z., Zhang, S., Deng, H., and Liu, D. FedMask：通过异构掩码实现计算与通信高效的联邦学习。载于《第20届ACM嵌入式网络传感器系统会议（SenSys '22）论文集》（2022），ACM，第1–14页。[12] Islam, M. A., Rahman, M. M., and Saha, A. FedClust：通过基于权重的相似度实现高效的联邦学习客户端聚类。《IEEE人工智能汇刊》（2024）。[13] Kapoor, A., and Kumar, D. FLCAP：面向异构且可扩展系统的聚类自适应剪枝联邦学习。载于《IEEE/CVF计算机视觉与模式识别会议（CVPR）研讨会论文集》（2023），第3386–3395页。[14] Kim, T., Park, J., and Lee, S. 面向物联网应用的无线传感器网络：全面综述。《传感器》21，6（2021），1–28。[15] Kwapisz, J. R., Weiss, G. M., and Moore, S. A. 使用手机加速度计进行活动识别。《ACM SIGKDD探索通讯》12，2（2011），74–82。[16] Li, A., Sun, J., Wang, B., Duan, L., Li, S., Chen, Y., and Li, H. LotteryFL：基于彩票假设的个性化、高通信效率联邦学习。《IEEE并行与分布式系统汇刊》33，4（2022），975–988。基于彩票假设的EfficientFL架构的核心参考文献。[17] Li, N., Wang, X., Du, X., Wang, C., Cai, P., and Wang, T. SAFL：通过客户端专属聚类和SCSI引导的模型剪枝实现结构感知个性化联邦学习。载于《IEEE第31届并行与分布式系统国际会议（ICPADS）论文集》（2025），IEEE。[18] Li, T., Hu, S., Beirami, A., and Smith, V. Ditto：通过个性化实现公平且鲁棒的联邦学习。《第38届国际机器学习会议（ICML）论文集》139（2021），6357–6368。[19] Li, T., Sahu, A. K., Zaheer, M., Sanjabi, M., Talwalkar, A., and Smith, V. 异构网络中的联邦优化。《机器学习与系统论文集》2（2020），429–450。[20] Li, X., Zhang, Y., Wang, J., and Chen, K. FedCHAR：表征联邦学习中的通信异构性。《IEEE移动计算汇刊》（2023）。[21] Li, Y., Wang, X., and An, L. 基于层次聚类的个性化联邦学习，用于鲁棒且公平的人体活动识别。《ACM智能系统与技术汇刊》（2023）。[22] Lin, R., et al. 利用联邦学习提升神经网络效率。载于Interspeech（2022）。[23] Lu, Y., Yin, H., et al. 面向6G边缘智能交通的自适应剪枝聚类联邦学习。《IEEE智能交通系统汇刊》26，10（2025），17790–17799。发表于2025年2月21日。

[24] McMahan, B.、Moore, E.、Ramage, D.、Hampson, S.与y Arcas, B. A. 去中心化数据上的深度网络通信高效学习。载于第*20*届人工智能与统计国际会议论文集（2017年），PMLR，第1273–1282页。[25] Miralles, F.、Fernandez-Marqes, J.与Lane, N. D. FLCAP：基于聚类聚合与个性化剪枝的联邦学习。*IEEE*国际分布式计算系统会议（*ICDCS*）论文集（2023年），第456–467页。[26] Ouyang, W.、Liu, Y.与Han, Z. ClusterFL：面向人体活动识别的相似性感知联邦学习系统。载于*ACM*嵌入式网络传感器系统会议（*SenSys*）论文集（2021年），第1–14页。[27] Ouyang, X.、Xie, Z.、Zhou, J.、Xing, G.与Huang, J. ClusterFL：面向人体活动识别的基于聚类的联邦学习系统。*ACM*传感器网络汇刊*19*，第1期（2023年），17:1–17:32。[28] Pandolfo, S. 高频数据的波动率分析：已实现波动率与HAR模型。硕士论文，威尼斯卡福斯卡里大学，2020年。导师：Stefano Federico Tonellato教授。[29] Wu, T.、Song, C.与Zeng, P. 基于模型剪枝的资源受限边缘设备高效联邦学习。《复杂与智能系统》*9*（2023年），第6999–7013页。[30] Xu, J.、Li, T.、Sahu, A. K.、Talwalkar, A.与Smith, V. FedSNIP：通过稀疏神经网络训练实现通信高效的联邦学习。载于第*38*届国际机器学习会议（*ICML*）论文集（2021年），第11435–11444页。[31] Zeng, Y.、Liu, Z.、Wang, J.与Chen, K. StoCFL：面向非独立同分布数据的随机聚类联邦学习。*IEEE*神经网络与学习系统汇刊（2025年）。[32] Zhang, Y.、Chen, H.、Lin, Z.、Chen, Z.与Zhao, J. Fedac：面向异质数据的自适应聚类联邦学习框架。*arXiv*预印本*arXiv:2403.16460*（2024年）。[33] Zhu, L.、Liu, X.与Wang, J. FedResCuE：统计异质性条件下带救援机制的联邦学习。载于*AAAI*人工智能会议论文集（2022年），第36卷，第4087–4095页。

8 附录

8.1 算法

在本节中，我们已通过算法1中所述的伪代码对CA-AFP 框架进行了说明。

8.2 分数系数的完整消融实验结果

表 4 报告了所有评估过的权重配置和数据集场景下重要性分数的完整消融实验结果。

8.3 实验设置

本节介绍实验中使用的配置，详见表 5–10。

表 5：UCI HAR 实验的超参数配置

方法	参数	值
通用设置	分阶段轮数 (P_1, P_2, P_3)	0, 0, 50
	本地训练轮数 (E)	3
	批大小 (B)	32
	优化器	Adam ($lr = 1e^{-3}$)
	微调轮数 (P_4)	3
CA-AFP（我们的方法）	目标稀疏度	自适应（开始 70%）
	评分权重 (α， β， γ)	0.5, 0.25, 0.25
	掩码更新率（更替率）	5%
	簇数量（K）	3
FedSNIP	目标稀疏度	70%
EfficientFL	目标稀疏度	70%
FedCHAR	簇（K）	3
	预热轮次	10
ClusterFL	簇（K）	3
	正则化（ρ）	$5e^{-3}$
	双层参数 (α, β)	$1e^{-3}$, $5e^{-4}$

算法 1 CA-AFP：簇感知自适应联邦剪枝 要求：客户端 K、簇 C、预热轮次 $T_1 \geq 0$、稳定化轮次 $T_2 \geq 0$、剪枝轮次 T_3、剪枝频率 f、目标稀疏度 S_{target}、剪枝率 ρ确保：个性化模型 $\{v_k\}$ 1：初始化全局模型 $w^{(0)}$2：阶段 1：预热训练 3：如果 $T_1 > 0$ 则 4：对于 $t = 1$ 到 T_1 执行 5：服务器广播 $w^{(t)}$6：对于所有客户端 k 执行 7：客户端进行本地训练并返回 $w_k^{(t)}$ 8：结束循环 9：$w^{(t+1)} \leftarrow \sum_k \frac{|D_k|}{|D|} w_k^{(t)}$ 10：结束循环 11：结束条件 12：客户端聚类（始终执行）13：令 $w^{\text{ref}} \leftarrow w^{(T_1)}$ ▷ $w^{(0)}$ 如果 $T_1 = 0$14：对于所有客户端 k 执行 15：客户端从 w^{ref}开始进行一个本地训练轮次 16：计算更新量 Δw_k17：结束循环 18：服务器将 $\{\Delta w_k\}$ 聚类为 C 个簇 19：对于所有簇 c 执行 20：$w_c^{(0)} \leftarrow w^{\text{ref}}$21：$M_c^{(0)} \leftarrow 1$22：结束循环 23：阶段 2：稠密簇稳定化 24：如果 $T_2 > 0$ 则 25：对于 $t = 1$ 到 T_2 执行 26：对于所有簇 c 执行 27：服务器广播 $w_c^{(t)}$28：对于所有客户端 $k \in C_c$ 执行 29：客户端最小化 $L_k^c(w_c)$ 并返回 $w_{ck}^{(t)}$， 30：结束循环 31：$w_c^{(t+1)} \leftarrow \sum_{k \in C_c} \frac{|D_k|}{|D_c|} w_{ck}^{(t)}$， 32：结束循环 33：结束循环 34：结束条件 35：阶段 3：剪枝–修复训练 36：对于 $t = 1$ 到 T_3 执行 37：对于所有簇 c 执行 38：如果 $t \bmod f = 0$ 则 39：客户端使用 $w_c^{(t)} \odot M_c^{(t)}$计算梯度 40：计算重要性分数 $\text{Score}_w = \alpha \text{Mag}_w + \beta \text{Coh}_w + \gamma \text{Con}_w$41：计算稀疏度 $S^{(t)}$42：$R \leftarrow \lfloor (T_3 - t)/f \rfloor$43 $\Delta S \leftarrow S_{\text{target}} - S^{(t)}$ R：( )/ : 44 $N_{\text{churn}} \leftarrow \rho \quad N_{\text{active}} \cdot : \cdot$46：$N_{\text{prune}} \leftarrow N_{\text{churn}} + N_{\text{deficit}}$ 45 $N_{\text{deficit}} \leftarrow \Delta S \quad N_{\text{total}}$ 47 $N_{\text{grow}} \leftarrow N_{\text{churn}}$: : 剪除得分最低的已激活权重 49：重新生 48 长梯度最高的未激活权重 50：更新掩码 $M_c^{(t)}$51：结束条件 52：服务器广播 $(w_c^{(t)}, M_c^{(t)})$ 53：对于所有客户端 $k \in C_c$ 执行 54：在掩码 $M_c^{(t)}$约束下最小化 $L_k^c(w_c)$ 55：返回 $w_{ck}^{(t)}$ ， 56：结束循环 57：$w_c^{(t+1)} \leftarrow \left(\sum_{k \in C_c} \frac{|D_k|}{|D_c|} w_c^{(t)}, k\right) \odot M_c^{(t)}$58：结束循环 59：结束循环 60：阶段 4：个性化微调（可选）61：对于所有簇 c 执行 62：$w_c^* \leftarrow w_c^{(T_3)} \odot M_c^{(T_3)}$ 63：对于所有客户端 $k \in C_c$ 执行 64：$v_k \leftarrow w_c^*$65：客户端最小化 $L_k^{CE}(v_k; D_k)$，满足 $v_k \odot M_c^{(T_3)} = v_k$66：结束循环 67： 结束循环 68：返回 $\{v_k\}$

表 4：UCI HAR 和 WISDM 数据集上重要性评分权重的综合消融研究（α、β、γ）。注：“极端非独立同分布”未纳入，因为未在 WISDM 上进行评估。

	权重			UCI HAR			WISDM		
场景	α	β	γ	Acc	公平性得分		Acc	公平性得分	
漂移	1.00	0.00	0.00	0.6860	0.3526	1.9453	0.6129	0.3820	1.6047
漂移	0.00	1.00	0.00	0.6339	0.3148	2.0135	0.5915	0.3456	1.7116
漂移	0.00	0.00	1.00	0.6475	0.3264	1.9840	0.5763	0.3312	1.7399
漂移	0.50	0.25	0.25	0.6939	0.3452	2.0102	0.5774	0.3396	1.7001
漂移	0.25	0.50	0.25	0.6853	0.3546	1.9323	0.5974	0.3671	1.6273
漂移	0.25	0.25	0.50	0.6654	0.3270	2.0350	0.6169	0.3671	1.6807
漂移	0.33	0.33	0.34	0.6967	0.3506	1.9870	0.5797	0.3420	1.6950
标准	1.00	0.00	0.00	0.8964	0.1550	5.7816	0.9143	0.1037	8.8198
标准	0.00	1.00	0.00	0.8548	0.1416	6.0361	0.9040	0.1019	8.8708
标准	0.00	0.00	1.00	0.8139	0.1549	5.2535	0.8906	0.1211	7.3559
标准	0.50	0.25	0.25	0.8796	0.1691	5.2014	0.9031	0.1095	8.2486
标准	0.25	0.50	0.25	0.8971	0.1396	6.4262	0.8939	0.1152	7.7587
标准	0.25	0.25	0.50	0.9119	0.0981	9.2963	0.9157	0.1051	8.7110
标准	0.33	0.33	0.34	0.8769	0.1571	5.5809	0.9103	0.1145	7.9534
噪声客户端	1.00	0.00	0.00	0.7565	0.2111	3.5833	0.7456	0.2204	3.3832
噪声客户端	0.00	1.00	0.00	0.6890	0.2006	3.4351	0.7449	0.2341	3.1822
噪声客户端	0.00	0.00	1.00	0.6449	0.1800	3.5821	0.7322	0.2082	3.5170
噪声客户端	0.50	0.25	0.25	0.7442	0.1903	3.9101	0.7490	0.2333	3.2100
噪声客户端	0.25	0.50	0.25	0.7458	0.1949	3.8267	0.7335	0.2164	3.3894
噪声客户端	0.25	0.25	0.50	0.6860	0.1963	3.4957	0.7436	0.2072	3.5893
噪声客户端	0.33	0.33	0.34	0.7130	0.1843	3.8680	0.7560	0.2189	3.4542

表6：WISDM实验配置

方法	参数	值
通用设置	分阶段轮数（P_1，P_2，P_3）	0, 0, 50
	本地训练轮数（E）	3
	批大小（B）	32
	优化器	Adam（$lr = 1e^{-3}$）
	微调轮数（P_4）	**25**
CA-AFP（我们的方法）	目标稀疏度	自适应（开始于 **70%**）
	评分权重（α，β，γ）	0.25，0.25，0.50
	掩码更新率（更替率）	5%
	簇数量（K）	3
FedSNIP	目标稀疏度	70%
EfficientFL	目标稀疏度	70%
FedCHAR	簇（K）	3
	预热轮次	10
ClusterFL	簇（K）	3
	正则化（ρ）	$5e^{-3}$
	双层参数（α, β）	$1e^{-3}, 5e^{-4}$

表7：CA-AFP消融研究的阶段配置（WISDM）

消融研究	P_1（预热）	P_2（适应）	P_3（剪枝）	P_4（微调）	开始稀疏度
研究A：初始稀疏度（变化 S_{start}）	0	0	40	**3**	变化（0.0 − 0.9）
研究B：聚类后（变化 P_2）	0	变化 {0,5,. . .,20}	40	**3**	0.5（50%）
研究C：初始预热（变化中 P_1）	变化 {0,5,. . .,20}	0	40	**3**	0.5（50%）

表8：重要性分数消融研究的配置

组件	参数	值 / 描述
评分权重 (α, β, γ)	**网格搜索**（7 种组合）	(1.0, 0.0,0.0) – 纯幅度 (0.0, 1.0,0.0) – 纯连贯性 (0.0, 0.0,1.0) – 纯一致性 (0.50, 0.25,0.25) – 幅度主导 (0.25, 0.50,0.25) – 连贯性主导 (0.25, 0.25,0.50) – 一致性主导 (0.33, 0.33,0.34) – 均衡
数据场景	**标准** **漂移** **噪声客户端**	UCI HAR 和 WISDM 的自然用户划分 40% 的客户端带有 100% 的随机标签（传感器故障） 40%的客户端带有30%的随机标签
稀疏度计划	初始稀疏度（S_{start}） 目标稀疏度（S_{target}） 演化策略	**30%**（0.3） **70%**（0.7） 动态剪枝 (0.3 → 0.7)
训练配置	总轮数 (P_3) 本地训练轮数 (E) 微调（P_4） 剪枝频率	**15 轮** **1 个训练轮次** **0 个训练轮次** 每 5 轮

表9：微调消融配置

组件	参数	值
数据集	场景 划分	WISDM（标准） 自然用户划分
阶段1、2和3	轮数 (P_1, P_2, P_3) 稀疏度目标 每轮客户端数	(0, 0, 50) 70% 10
阶段4：个性化	轮数（P_4） 优化器	{1, 5, 10, 15, 20, 25} Adam（$lr = 1e^{-3}$）
对比方法	CA-AFP Global-FT（CA-AFP 的变体） FedSNIP EfficientFL FedCHAR ClusterFL	$K = 3, S = 70\%$ $K = 1$ (无聚类), $S = 70\%$ $K = 1, S = 70\%$ $K = 1, S = 70\%$ $K = 3, S = 0\%$ $K = 3, S = 0\%$

表10：非独立同分布实验配置

参数	配置
数据集	WISDM（活动识别）
客户端总数	36（异构）
全局轮次	50
目标稀疏度	70%（0.7）
每个场景的运行次数	5个独立随机种子
非独立同分布场景	
极端非独立同分布	客户端仅限于**1个类别**
中度非独立同分布	客户端仅限于**2个类别**
轻度非独立同分布	客户端仅限**3 个类别**