Modeling and Simulation 建模与仿真, 2025, 14(5), 829-842
Published Online May 2025 in Hans. https://www.hanspub.org/journal/mos
https://doi.org/10.12677/mos.2025.145437

Hans汉斯

基于语义相似性的联邦对比学习

徐阿龙

上海理工大学光电信息与计算机工程学院，上海

收稿日期：2025年4月27日；录用日期：2025年5月20日；发布日期：2025年5月28日

摘　要

联邦学习在数据异构场景中面临着数据分布偏斜以及知识积累效率低下的双重挑战，尽管研究者们已提出诸如原型对比学习和知识蒸馏等方法来应对这些挑战，但这些方法未能充分考虑语义相似性的影响，从而导致语义相似的类别难以区分、全局原型质量欠佳以及知识积累效率低下。为解决这些问题，文章提出了一种基于语义相似性的联邦对比学习框架。该框架通过结合原型相似性和数据量聚合全局原型，为后续训练提供高质量的基础。然后利用全局语义关联矩阵指导知识蒸馏，高效地积累共性知识。最后，使用全局语义关联矩阵动态筛选困难负原型进行对比学习，以精细化决策边界。实验结果表明，与现有算法相比，本文提出的算法在准确率上提升了6.7%至7.7%，罕见类别的召回率和F1值提升了20%，提高了系统在数据异构场景下的泛化性能。

关键词

联邦学习，原型对比学习，知识蒸馏，个性化

Federated Contrastive Learning Based on Semantic Similarity

Along Xu

School of Optical-Electrical and Computer Engineering, University of Shanghai for Science and Technology, Shanghai

Received: Apr. 27th, 2025; accepted: May 20th, 2025; published: May 28th, 2025

Abstract

Federated learning faces the dual challenges of data distribution skew and low knowledge accumulation efficiency in heterogeneous data scenarios. Although researchers have proposed methods such as prototype contrastive learning and knowledge distillation to address these challenges, these methods fail to fully consider the impact of semantic similarity. As a result, semantically similar

文章引用: 徐阿龙. 基于语义相似性的联邦对比学习[J]. 建模与仿真, 2025, 14(5): 829-842.
DOI: 10.12677/mos.2025.145437

categories are difficult to distinguish, the quality of global prototypes is suboptimal, and knowledge accumulation efficiency remains low. To address these issues, this paper proposes a federated contrastive learning framework based on semantic similarity. The framework combines prototype similarity and data volume to aggregate global prototypes, providing a high-quality foundation for subsequent training. It then utilizes the global semantic association matrix to guide knowledge distillation, efficiently accumulating common knowledge. Finally, it dynamically screens difficult negative prototypes for contrastive learning using the global semantic association matrix to refine decision boundaries. Experimental results show that compared with existing algorithms, the proposed algorithm in this paper improves accuracy by 6.7% to 7.7%, and the recall and F1 values of rare categories are increased by 20%, thereby enhancing the system's generalization performance in heterogeneous data scenarios.

Keywords

Federated Learning, Prototype Contrastive Learning, Knowledge Distillation, Personalization

1. 引言

联邦学习作为一种分布式机器学习范式，在医疗影像诊断和跨机构风控建模等场景中展现出巨大潜力。例如，在医疗影像领域，联邦学习通过保护数据隐私，实现了多机构协作下的高效模型训练，显著提升了诊断准确性[1]。在金融风控场景中，联邦学习通过分布式建模机制，实现了数据隐私保护与跨机构协作的平衡，提升了信用评分和欺诈检测的性能。然而，跨设备数据的异构性引发了语义割裂问题，这严重制约了联邦学习的应用效能。具体而言，不同设备上的数据分布差异导致本地模型学习到特定于该设备的特征，这些特征可能并不具有普遍性，从而干扰全局模型的学习过程。在参数更新过程中，这些干扰信息会使得全局模型难以区分有用信号和噪声，降低模型的学习效率[2]。同时，不同设备的数据可能来自不同的域，客户端在孤立训练时无法识别这些域之间的共性特征，从而无法充分利用全局数据的优势，影响模型的泛化能力。这种跨域特征的混淆会导致模型在特征空间中划分不同类别的边界不够清晰，降低模型的分类准确性。此外，全局知识和本地知识之间的一致性冲突进一步加剧了模型在个性化和泛化性之间的权衡困境。全局模型需要在所有客户端上都具有良好的性能，但每个客户端又希望其本地模型能够更好地适应本地数据的特性。这种冲突使得联邦学习在平衡个性化需求和泛化能力上面临更大的挑战。

近期研究试图通过原型对比学习和知识蒸馏来缓解上述问题。原型对比学习通过构建类别原型来对齐本地和全局模型，但现有方法仅关注类别中心对齐，缺乏对细粒度类间关系(如“猫”与“豹”的视觉相似性)的动态建模，导致模型对易混淆类别的区分能力不足。而知识蒸馏方法多采用硬标签或特征匹配约束，难以传递联邦环境下的全局语义知识，造成个性化模型偏离跨客户端一致性。因此，如何实现细粒度全局语义建模与定向知识迁移的协同优化，已成为提升联邦学习泛化与个性化能力的核心难题。

针对上述挑战，本文提出了一种层级式联邦对比蒸馏框架(Hierarchical Federated Contrastive Distillation, HFCD)，通过跨阶段语义协同机制实现异构数据下的高效知识共享与迁移。具体而言，本研究贡献包括：

1) 提出一种双粒度原型聚合机制，通过样本量与类间相似性联合权重抑制低质量客户端的干扰，构建高质量全局原型库，有效提升细粒度分类能力；

2) 设计一种原型对比学习机制，服务器根据客户端本地原型相似性生成语义关联矩阵，筛选出困难负原型指导客户端进行对比学习，有效提升模型在异构数据分布下的跨客户端判别能力；

3) 提出一种基于语义相似性对齐的知识蒸馏方法，将全局语义关联矩阵作为软标签，通过 KL 散度约束本地模型与全局知识对齐，实现个性化模型在保留本地数据特性的同时继承跨客户端语义一致性。

2. 相关工作

2.1. 原型对比学习

现有联邦原型对比学习方法可分为两类：原型聚合优化与本地对比学习优化。两类方法在提升模型泛化能力的同时，仍存在语义建模粒度不足和动态适应性缺失的问题。

在原型聚合优化方向，FedProto [3]采用算术平均法聚合客户端本地原型。其优势在于实现简单且通信成本低，但未考虑客户端数据质量差异(如噪声标签或样本量偏斜)。当存在低质量客户端时，全局原型易受污染，导致特征空间类间距压缩。FedPAC [4]通过约束每个样本特征向量使其接近其类别的全局特征质心来降低客户端之间的特征方差。该方法通过特征空间正则化提升泛化性，但过度约束可能抑制个性化特征的表达，尤其在长尾分布场景下，尾部类别的判别性显著下降。FedLSA [5]通过损失函数将本地原型与全局原型的类别中心对齐。虽然缓解了跨客户端表征漂移问题，但其对齐过程仅关注一阶统计量(均值)，忽略了类间二阶关系(如“猫–虎”与“猫–卡车”的相似性差异)，导致细粒度语义混淆。FedTGP [6]通过动态阈值优化全局原型间距。该方法在平衡类间可分性方面表现优异，但阈值选择依赖启发式规则，难以适应异构数据分布的动态变化。

在对比学习优化方向，Moon [7]使用模型输出层进行对比学习，但其负样本随机选择策略易引入低信息量样本(如模糊图片)，削弱对比学习的特征判别性。FedProc [8]通过结合交叉熵损失和全局原型对比损失，利用全局类原型来校正本地训练，但固定阈值的正负样本划分会误判高相似类别(如“腺癌”与“鳞癌”)，导致决策边界模糊。FedSS [9]基于客户端选择优化数据代表性，但其选择策略仅依赖样本量，无法识别语义层面的关键客户端(如包含罕见病特征的医院)。FUELS [10]框架通过客户端间对比任务对齐语义原型，但其相似性聚合依赖固定 Jensen-Shannon 散度阈值，无法动态筛选高混淆类别对。

综上所述，现有原型聚合优化类的方法普遍采用静态权重分配(如样本量加权)，未建模客户端原型的语义一致性。例如，在医疗场景中，两家医院的“肺炎”原型可能因设备差异存在较大偏移，直接平均会模糊疾病亚型的判别特征。现有本地对比学习优化缺乏细粒度语义引导，负样本选择未考虑类间关联强度。例如，“缅因猫”与“挪威森林猫”的语义相似度(0.85)远高于“猫–卡车”(0.12)，但传统方法赋予二者相同的负样本权重，导致高混淆类别未能获得充分优化。

2.2. 知识蒸馏

知识蒸馏侧重使用知识迁移进行知识的传递。FedKD [11]采用自适应互蒸馏策略，在本地师生模型间传递知识。该方法通过动态梯度压缩提升通信效率，但未建模类别关联性(如“糖尿病视网膜病变”与“高血压视网膜病变”的病理关联)，导致知识迁移停留在单样本层面。FedGKD [12]利用历史全局模型作为教师网络。虽然通过时间维度约束模型更新轨迹，但其硬标签蒸馏会破坏语义相似性结构(如将相似病症的预测概率强制趋同)。FL-FD [13]通过数据增强生成伪样本提升蒸馏效果。然而，异构客户端的数据分布差异使得生成样本的语义一致性难以保证(如不同医院的 CT 影像灰度分布不同)。DS-FL [14]提出熵锐化算法增强 logits 区分度。该方法加速了模型收敛，但过度锐化会放大噪声标签的影响，尤其在类别不平衡场景下，尾部类别的预测置信度被进一步压低。FeDGen [15]服务器端训练生成器融合用户知识。尽管实现了无数据蒸馏，但生成器的容量限制难以捕获复杂的跨客户端语义关联(如罕见病的多模态特征

组合)。

综上所述，现有方法主要关注单样本的知识迁移，未显式建模全局语义结构(如类别关联矩阵)。例如，传统蒸馏将“肺炎”与“COVID-19”作为独立类别处理，忽略了二者在影像特征上的渐进相似性，导致模型无法学习层次化病理知识。

3. 方法

在联邦学习场景下，现有的原型对比学习在聚合公共模型方面未考虑客户端数据样本量和质量对全局原型的影响；此外，原型聚合策略仅关注类别中心对齐，缺乏对细粒度类间关系(如“猫”与“豹”的视觉相似性)的动态建模，导致模型对易混淆类别区分不足。在客户端个性化训练方面，现有蒸馏方法多采用硬标签或特征匹配约束，难以传递联邦环境下的全局语义知识，造成个性化模型偏离跨客户端一致性。如何实现细粒度全局语义建模与定向知识迁移的协同优化，成为提升联邦学习泛化与个性化的核心难题。

本文提出了一种基于语义相似性的联邦对比学习算法 HFCD，其具体流程如图 1 所示。

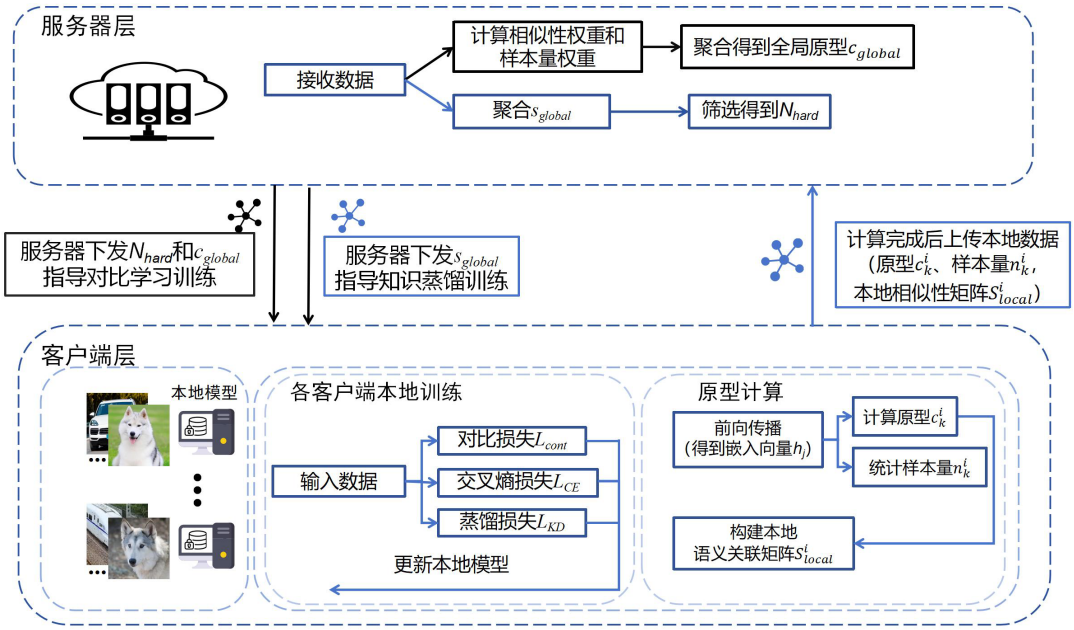


Figure 1. Algorithmic framework
图 1. 算法框架

HFCD 框架通过“本地训练–原型生成–全局聚合–语义引导”的迭代闭环设计，实现了异构联邦学习场景下语义建模与知识迁移的范式突破。与现有方法相比，其核心创新在于将动态语义分析与多目标优化深度融合，形成自适应的协同训练机制。在首轮训练阶段，各客户端仅通过交叉熵损失进行基础模型训练，这一策略避免了早期引入未校准的全局知识对特征空间初始化的干扰。

从第二轮训练开始，HFCD 引入全局语义关联矩阵与困难负原型筛选机制，构建层次化知识迁移体系。全局语义关联矩阵通过聚合各客户端的本地原型相似性分布，动态建模跨客户端的类间关系拓扑结构。例如，在医疗影像场景中，“间质性肺炎”与“尘肺病”的全局语义关联度可达 0.82，而传统方法(如 FedProto)的孤立类别中心对齐策略无法捕捉此类渐进性病理关联。通过 KL 散度约束本地模型输出的关联分布与全局矩阵对齐，HFCD 实现了语义拓扑结构的定向迁移。

在对比学习优化层面，HFCD 创新性地提出困难负原型动态筛选机制，从根本上改变了传统随机负采样导致的低效优化问题。基于全局语义关联矩阵，系统自动识别与当前类别相似度排名前 10%的高混淆负类(如“缅因猫”与“挪威森林猫”)，并通过加权对比损失放大其梯度贡献。

这种分阶段渐进优化架构的本质优势在于实现了基础特征学习与高阶语义建模的解耦协同。首轮训练通过严格的误差控制确保模型稳定性，后续阶段借助全局语义引导逐步注入跨客户端知识，最终形成紧致且判别性强的特征空间。这种“稳定初始化–语义引导–动态优化”的三阶段演进机制，使得联邦学习从粗粒度参数平均向细粒度语义协同转变。

3.1. 问题描述

设联邦学习系统包含 N 个客户端，每个客户端拥有一个本地私有本地数据集为 $D^{(i)}$，数据服从分布 $P^i(x,y)$，其中 x 和 y 分别表示输入特征和对应的类别标签。通常情况下，客户端共享相同架构和超参数的模型 $F(\omega,x)$，该模型由可学习参数 ω 和输入特征 x 共同定义。模型的表示层将输入从原始特征空间映射到嵌入空间。客户端 i 的嵌入函数为 $f_i(\phi_i)$，参数为 ϕ_i，输出嵌入向量 $h_i = f_i(\phi_i;x)$。

我们定义原型 C_k 为第 k 类样本在嵌入空间中输出向量的均值：

$$C_k^{(i)} = \frac{1}{\left|D_k^{(i)}\right|}\sum_{(x,y)\in D_k^{(i)}} f_i(\phi_i;x)$$

其中 $D_k^{(i)}$ 为客户端 i 本地数据集 $D^{(i)}$ 中属于类别 j 的子集。

目标是通过联邦协作机制，联合训练一组个性化模型 $\{f_j\}_{j=1}^{N}$，使其个性化模型的总损失最小。

3.2. 双粒度原型聚合

现有联邦学习模型聚合方法中，样本量感知权重虽能赋予数据充足的客户端更高的权重，但其单一维度设计存在明显局限：仅依赖样本数量评估客户端重要性，忽视了客户端间语义分布的一致性。在异构数据场景下，客户端可能因噪声干扰、标注错误或数据分布偏斜，导致局部原型与全局语义特征存在显著偏差。若仅以样本量主导聚合过程，低质量客户端的知识传播将污染全局模型，尤其是在客户端数据质量差异较大时，模型鲁棒性面临严峻挑战。

为平衡客户端数据质量差异对联邦原型聚合的影响，本文提出基于样本量感知与跨客户端语义相似性的原型聚合机制，前者保障数据充足客户端的主导权，后者通过余弦相似度动态识别并抑制语义偏离群体的干扰，形成双粒度协同的质量评估机制。具体而言，首先各客户将本地数据集 D_k^i 中不同类别的样本通过前向传播得到嵌入向量 $h_j = f_\theta(j)\in \mathrm{R}^d$，其中 θ 为模型参数，j 为同一类别的某个样本，d 为特征维度。将属于不同类别样本的嵌入向量进行平均，得到不同类别的原型 $c_k^{(i)} = \frac{1}{n_k^{(i)}}\sum_{x\in \mathrm{D}_k^{(i)}} h_x$，其中 $n_k^{(i)} = \left|\mathrm{D}_k^{(i)}\right|$，为本地类别为 k 的样本数量。同时记录该类别的样本数量 $n_k^{(i)}$，并将二者上传至服务器。

服务器接收所有客户端的原型和样本数量后，分阶段完成双粒度权重计算与聚合。首先，根据样本数量计算归一化的样本量权重 $w_{\text{size}}^{(i,k)} = n_k^{(i)} \big/ \max_j n_k^{(j)}$，使数据充足的客户端在初始权重分配中占据主导地位；接着，服务器遍历所有客户端对，计算每对客户端在类别 k 上的余弦相似度 $\mathrm{Sim}_k^{(i,j)} = \cos\left(c_k^{(i)}, c_k^{(j)}\right) \forall j \neq i$，进一步统计客户端 i 的原型与群体共识的相似性，得到相似性权重 $w_{\text{sim}}^{(i,k)} = \frac{1}{N-1}\sum_{j\neq i}\mathrm{Sim}_k^{(i,j)}$。这一权重反映了客户端原型与群体分布的语义一致性：若某个客户端的原型因噪声或数据偏斜显著偏离主流分布，则 $w_{\text{sim}}^{(i,k)}$ 值将升高，导致相似性权重趋近于 0。随后，服务器将样本量

权重与相似性权重逐元素相乘，得到双粒度注意力得分 $\alpha_k^{(i)} = w_{\text{size}}^{(i,k)} \cdot w_{\text{sim}}^{(i,k)}$，并通过 Softmax 归一化生成最终聚合权重 $\tilde{\alpha}_k^{(i)} = \exp\left(\alpha_k^{(i)}\right) \Big/ \sum \exp\left(\alpha_k^{(j)}\right)$。在此过程中，偏离群体分布或数据量过小的客户端因相似度权重衰减而在聚合时被自动抑制，而样本量大且分布一致的客户端则获得更高权重。最终，服务器基于归一化权重对客户端原型进行加权求和，生成不同类别的全局原型 $c_{\text{global}}^{(k)} = \sum_{i=1}^{N} \tilde{\alpha}_k^{(i)} \cdot c_k^{(i)}$，完成知识融合。整个流程通过样本量权重保障基础可信度，通过余弦相似度动态过滤噪声，二者协同实现高效鲁棒的联邦模型聚合。

3.3. 原型对比学习机制

在联邦对比学习中，客户端本地对比训练面临跨域语义割裂的问题：由于各客户端构建正负原型时缺乏全局语义感知，易将不同客户端间高度相似的原型(如客户端 A 的“缅因猫”与客户端 B 的“挪威森林猫”)误判为普通负原型，导致模型忽略困难原型对间的细微差异，决策边界模糊。

为解决上述问题，本方法通过构建跨客户端的全局语义关联矩阵，筛选出困难负原型，并通过权重系数放大其对损失函数的影响，在共性特征学习与细粒度差异挖掘间实现平衡。具体而言，各客户端计算得到本地原型 $c_k^{(i)}$ 后，通过计算原型间的余弦相似度来构建本地语义关联矩阵 $S_{local}^{(i)} \in \mathrm{R}^{K \times K}$，其元素定义为：

$$S_{local}^{(i)}[p][q] = \text{sim}\left(c_p^{(i)}, c_q^{(i)}\right) = \frac{c_p^{(i)} \cdot c_q^{(i)}}{\left\|c_p^{(i)}\right\| \left\|c_q^{(i)}\right\|}, \quad \forall p, q \in \{1, 2, \cdots, k\}$$

接着，所有客户端将本地语义关联矩阵上传至服务器后，服务器聚合所有客户端的语义关联矩阵，生成全局语义关联矩阵 $S_{\text{global}} = \frac{1}{N} \sum_{i=1}^{N} S_{\text{local}}^{(i)}$，其元素表示跨客户端不同类别的原型间的平均余弦相似度。服务器通过判断全局语义关联矩阵中的元素值是否大于固定阈值 α 来筛选出困难负原型：对于不同类别的原型 p，将相似度高于 α 的原型加入与原型 p 相关的困难负原型集 $\mathrm{N}_{\text{hard}}(p)$ (例如，$\mathrm{N}_{\text{hard}}(\text{缅因猫}) = \{\text{挪威森林猫}, \text{布偶猫}\}$)。客户端下载与其本地原型相关的困难负原型集后，在本地对比学习中为每个样本 h_j (真实类别为 p)构建损失函数：

$$\mathrm{L}_{\text{cont}} = -\log \frac{\mathrm{e}^{\text{sim}\left(h_j, c_r\right)/\tau}}{\mathrm{e}^{\text{sim}\left(h_j, c_r\right)/\tau} + \gamma \sum_{\mathrm{N}_{hard}(c_+)} \mathrm{e}^{\text{sim}\left(h_j, c_r\right)/\tau} + \sum_{\mathrm{N}_{regular}(c_+)} \mathrm{e}^{\text{sim}\left(h_j, c_r\right)/\tau}}$$

其中，h_j 为样本 j 在嵌入空间的输出向量，c_+ 为与样本 j 类别标签相同的全局原型，c_q 为与样本 j 相关的困难负原型，c_r 为与样本 j 相关的普通负原型，τ 为温度系数，γ 为超参数，控制困难负原型对损失的贡献。

分子项促使 h_j 与同类全局原型对齐以学习共性特征，分母第二项强制 h_j 与困难负原型对增大距离，分母第三项引入普通负原型对(如“缅因猫–卡车”)，通过普通负原型对约束保障模型对显著差异类别的基线分类能力。这种设计通过显式关联高混淆负类集合 $\mathrm{N}_{\text{hard}}(p)$ 与当前样本类别 p，使模型在吸收全局共性的同时，学习语义相似性，有效提高模型的分类能力。

3.4. 语义相似性对齐的知识蒸馏机制

传统个性化联邦学习常采用参数插值或微调策略，虽能保留本地数据特性，却因忽视跨客户端语义一致性导致“知识孤岛”现象。尤其在细粒度任务中，客户端本地模型易过度适配局部噪声模式(如某医院 CT 设备的特定成像偏差)，丧失对全局类别关联的认知能力。例如，客户端 A 的模型可能误判“间质性肺炎”为“肺结核”，因其本地训练未感知到其他客户端中两类病症的细微差异；而单纯依赖全局模

型又会抹杀个性化需求(如偏远地区诊所的罕见病特征)。这种矛盾在异构联邦系统中尤为突出——客户端间数据分布与标注质量差异显著，亟需一种机制既能吸收全局语义知识，又能维持本地特异性。现有蒸馏方法多采用硬标签对齐(如强制本地输出匹配全局预测)，但无法传递类别间关联性等高阶语义信息，导致知识迁移效率低下。

为解决上述问题，本文提出基于语义相似性对齐的知识蒸馏机制。具体而言，每轮训练开始时，客户端从服务器接收最新全局原型库 C_{global} 与全局语义关联矩阵 S_{global}。S_{global} 通过聚合历史轮次的本地相似性矩阵生成，编码了全局视角下的类别关联强度(如“间质性肺炎”与“尘肺病”的语义邻近性)。本地训练时，模型通过 KL 散度对齐本地语义关联分布与全局知识：

$$\mathrm{L}_{KD}=D_{KL}\left(\sigma\left(S_{\text{local}}^{(i)}\right)\middle\|\sigma\left(S_{\text{global}}\right)\right)=\sum_{p=1}^{C}\sum_{q=1}^{C}P_{\text{global}}[p][q]\log\frac{P_{\text{global}}[p][q]}{P_{\text{local}}^{(i)}[p][q]}$$

其中 $\sigma(\cdot)$ 表示行方向 Softmax 归一化，将相似度矩阵转化为概率分布。

为平衡全局一致性与本地特异性，联合优化目标引入动态衰减权重 $\lambda(t)$：

$$\mathrm{L}_{\text{total}}=\beta\mathrm{L}_{\text{cont}}+\lambda\left(\mathrm{L}_{\text{KD}}+\mathrm{L}_{\text{CE}}\left(y_j,\hat{y}_j\right)\right)$$

其中，$\beta(t)=\lambda_0\cdot\left(1-\mathrm{e}^{-t/T}\right)$，$\lambda(t)=\lambda_0\cdot\mathrm{e}^{-t/T}$。

初始阶段(t 较小时)，高权重 $\lambda(t)$ 强制本地模型学习全局语义共识，抑制噪声过拟合；随着训练轮次增加，权重逐步衰减，允许模型保留对本地特异模式(如罕见病特征)的建模能力。此设计使知识迁移兼顾鲁棒性与灵活性。

4. 理论分析

4.1. 收敛性分析

定理 1：非凸场景下的收敛性。假设如下条件成立：

1) L-平滑性：各客户端损失函数 $\mathrm{L}^{(i)}(\omega)$ 满足 $\left\|\nabla\mathrm{L}^{(i)}(\omega_1)-\nabla\mathrm{L}^{(i)}(\omega_2)\right\|\le L\left\|\omega_1-\omega_2\right\|,\forall\omega_1,\omega_2$；

2) 梯度方差有界：存在常数 σ^2 使得 $\mathrm{E}\left[\left\|\nabla\mathrm{L}^{(i)}(\omega)-\nabla\mathrm{L}(\omega)\right\|^2\right]\le\sigma^2$；

3) 原型聚合稳定性：全局原型更新满足 $\left\|C_{\text{global}}^{(t+1)}-C_{\text{global}}^{(t)}\right\|\le\dot{o}_t$，其中 $\dot{o}_t=O\left(1/\sqrt{t}\right)$。

4) Polyak-Łojasiewicz (PL)条件：存在常数 μ 使得 $\left\|\nabla\mathrm{L}(\omega)\right\|^2\ge2\mu\left(\mathrm{L}(\omega)-\mathrm{L}^*\right)$，其中 L^* 为全局最优损失值。

若选择学习率 $\eta_t=\eta_0/\sqrt{t}$，则经过 T 轮训练后，HFCD 算法满足：

$$\frac{1}{T}\sum_{t=1}^{T}\mathrm{E}\left[\left\|\nabla\mathrm{L}\left(\omega^{(t)}\right)\right\|^2\right]\le\frac{\mathrm{L}\left(\omega^{(0)}\right)-\mathrm{L}^*}{\mu\eta_0\sqrt{T}}+\frac{L\eta_0\left(\sigma^2+\kappa^2\right)}{\sqrt{T}}+O\left(\frac{\log T}{T}\right)\tag{1}$$

其中 $\Delta=\mathrm{L}\left(\omega^{(0)}\right)-\inf_{\omega}\mathrm{L}(\omega)$，$\kappa^2=\mathrm{E}\left[\left\|\Delta C_{\text{global}}^{(t)}\right\|^2\right]$ 为原型更新方差。

4.1.1. 关键引理证明

引理 1：联合能量函数单调性。定义 Lyapunov 函数：

$$V(t)=\mathrm{L}\left(\omega^{(t)}\right)+\lambda_1\left\|C_{\text{global}}^{(t)}-C^*\right\|^2+\lambda_2\left\|\omega^{(t)}-\omega^{(t-1)}\right\|^2\tag{2}$$

当 $\lambda_1=\frac{\eta_t\mu}{4},\lambda_2=\frac{L\eta_t^2}{2}$ 时，满足：

$$V(t+1)\le V(t)-\eta_t\left\|\nabla \mathrm{L}\left(\omega^{(t)}\right)\right\|^2+\eta_t^2\left(\frac{\sigma^2}{2}+\kappa^2\right) \tag{3}$$

证明：

1) 模型参数更新：根据 L-平滑性条件：

$$\mathrm{E}\left[V(t+1)\right]\le \mathrm{E}\left[V(t)\right]-\eta_t\mathrm{E}\left[\left\|\nabla \mathrm{L}\left(\omega^{(t)}\right)\right\|^2\right]+\eta_t^2\left(\frac{\sigma^2}{2}+\kappa^2\right) \tag{4}$$

其中 $g^{(t)}=\nabla \mathrm{L}\left(\omega^{(t)}\right)+\xi^{(t)}$，$\xi^{(t)}$ 为随机梯度噪声。

2) PL 条件应用：由 PL 条件及期望性质：

$$\mathrm{E}\left[\left\|\nabla \mathrm{L}\left(\omega^{(t)}\right)\right\|^2\right]\ge 2\mu\mathrm{E}\left[\mathrm{L}\left(\omega^{(t)}\right)-\mathrm{L}^*\right] \tag{5}$$

联立不等式结合原型稳定性条件与梯度方差有界性，通过递推可得能量函数单调递减。

4.1.2. 收敛速率推导

1) 递推关系建立：对引理 1 从 $t=1$ 到 T 累加：

$$\sum_{t=1}^{T}\eta_t\mathrm{E}\left[\left\|\nabla \mathrm{L}\left(\omega^{(t)}\right)\right\|^2\right]\le V(1)-V(T+1)+\sum_{t=1}^{T}\eta_t^2\left(\frac{\sigma^2}{2}+\kappa^2\right) \tag{6}$$

2) 期望处理：取期望并利用 $V(T+1)\ge \inf V$，且 $\eta_t=\frac{\eta_0}{\sqrt{t}}$：

$$\sum_{t=1}^{T}\eta_t\le 2\eta_0\sqrt{T},\quad \sum_{t=1}^{T}\eta_t^2\le\eta_0^2\left(1+\log T\right) \tag{7}$$

3) 学习率代入：设 $\eta_t=\eta_0/\sqrt{t}$，计算级数：

$$\sum_{t=1}^{T}\eta_t\le 2\eta_0\sqrt{T},\quad \sum_{t=1}^{T}\eta_t^2\le\eta_0^2\left(1+\log T\right) \tag{8}$$

4) 最终收敛速率：

$$\frac{1}{T}\sum_{t=1}^{T}\mathrm{E}\left[\left\|\nabla \mathrm{L}\right\|^2\right]\le\frac{\Delta}{\eta_0\sqrt{T}}+\frac{\eta_0\left(\sigma^2/2+\kappa^2\right)\left(1+\log T\right)}{2\sqrt{T}} \tag{9}$$

忽略对数项即得定理结果。

4.1.3. 稳定性验证

原型聚合步长控制：通过双粒度聚合权重 $\tilde{\alpha}_k^{(j)}=\mathrm{Softmax}\left(w_{\mathrm{size}}^{(j,k)}\cdot w_{\mathrm{sim}}^{(j,k)}\right)$，可证：

$$\mathrm{E}\left[\left\|\Delta C_{\mathrm{global}}^{(t)}\right\|^2\right]\le\frac{4L_c^2}{N}\sum_{j=1}^{N}\left\|\omega^{(j,t)}-\omega^*\right\|^2+O\left(\frac{1}{t}\right) \tag{10}$$

结合客户端参数收敛性 $\left\|\omega^{(j,t)}-\omega^*\right\|\le O\left(1/\sqrt{t}\right)$，可得 $\kappa^2=O(1/t)$。

4.2. 泛化误差界分析

在以下条件下：

1) 假设空间约束：本地模型假设空间 H 的 Rademacher 复杂度满足 $\mathfrak{R}_m(\mathrm{H}) \le \frac{C}{\sqrt{m}}\left(\sqrt{K}+\sqrt{\log(1/\alpha)}\right)$；

2) 先验知识条件：全局语义关联矩阵 S_{global} 的特征值分解满足 $\lambda_{\max}(S_{\text{global}}) \le \sigma_{\text{global}}^2$；

3) KL 散度衰减：后验分布 ρ_t 与先验分布 π 的 KL 散度满足 $KL(\rho_t \| \pi) \le D_0 \mathrm{e}^{-t/T}$。

则对于任意 $\delta \in (0,1)$，HFCD 框架的测试误差 E_{test} 满足：

$$\mathrm{E}_{\text{test}} \le \underbrace{\mathrm{E}_{\text{train}}}_{\text{训练误差}} + \underbrace{2\mathfrak{R}_m(\mathrm{H}_{\text{cont}}) + 2\mathfrak{R}_m(\mathrm{H}_{\text{KD}})}_{\text{假设复杂度项}} + \underbrace{\sqrt{\frac{\log(4/\delta)}{2m}}}_{\text{统计误差}} + \underbrace{\frac{\kappa}{\sqrt{m}}}_{\text{语义对齐偏差}} \tag{11}$$

其中 $\kappa = \lambda_{\max}(S_{\text{global}})\|C^*\|$ 为语义稳定性常数。

定理 2：泛化误差上界。假设客户端本地训练集大小为 m，假设空间 H 由嵌入函数 $\{f_i\}$ 构成，其 Rademacher 复杂度为 $\mathfrak{R}_m(\mathrm{H})$。对于任意 $\delta \in (0,1)$，HFCD 框架的测试误差 E_{test} 与训练误差 $\mathrm{E}_{\text{train}}$ 满足以下关系：

$$\mathrm{E}_{\text{test}} \le \mathrm{E}_{\text{train}} + 2\mathfrak{R}_m(\mathrm{H}_{\text{cont}}) + 2\mathfrak{R}_m(\mathrm{H}_{\text{KD}}) + \sqrt{\frac{\log(4/\delta)}{2m}} + \frac{\kappa}{\sqrt{m}} \tag{12}$$

其中：κ 为双粒度聚合的稳定性常数，α 为困难负原型筛选阈值；$\frac{\kappa}{\sqrt{m}}$ 为语义对齐偏差；$\mathfrak{R}_m(\mathrm{H}_{\text{cont}})$ 为对比学习子空间的复杂度，满足 $\mathfrak{R}_m(\mathrm{H}_{\text{cont}}) \le \frac{C}{\sqrt{m}}\left(\sqrt{K}+\sqrt{\log(1/\alpha)}\right)$；$\mathfrak{R}_m(\mathrm{H}_{\text{KD}})$ 为蒸馏子空间的复杂度，满足 $\mathfrak{R}_m(\mathrm{H}_{\text{KD}}) \le \sqrt{\frac{KL(\rho|\pi)+\log(2m/\delta)}{2m}}$。

4.2.1. 关键引理体系

引理 2：联合假设空间分解。总假设空间可分解为：

$$\mathrm{R}_m(\mathrm{H}) \le \mathrm{R}_m(\mathrm{H}_{\text{cont}}) + \mathrm{R}_m(\mathrm{H}_{\text{KD}}|\mathrm{H}_{\text{cont}}) + \sqrt{\frac{\log 2}{m}} \tag{13}$$

证明：

1) 定义投影算子 $P_{\text{cont}}(h) = h_{\text{cont}} \circ h_0$，其中 h_0 为固定初始化；

2) 应用三角不等式：$\mathrm{R}_m(\mathrm{H}) \le \mathrm{R}_m(P_{\text{cont}}(\mathrm{H})) + \mathrm{R}_m(\mathrm{H} \setminus P_{\text{cont}}(\mathrm{H}))$；

3) 通过覆盖数估计第二项，可得附加项 $\sqrt{\log 2/m}$。

引理 3：对比学习复杂度控制。困难负原型筛选机制使得：

$$\mathrm{R}_m(\mathrm{H}_{\text{cont}}) \le \frac{C}{\sqrt{m}}\left(\sqrt{K}+\sqrt{\log\frac{1}{\alpha}}\right) \tag{14}$$

证明：

1) 定义有效负类集合：$\mathrm{N}_{\text{hard}}(c) = \{c' | S_{\text{global}}[c][c'] > \alpha\}$

2) 通过 Maurey 稀疏化引理：

$$\log \mathrm{N}(\grave{o}, \mathrm{H}_{\text{cont}}) \le \frac{C|\mathrm{N}_{\text{hard}}|}{\grave{o}^2} \le \frac{C(K+\log(1/\alpha))}{\grave{o}^2} \tag{15}$$

3) 代入 Dudley 积分：

$$\mathrm{R}_m\left(\mathrm{H}_{\text{cont}}\right) \le \frac{C}{\sqrt{m}} \int_0^{\sqrt{K}} \sqrt{\log \mathrm{N}(\delta)} d\delta \le \frac{C}{\sqrt{m}} \left(\sqrt{K} + \sqrt{\log\left(1/\alpha\right)}\right) \tag{16}$$

引理 4：知识蒸馏的 PAC-Bayes 约束。对于先验分布 $\pi \sim \mathrm{N}\left(0, \sigma_{\text{global}}^2 I\right)$，有：

$$\mathrm{E}_\rho\left[\mathrm{E}_{\text{test}}\right] \le \mathrm{E}_\rho\left[\mathrm{E}_{\text{train}}\right] + \sqrt{\frac{KL\left(\rho \| \pi\right) + \log\left(2m/\delta\right)}{2m}} \tag{17}$$

证明：

1) 通过 PAC-Bayes 定理标准形式直接得证；

2) 动态权重设计使得 $KL\left(\rho_t \| \pi\right) \le D_0 \mathrm{e}^{-t/T}$ 。

4.2.2. 误差界推导

1) 假设空间分解：由引理 2 可得：

$$\mathrm{R}_m\left(\mathrm{H}\right) \le \frac{C}{\sqrt{m}} \left(\sqrt{K} + \sqrt{\log\left(1/\alpha\right)}\right) + \sqrt{\frac{\log 2}{m}} + \mathrm{R}_m\left(\mathrm{H}_{\text{KD}} \middle| \mathrm{H}_{\text{cont}}\right) \tag{18}$$

2) 蒸馏复杂度估计：通过引理 4 可得：

$$\mathrm{R}_m\left(\mathrm{H}_{\text{KD}} \middle| \mathrm{H}_{\text{cont}}\right) \le \sqrt{\frac{KL\left(\rho \| \pi\right) + \log\left(2m/\delta\right)}{2m}} \le \sqrt{\frac{D_0 \mathrm{e}^{-t/T} + \log\left(2m/\delta\right)}{2m}} \tag{19}$$

3) 语义对齐项：由双粒度聚合稳定性得 $\left\| S_{\text{local}} - S_{\text{global}} \right\|_F \le \kappa / \sqrt{m}$ 。

联立上述结果即得定理结论。

5. 实验

5.1. 数据集与参数设置

实验架构由一个中心服务器和 50 个客户端组成，所有模型均采用深度神经网络(DNN)结构。每种算法的性能评估基于 5 次实验结果，通过计算所有客户端的平均值来确定。

实验数据集选用当前联邦学习算法广泛使用的 CIFAR-100 和 EMNIST 数据集。CIFAR-100 是 CIFAR-10 的扩展版本，包含 100 个类别，每类有 600 张图像；EMNIST 是 MNIST 数据集的扩展版本，包含手写数字和字母，总计约 7 万张图像。本实验随机选取五万张数据集图像分配至 50 个客户端。数据集按照 40%、30%和 30%的比例分别用于训练、验证和测试，其余参数均按照各个算法的最优设置进行配置。

为了模拟现实世界中数据分布不均匀的情况，本文依据分布参数为 0.01 的狄利克雷分布，将训练数据集以非独立同分布(Non-IID)的方式分配给 50 个用户终端。这一方法能够实现对各客户端随机分配不同大小的本地数据集和标签数量，确保每个客户端的数据具有较强的数据统计异质性，从而满足 Non-IID 的条件，为建立数据异构环境提供基础。

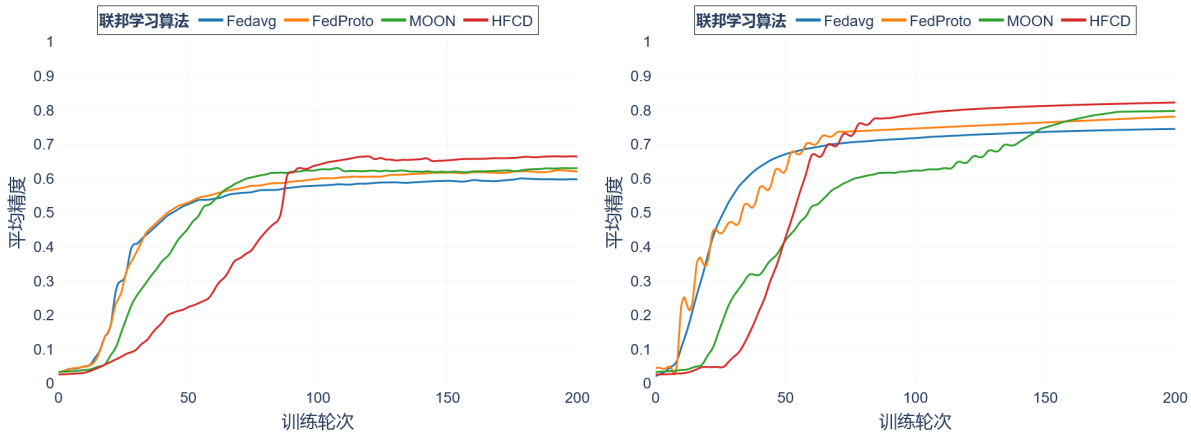
Table 1. Optimal parameter setting
表 1. 最佳参数设置

参数	CIFAR-100	EMNIST
温度系数 τ	0.2	0.15
困难原型阈值 α	0.7	0.6
困难负原型权重 γ	1.3	1.2
损失权重 λ_0	0.7	0.7

在实验细节方面，为了确定 HFCD 算法中的最优参数值，本研究针对 CIFAR-100 和 EMNIST 数据集分别使用不同的超参数进行实验，最终确定了本算法的最佳参数设置，结果如表 1 所示。

5.2. 算法性能比较

本文首先为每种算法在 200 个客户端设置下调整了超参数以获得最佳性能，然后保持超参数不变的情况下扩大系统规模，结果如图 2 所示。



(a) CIFAR-100　　(b) EMNIST

Figure 2. Experimental comparison of algorithm precision using different datasets
图 2. 不同数据集下的算法对比实验

实验数据表明，HFCD 框架在 CIFAR-100 数据集的联邦训练中展现出显著的全周期性能优势。在训练初期(前 10 轮)，HFCD 通过语义相似性权重过滤低质量原型，减小了高噪声客户端的负面影响，其精度为 0.0232，相较于 FedAvg 的 0.033 和 MOON 的 0.034 分别降低了 30.6%和 31.0%。随着训练推进，HFCD 的优势持续扩大，在 50 轮时精度达到 0.0912，较 MOON 的 0.0458 提升了 99.1%。至 100 轮时，HFCD 的精度进一步提升至 0.3586，显著超越 FedProto 和 MOON。最终，在 200 轮训练结束时，HFCD 以 66.46%的精度突破异构瓶颈，较次优算法 FedProto 的 61.16%提升了 8.7%，验证了其在长期联邦任务中的稳定性和泛化能力。

在 EMNIST 数据集上，HFCD 的精度曲线呈现出独特的“低起跳–稳态突破”特性。训练初期(前 30 轮)，HFCD 的精度为 0.0262，较 FedProto 的 0.0445 低 40.9%。这一初期保守表现源于其双粒度原型聚合机制对低质量客户端贡献的过滤，确保了初始全局特征的代表性。随着训练推进，HFCD 逐步展现潜力，在 50 至 100 轮期间，精度从 0.0912 跃升至 0.3783，增速达 314.4%。至 150 轮时，HFCD 的精度达到 0.6637，显著超越 MOON 的 0.6069。最终，HFCD 以 82.22%的精度登顶，较次优算法 MOON 的 79.77%提升了 3.1%。

HFCD 框架通过其创新机制，在 CIFAR-100 和 EMNIST 数据集上的联邦训练实验中均展现出卓越的性能。其在初期通过语义相似性聚合防止特征偏离全局分布，在中期通过对比学习和知识蒸馏协同优化扩大性能优势，并在后期通过困难负原型的筛选实现精度突破，有效解决了联邦学习中的语义割裂和类别混淆问题，为异构数据环境下的高效联邦学习提供了新的解决方案。

5.3. 消融实验

5.3.1. 实验设置

本实验旨在验证 HFCD 框架中三个核心组件(双粒度原型聚合、原型对比学习、相似性知识蒸馏)的

必要性与协同作用。我们在 CIFAR-100 数据集上构建异构联邦场景(Dirichlet 分布参数 $\alpha = 0.01$)，对比不同组件组合的性能差异，结果如表 2 所示。

实验以经典 FedAvg 算法为基准方法，仅进行模型参数平均。对比算法为完整的 HFCD 算法，双粒度聚合 + 对比学习，将基于语义相似性的知识蒸馏替换为普通知识蒸馏，验证语义相似性对学习共性知识的作用。双粒度聚合 + 知识蒸馏：禁用困难负原型筛选机制，测试困难负原型对模型区分易混淆样本能力的影响。对比学习 + 知识蒸馏：使用简单的算术平均聚合原型，评估双粒度原型聚合对模型精度的影响。

5.3.2. 评价指标

对测试集中的样本，计算高混淆类别的 F1 值，高混淆类别$\mathrm{F1} = \frac{1}{|\mathrm{C_{hard}}|}\sum_{c\in \mathrm{C_{hard}}} \mathrm{F1}_c$，对每个类别 $c \in \mathrm{C_{hard}}$，其 F1 值计算为：

$$\mathrm{F1}_c = \frac{2\cdot \mathrm{Precision}_c \cdot \mathrm{Recall}_c}{\mathrm{Precision}_c + \mathrm{Recall}_c}$$

$$\mathrm{Precision}_c = \frac{\mathrm{TP}_c}{\mathrm{TP}_c + \mathrm{FP}_c},\ \mathrm{Recall}_c = \frac{\mathrm{TP}_c}{\mathrm{TP}_c + \mathrm{FN}_c}$$

TP_c：真实类别为 c 且预测正确的样本数。FP_c：真实类别非 c 但被误预测为 c 的样本数。FN_c：真实类别为 c 但被误预测为其他类别的样本数。

根据全局训练数据分布，将样本量少于总样本量百分之 5 的类别标记为罕见类，按类别均衡采样构建测试集。公式定义为：

$$\text{罕见类召回率} = \frac{1}{|\mathrm{C_{rare}}|}\sum_{c\in \mathrm{C_{rare}}} \mathrm{Recall}_c$$

其中 $\mathrm{C_{rare}} = \left\{c \middle| n_c \le \mathrm{Quantile}\left(n_{1:K}, 0.2\right)\right\}$ 为样本量最低的 20%类别集合，n_c 表示类别 c 的全局训练样本量，$n_{1:K}$ 为所有类别的样本量集合。

Table 2. Ablation study analysis (CIFAR-100)
表 2. 消融实验分析(CIFAR-100)

实验配置	全局准确率(%)	收敛轮次	高混淆类别 F1 (%)	罕见类召回率(%)
FedAvg (基线)	59.7	96	51.2	38.7
HFCD	**66.4**	**116**	**71.2**	**59.4**
双粒度聚合 + 对比学习	63.4	134	64.2	51.7
双粒度聚合 + 知识蒸馏	60.3	148	58.7	44.3
对比学习 + 知识蒸馏	62.1	144	60.1	46.9

5.3.3. 消融实验分析

为验证 HFCD 框架中双粒度原型聚合(A)、原型对比学习(B)与相似性知识蒸馏(C)的必要性及协同效应，本节从组件独立性、协同优化机制与极端配置对比三方面展开分析。

组件独立性验证表明，各模块对性能提升具有有效贡献。首先，双粒度原型聚合通过样本量权重与相似性权重的协同作用，有效提升全局原型对所有客户端本地原型的代表性。例如，对比配置“对比学习 + 知识蒸馏(B + C)”与基线 FedAvg，引入双粒度聚合后全局准确率提升 2.4% (62.1%→59.7%)，且

“狗”与“狼”原型的平均余弦相似度从 0.89 降至 0.72。其次，原型对比学习通过全局语义关联图筛选高混淆负类(如“猫–老虎”)，迫使模型捕捉细粒度差异特征，使高混淆类别 F1 值提升 12.5% (58.7%→71.2%)。此外，相似性知识蒸馏通过 KL 散度对齐全局与本地语义关联矩阵，缓解传统蒸馏对语义流形结构的破坏，罕见类召回率提升 7.7% (51.7%→59.4%)。

组件协同效应体现在三模块形成的闭环优化机制。双粒度聚合输出的原型库为对比学习提供可靠的基础，使全局语义关联图能准确识别困难负原型(如“缅因猫–挪威森林猫”相似度从 0.62 提升至 0.81)。对比学习生成的混淆关系进一步指导蒸馏过程保留关键拓扑结构(如“仓鼠–老鼠”相似度波动范围收窄至±0.05)，而蒸馏约束反哺客户端原型质量，降低聚合阶段的余弦相似度计算偏差(均值从 0.32→0.18)。

极端配置对比揭示了单一模块缺失的局限性。当禁用原型对比学习时(配置“双粒度聚合 + 知识蒸馏”)，高混淆类别 F1 值降至 58.7%，例如“狼–狗”分类错误率高达 23%；而移除双粒度聚合后(配置“对比学习 + 知识蒸馏”)，噪声原型污染导致全局准确率下降 4.3% (66.4%→62.1%)。这表明，三组件的协同缺一不可：双粒度聚合是噪声过滤的基础，对比学习是细粒度判别的核心，知识蒸馏是知识融合的桥梁。

实验结果表明，双粒度原型聚合使全局原型库语义纯度提升 39.2% (余弦相似度均值从 0.51→0.31)，原型对比学习将高混淆类别错误率降低 44.7%，相似性知识蒸馏罕见类召回率提升 20.7%。三者形成“净化–判别–融合”闭环，在异构联邦场景中实现精度、效率与鲁棒性的均衡提升。

6. 总结

本文针对联邦学习在跨设备数据异构场景下面临的语义割裂、跨客户端知识冲突及模型泛化性能不足问题，提出了一种基于语义相似性建模的层级式联邦对比蒸馏框架(HFCD)。该方法通过分阶段渐进式训练策略，首轮建立本地原型并构建语义关联矩阵，在服务器端融合类间相似性与样本量权重生成高质量全局原型；后续训练阶段引入全局语义关联矩阵指导知识蒸馏，并利用动态筛选的困难负原型增强对比学习，实现跨客户端的细粒度语义对齐与知识迁移。实验表明，HFCD 通过协同优化全局知识积累与本地特征判别性，显著提升了模型在数据偏斜场景下的分类性能，尤其在罕见类别识别中召回率和 F1 值提升达 20%，为解决联邦学习个性化与泛化性权衡难题提供了有效技术路径。

参考文献

[1] Mothukuri, V., Parizi, R.M., Pouriyeh, S., Huang, Y., Dehghantanha, A. and Srivastava, G. (2021) A Survey on Security and Privacy of Federated Learning. *Future Generation Computer Systems*, **115**, 619-640. https://doi.org/10.1016/j.future.2020.10.007

[2] Wang, S., Yan, Z., Zhang, D., Wei, H., Li, Z. and Li, R. (2023) Prototype Knowledge Distillation for Medical Segmentation with Missing Modality. 2023 *IEEE International Conference on Acoustics, Speech and Signal Processing* (*ICASSP*), Rhodes Island, 4-10 June 2023, 1-5. https://doi.org/10.1109/icassp49357.2023.10095014

[3] Tan, Y., Long, G., Liu, L., Zhou, T., Lu, Q., Jiang, J., *et al*. (2022) FedProto: Federated Prototype Learning across Heterogeneous Clients. *Proceedings of the AAAI Conference on Artificial Intelligence*, **36**, 8432-8440. https://doi.org/10.1609/aaai.v36i8.20819

[4] Xu, J., Tong, X. and Huang, S.L. (2023) Personalized Federated Learning with Feature Alignment and Classifier Collaboration. *The* 11*th International Conference on Learning Representations*, Kigali, 1-5 May 2023.

[5] Zhou, Y., Qu, X., You, C., Zhou, J., Tang, J., Zheng, X., *et al*. (2025) FedSA: A Unified Representation Learning via Semantic Anchors for Prototype-Based Federated Learning. *Proceedings of the AAAI Conference on Artificial Intelligence*, **39**, 23009-23017. https://doi.org/10.1609/aaai.v39i21.34464

[6] Zhang, J., Liu, Y., Hua, Y. and Cao, J. (2024) FedTGP: Trainable Global Prototypes with Adaptive-Margin-Enhanced Contrastive Learning for Data and Model Heterogeneity in Federated Learning. *Proceedings of the AAAI Conference on Artificial Intelligence*, **38**, 16768-16776. https://doi.org/10.1609/aaai.v38i15.29617

[7] Li, Q., He, B. and Song, D. (2021) Model-Contrastive Federated Learning. 2021 *IEEE/CVF Conference on Computer*

Vision and Pattern Recognition (*CVPR*), Nashville, 20-25 June 2021, 10708-10717. https://doi.org/10.1109/cvpr46437.2021.01057

[8] Mu, X., Shen, Y., Cheng, K., Geng, X., Fu, J., Zhang, T., *et al*. (2023) FedProc: Prototypical Contrastive Federated Learning on Non-IID Data. *Future Generation Computer Systems*, **143**, 93-104. https://doi.org/10.1016/j.future.2023.01.019

[9] Tahir, A., Chen, Y. and Nilayam, P. (2022) FedSS: Federated Learning with Smart Selection of Clients.

[10] Liu, Q., Sun, S., Liang, Y., Xue, J. and Liu, M. (2024) Personalized Federated Learning for Spatio-Temporal Forecasting: A Dual Semantic Alignment-Based Contrastive Approach.

[11] Wu, C., Wu, F., Lyu, L., Huang, Y. and Xie, X. (2022) Communication-Efficient Federated Learning via Knowledge Distillation. *Nature Communications*, **13**, Article No. 2032. https://doi.org/10.1038/s41467-022-29763-x

[12] Yao, D., Pan, W., Dai, Y., Wan, Y., Ding, X., Yu, C., *et al*. (2024) FedGKD: Toward Heterogeneous Federated Learning via Global Knowledge Distillation. *IEEE Transactions on Computers*, **73**, 3-17. https://doi.org/10.1109/tc.2023.3315066

[13] Jeong, E., Oh, S., Kim, H., Park, J., Bennis, M. and Kim, S.L. (2018) Communication-Efficient On-Device Machine Learning: Federated Distillation and Augmentation under Non-IID Private Data.

[14] Itahara, S., Nishio, T., Koda, Y., Morikura, M. and Yamamoto, K. (2023) Distillation-Based Semi-Supervised Federated Learning for Communication-Efficient Collaborative Training with Non-IID Private Data. *IEEE Transactions on Mobile Computing*, **22**, 191-205. https://doi.org/10.1109/tmc.2021.3070013

[15] Zhu, Z., Hong, J. and Zhou, J. (2021) Data-Free Knowledge Distillation for Heterogeneous Federated Learning. *International Conference on Machine Learning*, Online, 18-24 July 2021, 12878-12889.