计 算 机 学 报
CHINESE JOURNAL OF COMPUTERS

面向异构联邦环境的原型引导型准确遗忘方法

蒋伟进[1)2)] 崔新雨[1)2)] 刘志华[1)2)] 陈伸有[1)2)] 胡佳龙[1)2)] 陈晓红[1)2)]

[1)](湖南工商大学计算机学院, 长沙 410205)

[2)](湘江实验室, 长沙 410205)

摘 要 联邦持续学习（Federated Continual Learning, FCL）在保护隐私的前提下实现分布式增量知识积累，具有广泛的应用前景。然而，异构数据环境下的模型漂移与灾难性遗忘相互叠加，仍是联邦持续学习面临的核心挑战。为此，本文提出了一种面向异构联邦环境的原型引导型准确遗忘方法。该方法旨在通过精细化的原型管理与选择性知识过滤，在提升模型精度的同时有效缓解知识遗忘。在客户端，本文设计了在线原型平衡与自适应原型反馈机制，旨在通过构建高判别性的特征流形，强化类间可分性与类内紧凑性，为历史知识保持提供稳定的表征基础。在服务器端，通过引入基于归一化流（Normalizing Flows, NF）的概率密度建模，对跨客户端上传的原型进行分布一致性评估，将低密度的离群特征视为不可靠知识，并进行自适应抑制或剔除，从而实现对错误记忆的精准过滤与动态净化。通过原型相关性评估识别并剔除由异构冲突引发的离群知识，从而提高全局知识库的可靠性。在 CIFAR-10($\alpha = 0.5$)、CIFAR-100($\alpha = 0.5$)及 Tiny-ImageNet($\alpha = 0.3$) 基准数据集上的实验结果表明，PAF-FCL 在三个数据集上的分类准确率分别达到 87.33%、80.62%和 74.49%；在遗忘抑制方面，该方法在 Tiny-ImageNet、α=0.3 的复杂异构场景下将遗忘率降至 5.03%；同时，每轮额外通信载荷为 0.64MB，表明该方法具有在资源受限边缘计算场景中应用的潜力。

关键词：联邦持续学习；灾难性遗忘；原型净化；归一化流；数据异构

中图法分类号 TP18

Prototype-guided Accurate Forgetting in Federated Continual Learning

JIANG wei-jin[1)2)] CUI xin-yu[1)2)] LIU zhi-hua[1)2)] CHEN shen-you[1)2)] HU jia-long[1)2)] CHEN xiao-hong[1)2)]

[1)]（Department of Computer Science, Hunan University of Technology and Business, Changsha 410205）

[2)](Xiangjiang Laboratory, Changsha 410205)

Abstract Federated Continual Learning (FCL) enables distributed incremental knowledge accumulation while preserving data privacy and has broad application potential. However, model drift and catastrophic forgetting are compounded in heterogeneous data environments, posing fundamental challenges to FCL. To address these issues, this paper proposes a Prototype-guided Accurate Forgetting method for Federated Continual Learning (PAF-FCL) in heterogeneous federated environments. The proposed method aims to improve model accuracy while effectively mitigating knowledge forgetting through refined prototype management and selective knowledge filtering. On the client side, we design an Online Prototype Equilibrium mechanism and an Adaptive Prototype Feedback mechanism to construct highly discriminative feature manifolds, thereby enhancing inter-class separability and intra-class compactness and providing a stable representational basis for retaining historical knowledge. On the server side, we introduce Normalizing Flow (NF)-based probability density modeling to evaluate the distributional consistency of prototypes uploaded by different clients. Low-density outlier features are regarded as unreliable knowledge and are adaptively suppressed or discarded, enabling precise filtering and dynamic purification of erroneous memories. By evaluating prototype relevance and identifying and removing outlier knowledge caused by heterogeneous conflicts, the proposed method further improves the reliability of the global knowledge repository. Experimental results on three benchmark datasets,

本课题得到国家自然科学基金(61772196)；湖南省自然科学基金(2020JJ4249)，湖南省教育厅科学研究重点项目(24A0446, 24A0753)资助. **蒋伟进**，博士，教授，博士生导师，中国计算机学会（CCF）高级会员，主要研究领域为边缘计算、联邦学习、网络空间安全。**崔新雨**(通信作者)，硕士，CCF 学生会员，主要研究领域为联邦学习、隐私保护。**刘志华**，硕士，CCF 学生会员，主要研究领域为联邦学习、隐私保护。**陈伸有**，硕士，CCF 学生会员，主要研究领域为联邦学习、隐私保护。**胡佳龙**，硕士，CCF 学生会员，主要研究领域为数据动态定价、联邦学习。**陈晓红**，博士，教授，博士生导师，中国工程院院士，主要研究领域为金融工程、管理科学与工程、网络空间安全。

CIFAR-10 ($\alpha = 0.5$), CIFAR-100 ($\alpha = 0.5$), and Tiny-ImageNet ($\alpha = 0.3$), demonstrate that PAF-FCL achieves classification accuracies of 87.33%, 80.62%, and 74.49%, respectively. In terms of forgetting mitigation, PAF-FCL reduces the forgetting rate to 5.03% on Tiny-ImageNet under the challenging heterogeneous setting of ($\alpha = 0.3$). Moreover, the method incurs only 0.64 MB of additional communication overhead per round, indicating its potential for deployment in resource-constrained edge computing environments.

Key words Federated Continual Learning; Catastrophic Forgetting; Prototype Purification; Normalizing Flows; Data Heterogeneity

1 引言

人工智能的发展正在从单任务智能向具备持续学习与自适应能力的通用智能（Artificial General Intelligence, AGI）演进。类比于人类在一生中不断学习新知识并保持旧知识的能力，持续学习（Continual Learning, CL）旨在使模型能够在连续到来的任务序列中不断学习新任务，同时维持对旧任务的性能[1]-[2]。然而，由于在新任务学习过程中无法访问旧任务数据，模型往往会出现“灾难性遗忘”（Catastrophic Forgetting）现象，即新知识的获取导致旧知识丧失[3-4]。这一问题严重限制了深度学习模型在动态、非平稳环境中的应用。

与此同时，随着数据隐私保护和分布式计算的需求不断增强，联邦学习（Federated Learning, FL）成为了一种重要的协同训练范式[5-6]。在联邦学习框架中，各客户端无需共享本地原始数据，通过协同参与模型训练，实现隐私保护条件下的知识共享与模型优化。然而，联邦学习也面临着显著的挑战：由于各客户端数据的非独立同分布（Non-Independent and Identically Distributed, Non-IID），模型聚合后的全局性能常常下降[7]。联邦持续学习[8]（Federated Continual Learning, FCL）的提出就是为了解决传统联邦学习的缺点，它将联邦学习与持续学习结合，使模型能够在多客户端、多任务环境中实现跨域、跨时间的协同学习[9]。

现有的 FCL 方法主要致力于减少灾难性遗忘，通过参数隔离[8]、知识回放[10]或正则化[11]等手段保留旧任务信息[12]。然而，在异构联邦环境下，不同客户端的数据分布往往存在显著差异，导致模型记忆中可能包含偏置特征或无关信息。这种错误记忆不仅不能提升性能，反而可能因特征偏差（Feature Bias）而削弱模型泛化能力。为此，Wuerkaixi A 等[13]提出了准确遗忘（Accurate Forgetting）的概念，通过引入可逆的归一化流模型（Normalizing Flow, NF）[14-15]来估计特征间的相关性，从而选择性地遗忘低置信度或有害特征，实现对跨任务知识的精准筛选与整合。

传统的持续学习研究往往关注全局模型的知识保持，却忽视了客户端本地特征学习的质量问题。实际上，即使在单客户端的持续学习中，模型也容易产生快捷方式学习（Shortcut Learning）现象，即模型倾向于依赖表层或相关性虚假的特征[16]。这种问题在联邦场景中尤为严重，因为每个客户端的局部数据分布受限且多样，极易导致非鲁棒特征的学习与传播。为缓解这一问题，Wei 等[17]提出了在线原型学习框架，通过在线原型平衡（Online Prototype Equilibrium, OPE）与自适应原型反馈（Adaptive Prototype Feedback, APF）机制，使模型能够在单次数据流中学习具有代表性与判别性的特征，从而有效抵御快捷方式学习与灾难性遗忘[18]。

随着任务数量不断增加，传统方法在保持历史任务性能方面面临挑战。由于客户端存储能力受限，模型训练过程通常依赖最新任务的数据进行更新，从而削弱了对早期任务知识的保持能力，导致旧任务性能明显下降。当客户端之间的数据异构性进一步加剧时，该问题会变得更加突出[19]。

基于此，本文提出一种面向异构联邦环境的原型引导型准确遗忘方法（Prototype-guided Accurate Forgetting in Federated Continual Learning, PAF-FCL）。该框架旨在从前端特征学

习与后端知识筛选两个层面协同优化 FCL 过程。

（1)在客户端阶段，采用在线原型学习机制，通过 OPE 与 APF 实现类内紧凑、类间分离的特征提取，提升本地特征表示的稳定性与泛化性；

（2）在服务器阶段，引入准确遗忘机制，利用 NF 模型对来自不同客户端的原型特征进行概率建模与相关性估计，从而识别并遗忘有偏或不相关特征；

（3）在联邦融合阶段，每轮通信周期完成学习—整合—遗忘的优化，使全局模型逐步形成结构稳定、特征可迁移的多任务表示空间。

主要贡献如下：

（1)提出了一种面向异构联邦环境的原型引导型准确遗忘框架 PAF-FCL，实现了对本地特征学习与全局知识筛选的双层优化；

（2）在客户端侧引入在线原型学习机制，以抑制快捷方式学习，增强特征的判别性与可泛化性。在服务器侧采用基于归一化流的概率建模，实现对全局特征分布的准确估计与有害特征的自适应遗忘；

（3）在多个基准数据集上进行的实验表明，PAF-FCL 在准确率、稳定性及抗遗忘能力方面均优于现有的主流 FCL 方法。

2 相关工作

2.1 联邦学习

联邦学习（Federated Learning，FL）是一种分布式协同训练范式，由多个客户端与中央服务器共同组成[5]。各参与客户端基于本地数据独立更新模型参数，并在预设的通信轮次中将更新结果发送至服务器进行统一聚合，从而在保护数据隐私的前提下实现全局优化。由于联邦学习对数据隐私保护的特性，其应用场景非常广泛，包括金融[20]、医疗[21]和物联网[22]等。然而，联邦学习普遍存在统计异质性（statistical heterogeneity）问题，即不同客户端间数据分布差异较大，导致模型聚合后性能退化[23]。为应对该问题，研究者提出了多种改进策略：Mohri 等[24]通过优化全局模型在任意目标分布上的表现实现公平性聚合；Zhu 等人[25]采用生成器对客户端特征进行建模与融合，以实现跨域知识引导。然而，当不同客户端任务完全不一致时，如何在异构任务间实现协同学习仍是开放性难题。

2.2 持续学习

持续学习（Continual Learning，CL）旨在使模型能够在非平稳的数据流或任务序列中持续学习新知识，同时保持对旧任务的记忆能力。Li 等[26]通过知识蒸馏损失限制新旧任务输出的变化，从而在学习新任务时保持旧任务知识的一致性。Kirkpatrick 等[27]引入参数重要性估计机制，使模型在更新参数时优先保留重要权重，从而缓解遗忘问题。Rusu 等人[28]为新任务引入独立的参数分支，从而在不干扰旧任务的前提下实现增量学习。Yoon 等[8]进一步在联邦场景中提出参数分解策略，将权重划分为全局共享参数与任务特定参数，以实现跨任务的知识迁移。重放类方法通过维护一个记忆缓冲区（Memory Buffer）或生成模型来重现旧任务数据，以减轻灾难性遗忘。MIR[29]根据梯度干扰程度选择重放样本；ASER[30]则基于 Shapley 值优化样本选择策略；而生成式重放（Generative Replay）利用生成模型重构旧任务数据，在缺乏真实样本时维持旧任务知识[31]。

2.3 联邦持续学习

近年来，随着数据合规要求提升与多机构协作需求增加，FL 通过在客户端本地训练并上传模型更新，在不直接共享原始数据的前提下完成多方协作建模，经典范式通常以联邦平均（Federated Averaging，FedAvg[5]）为代表。CL 关注模型在任务序列中学习新知识的同时保持旧知识，核心困难是灾难性遗忘（catastrophic forgetting）。Rebuffi 等人[32]结合样本记忆与蒸馏来维持旧类判别能力。然而，实际应用中任务会持续演化且模型需要不断吸收新知识，这使得持续学习与联邦学习的结合成为必然趋势，即联邦持续学习（Federated Continual Learning，FCL）[8]。在 FCL 场景下，各客户端不仅数据分布不同，还可能面对不同的任务顺序与任务相关性，形成典型的异构环境[33]。此时，模型既要学习新任务、保持旧任务能力，又要在跨客户端协作中获得稳定泛化，这对本地训练与跨轮聚合均提出更高要求。

与传统观点不同，在异构 FCL 中“遗忘”并不总是负面现象。由于客户端之间任务可能无关或弱相关，强行保留所有历史知识可能引入偏置迁移，甚至降低当前任务与全局协作的性能。因此，近期研究 Wuerkaixi A 等人[13]提出异构联邦持续学习的准确遗忘，通过在特征空间进行生成

回放并对历史知识进行相关性筛选。Yoon 等[8]最早给出 FCL 的定义，其提出的方法将模型参数拆分为通用参数和任务特定参数，客户端之间共享通用参数，且通过注意力机制选择性的学习其他客户端上对自身有用的任务特定参数[34]。

3 面向异构联邦环境的原型引导型准确遗忘方法

3.1 问题设置

本文考虑由N个客户端组成的联邦学习系统，各客户端仅在本地数据上训练并参与协作聚合，但不同客户端的本地数据特征-标签联合分布$p(x,y)$，x 表示输入样本，y表示对应标签，可能不相同，从而形成统计异质性。在持续学习中，任务按步骤 r=1,…,R 顺序到来，其中 R 表示任务总数；以$t = 1,2,\ldots,T$表示联邦通信轮次。学习第 t 步时无法直接访问此前步骤的数据。其目标可写为在获得θ^r的同时兼顾截至 t 的任务风险：

$$\min_{\theta}\sum_{t=1}^{T}\sum_{r=1}^{R}L_r(\theta^t;\mathrm{T}_r) \tag{1}$$

其中，θ^t表示第t轮通信后的全局模型参数，$L(\cdot)$第r个任务上的模型损失。该目标旨在学习新任务的同时保持对历史任务的性能。

FCL 将上述持续学习约束引入联邦场景，客户端k在第t轮可访问当前数据集D_k^t。进一步地，引入记忆池M_k^t用于存储客户端k在第t轮的历史样本。不同客户端之间、以及同一客户端不同时刻的任务之间不保证相关，甚至可能相互矛盾。因此，系统在隐私约束下的目标是协同学习全局参数θ^t，使所有客户端在截至当前已见任务上取得尽可能好的整体性能：设客户端k的本地任务历史为$\mathcal{T}_1,\ldots,\mathcal{T}_r$，全局模型$\theta$在其上的损失为$L(\theta;\mathcal{T}_t)$。目标：

$$\min_{\theta}\sum_{k=1}^{N}\sum_{r=1}^{R}L_r(\theta^t;\mathrm{T}_r) \tag{2}$$

用于衡量模型在各任务上的性能表现。

在上述无限任务池设定下，每个客户端本地的学习过程可形式化地定义为在线类增量学习（online Class-Incremental Learning, online CIL）：数据以单次遍历流形式到达$\mathrm{D}=\{D^1,\ldots,D^r\}$，其中$D_k^r=\{(x_i,y_i)\}_{i=1}^{n_t}$，$D_k^r$ 表示客户端 k 在任务 r 上的数据。标签满足$y_i\in C_r$，且不同任务的类别集合C_r两两不交。基于回放式方法会引入记忆库M_r存储少量历史数据；在任务 r 的每个时间步，模型接收用于训练的小批数据$X\cup X^b$，其中X来自D^t，X^b来自 M。此外采用单头（single-head）评估：由于测试时任务标识不可知，统一分类器需在所有已见类别中进行预测，目标是在数据仅被访问一次的条件下，使模型同时在新、旧类别上保持良好的性能。

3.2 客户端在线原型学习与自适应原型反馈

在 FCL 的在线类增量场景中，模型不仅需要在单次数据流中快速学习新类别特征，还需在缺乏完整历史数据的条件下维持对已见类别的判别能力。然而，受限于数据分布的动态变化与记忆库容量约束，模型在特征学习过程中容易出现两类问题：一是新旧类别之间的表示不均衡，导致模型过度偏向当前任务；二是类别间判别边界模糊，尤其在异构数据条件下，易产生类间混淆并加剧灾难性遗忘。

为此，本文从全局均衡性与局部判别性两个层面对原型学习机制进行设计。在整体层面，引入 OPE，通过对新旧类别原型进行统一约束，确保特征空间在持续学习过程中保持类内紧凑与类间分离的基本结构，从而缓解由任务切换引发的表示偏移问题；在细粒度层面，引入 APF，基于类别间原型距离动态识别易混淆类别，并通过针对性的数据增强强化其决策边界，以提升模型对困难类别的区分能力。

OPE 与 APF 在持续学习过程中构成了局部结构化与全局共识引导的协同模式。OPE 侧重于维持整体特征分布的稳定性，防止模型在持续学习过程中发生全局性偏移；而 APF 则聚焦于局部判别能力的增强，针对性修复潜在的类别混淆问题。二者协同作用，使模型在 FCL 过程中既能保持稳定的表示结构，又具备对复杂类别关系的自适应建模能力，从而为后续的全局知识筛选与准确遗忘提供高质量的特征基础。

本地训练过程遵循在线类增量学习范式：模型以小批方式单遍历数据集，每个批次包含新任务样本和少量记忆库样本用于重放。为后续实现在线原型学习，本文定义客户端 k 的本地模型参数为：

$$\theta_k=\left\{\theta_f,\theta_g,\theta_c\right\} \tag{3}$$

其中，$\theta_f,\theta_g,\theta_c$分别表示编码器、投影头和分类器参数。由编码器$f\left(\cdot;\theta_f\right)$、投影头$g\left(\cdot;\theta_g\right)$与分类器$c\left(\cdot;\theta_c\right)$三部分组成。对于经过数据增强

的输入样本 aug（x），其表示向量计算如下：

$$z = g_{\theta_g}(f_{\theta_f}(aug(x))) \tag{4}$$

其中，$aug(x)$表示数据增强操作，z表示样本在特征空间中的嵌入表示。

回放小批数据 X^b 的表示记为 Z^b。在步骤 t 的一个小批中，类别 c 的在线原型定义为小批次中同类样本类特征均值：

$$p_c = \frac{1}{\left|\left\{j : y_j = c\right\}\right|} \sum_{y_j = c} z_j \tag{5}$$

其中，j表示样本索引，c 表示类别索引，P_c表示第c类的原型向量，z_j 表示样本j的嵌入表示，|·|表示集合的基数。

由此得到当前小批的原型集合：

$$P = \{P_c\}_{c=1}^{C} \tag{6}$$

回放小批的原型集合：$P^b = \{p_c^b\}_{c=1}^{K_b}$，并有 $K = |P| \le |C|$。

3.2.1 在线原型平衡

对于当前批次的新类别原型集合 P^{new} 和记忆库的旧类别原型集合 P^{old}，分别计算对比损失：

$$\mathrm{L}_{OPE}^{new} = -\frac{1}{\left|P^{new}\right|} \sum_{i=1}^{\left|P^{new}\right|} \log \frac{\exp(P_i^{\cdot} \hat{P}_i / \delta)}{\sum_{j \neq i} \exp(P_i^{\text{嘘}} \hat{P}_j / \delta) + \sum_j \exp(P_i \hat{P}_j / \delta)} \tag{7}$$

其中，$\hat{P}_i$ 为原型 P_i 的增强视图，δ 为温度超参数。

在线原型模型的最终对比损失：

$$\mathrm{L}_{Diff}\left(P, P^b\right) = \frac{1}{2}(\ell(P, P^b) + \ell(P^b, P)) \tag{8}$$

其中，$\ell(\cdot,\cdot)$为原型级对比形式，P, P^b 分别为当前批次和回放批次的原型集合。

考虑到在在线增量学习中同时学习新类别和巩固已学知识，本文引入了 OPE 模块[17]，通过使用 L_{Diff} 从新类别和已见类别两方面入手学习具有代表性和区分性的特征：

$$\mathrm{L}_{OPE} = \mathrm{L}_{Diff}^{new}(P, \hat{P}) + \mathrm{L}_{Diff}^{seen}(P^b, \hat{P}^b) \tag{9}$$

其中，$\hat{P}$ 和 $\hat{P}^b$ 分别表示当前原型集合和回放原型集合经过增强后的视图。

3.2.2 自适应原型反馈

虽然 OPE 可以带来整体平衡，但其对各类别采取无差别的处理方式。实际上，不同类别间的混淆程度存在差异，模型应有针对性地关注易混淆的类别，用以巩固已学习的知识[35]。为此，本文引入 APF 机制，该机制依据在线原型反馈来识别易被误分类的类别，并强化其决策边界[36]。

对于记忆库 M 中的每一类别对，APF 基于上一时间步所有已见类别的在线原型，计算其相互间的距离，以此反映类别间的混淆情况。两个原型越接近，对应类别越容易被误分类。基于这一分析，本文的目标是增强这些易混淆类别之间的决策边界。为此，通过对称高斯核将原型距离矩阵转化为类别上的概率分布 $\boldsymbol{P}$：

$$\boldsymbol{P}_{i,j} = \frac{\exp\left(-\left\|p_{k,i}^b - p_{k,j}^b\right\|_2^2 / \left(2\sigma^2\right)\right)}{\sum_{a \neq b} \exp\left(-\left\|p_{k,a}^b - p_{k,b}^b\right\|_2^2 / \left(2\sigma^2\right)\right)}, i \neq j \tag{10}$$

其中，σ 为核宽度常数。

自适应原型反馈通过一种基于采样的混合增强策略实现。具体而言，APF 依据概率分布 $\boldsymbol{P}$，自适应地从记忆库 M 中倾向于选择易混淆类别的样本进行混合增强[37]。为减轻对当前在线原型均衡性的过度约束，本文在大小为 m 的回放数据 X^b 上设计了一种两阶段采样策略。首先，根据 $\boldsymbol{P}$ 选取 n_{APF} 样本，其中较高的 $\boldsymbol{P}_{a,b}$ 值意味着从类别 a 和 b 中抽取更多样本。$n_{APF} = \rho \cdot m$，ρ 为 APF 采样比例。其次，剩余的 $m - n_{APF}$ 样本则从整个记忆库中均匀随机抽取，以避免模型过度聚焦于易混淆类别，从而维护已学习到的类别间平衡。

对于混合增强，取两条回放样本 $\left(x_i, y_i\right), \left(x_j, y_j\right)$，构造：

$$\begin{aligned} \bar{x} &= \lambda x_i + (1-\lambda) x_i, \\ \bar{y} &= \lambda onehot\left(y_i\right) + (1-\lambda) onehot\left(y_i\right) \end{aligned} \tag{11}$$

其中，$\lambda \sim Beta(\eta, \eta)$为插值系数，$\left(x_i, x_j\right)$ 分别为来自两类的新样本，$\left(y_i, y_j\right)$ 为对应标签，*onehot* 表示独热编码。

通过 OPE 和 APF 两个模块的协同作用，模型能够学习到对快速学习具有判别性的特征，然而，由于在线原型难以覆盖全部实例级信息，各类别的特征表示可能不够紧凑。为进一步提升类

内聚合度，引入监督对比学习（supervised contrastive learning，SupCon）[38]以学习实例级别的表示：

$$L_{SupCon}=\frac{1}{|X|}\sum_{i=1}^{|X|}\left[-\frac{1}{|A(i)|}\sum_{j\in A(i)}\log\frac{\exp(z_i\cdot z_j/\tau)}{\sum_{a\in A(i)}\exp(z_i\cdot z_j/\tau)}\right] \tag{12}$$

其中，$A(i)$ 为与样本 i 同类的其他样本索引集合，τ为监督对比损失中的温度系数。z_i 为样本 x_i 经编码器和投影头得到的特征向量。

客户端本地训练的总损失为：

$$L_{client}=L_{OPE}+L_{cls}+L_{SupCon} \tag{13}$$

$$L_{cls}=cls\left(\bar{y},c\left(z\left(\bar{x};\theta_f,\theta_g\right);\theta_c\right)\right) \tag{14}$$

其中，L_{cls} 为分类交叉熵损失，L_{SupCon} 为对比损失。$(\phi(x;\theta))$ 为模型的分类概率输出，通过该损失，模型能够在特征空间中拉近同类样本，包含历史回放数据，推远异类样本，从而实现实例级的类内紧凑。

记忆库采用在线更新策略[17]，按照任务步 r 调整记忆库样本的保留概率。在每个任务步更新记忆库 M，通过从回放数据 X 中均匀随机选择样本放入 M，如果 M 满了，则从 M 中去除等量的样本，以维持固定容量：

$$M_k\leftarrow M_{k,c}\cup\left\{\left(x_i,y_i\right)\right\},with\ prob.\frac{|M_k|}{r} \tag{15}$$

其中，M_k 为客户端 k 的本地记忆库，$M_{k,c}$为客户端k中类别c的记忆样本集合。

当客户端在每轮通信结束时收到服务器广播的净化后全局原型记忆库M'后，执行以下同步操作。对于M'中的每个类别原型p_c^{global}；（1）若类别 c 已存在于客户端本地原型集合P_k^t中，则利用全局原型对本地原型进行加权融合：$p_{k,c}\leftarrow\beta\mu_{k,c}+(1-\beta)\mu_c^{\text{global}}$，其中$\beta\in(0,1)$为本地保留系数，控制本地统计与全局统计的权衡；（2）若类别 c 不在当前客户端参与训练的类别集合C_k^t内，则将该全局原型作为补充知识插入本地原型集合，并为其分配一个较小的初始权重，使其在后续训练中仅作为弱约束参与 OPE 与 APF 的计算。

在每一轮通信结束后，客户端在接收到服务器返回的全局模型参数及更新后的原型记忆库后，需要对本地记忆信息进行同步调整。具体而言，客户端首先将服务器下发的全局原型集合与本地记忆库进行对齐，对于类别标签一致的原型，采用加权融合策略更新本地原型表示，即结合本地原型与全局原型的特征均值，以平衡本地数据分布与跨客户端共享知识；对于仅存在于全局记忆库中的新类别原型，则直接引入本地记忆库，以增强模型对未观测类别的先验感知能力。

同时，对于服务器端在上一轮中已判定为低相关性的原型，客户端在本地记忆库中执行相应的同步删除或权重衰减操作，以避免过时或噪声知识在后续训练过程中持续干扰模型更新。在完成上述同步后，更新后的本地记忆库将参与下一轮的特征学习与原型构建过程。

3.3 服务器端聚合与准确遗忘

3.3.1 基于 NF 的原型相关性评估

PAF-FCL 将分类器拆分为特征提取器与分类头，并在特征空间训练条件 NF 用于生成回放，NF 记为 $s(\cdot;\psi)$，参数为ψ，其在特征空间工作。

潜变量映射与先验：

$$u=s(z;\psi),z=f(x),u\sim N(0,I) \tag{16}$$

其中，$(u\sim \mathrm{N}(0,I))$为标准正态噪声，$s(;\psi)$为可逆映射函数，用于将噪声映射为生成的潜变量 u。

生成回放是从先验样本$\bar{u}\sim N(0,I)$，经逆变换得到的生成特征$\bar{z}=s^{-1}(\bar{u};\psi)$，并与标签条件形成生成样本对$(\bar{r},\bar{y})$。每轮通信开始时，客户端可从上一轮训练好的 NF 生成若干合成特征及伪标签，将其并入本地训练集与真实样本一起训练。

通过 NF 模型评估各历史知识的可信度，服务器能够有选择地遗忘那些不可靠的、可能源自数据噪声或任务冲突的特征表示。

在客户端 k 的任务步 r，使用当前真实特征与历史生成特征共同训练 NF。由上一任务步保存的 NF 采样得到的历史特征集合为G_r，则 NF 的训练目标可写为负对数似然之和：

$$L_{eNF}\left(\psi;D_k^r,G_r\right)=$$
$$-\frac{1}{\left|D_k^r\right|}\sum_{(z,y)\in D_k^t}\log\boldsymbol{p}(z,y)-\frac{1}{|G_r|}\sum_{(z,y)\in G_r}\log\boldsymbol{p}(z,y) \tag{17}$$

其中，G_r 表示第 r 个任务对应的 NF 生成特征集

合。$\boldsymbol{p}(z,y)$为 NF 在特征空间给出的条件密度。

服务器收集各客户端上传的本地原型集合$\{P_k\}_{k=1}^{N}$，对于每个类别 c，计算聚合后的全局原型：

$$p_c=\frac{\sum_{k=1}^{N} n_k^c\cdot P_k^c}{\sum_{k=1}^{N} n_k^c} \tag{18}$$

其中，n_k^c为客户端 k 中类别 c 的样本数量，p_k^c为客户端 k 中类别 c 的原型，N 表示参与联邦训练的客户端总数。服务器端的全局模型库记为$M_{global}=\{P_k^c\}_{c=1}^{C}$。

模型参数用 FedAvg 聚合：

$$\theta^t=\sum_{k=1}^{N}\frac{n_k^t}{n^t}\theta_k^t \tag{19}$$

其中，$n^t=\sum_{k=1}^{N} n_k^t$为当前通信轮次中参与聚合的类别 c 的总样本数。

利用潜空间近似高斯的性质，用对角高斯近似当前任务分布，以样本估计其均值与协方差：

$$\begin{aligned}\mu_k^c&=\frac{1}{n_k^c}\sum_{i=1}^{n_k^c}u_i,\\ \Sigma_k^c&=\frac{1}{n_k^c}\sum_{i=1}^{n_k^c}\left(u_i-\mu_k^c\right)\left(u_i-\mu_k^c\right)^{\cdot}\end{aligned} \tag{20}$$

其中，u_i表示样本 i 的特征表示。

3.3.2 准确遗忘机制

准确遗忘机制对生成潜变量$\bar{u}$的相关性定义为其在该高斯下的概率密度：

$$\begin{aligned}\boldsymbol{P}_k^c\left(\bar{u}\right)&=\frac{1}{\left(2\pi\right)^{d/2}\left|\sum_k^c\right|^{1/2}}\\ &\exp\left(-\frac{\left(\bar{u}-\mu_k^c\right)^{\cdot}\left(\sum_k^c\right)^{-1}\left(\bar{u}-\mu_k^c\right)}{2}\right)\end{aligned} \tag{21}$$

其中，$\bar{u}$为该类当前特征分布的平均向量，μ_k和Σ_k分别为该类特征分布的均值和协方差矩阵。

该权重越小，表示生成知识在当前任务分布下越离群，应被抑制。

基于原型相关性得分，实施选择性遗忘策略：

$$\rho\left(p_k^c\right)=\frac{\boldsymbol{P}_k^c\left(\bar{u}\right)}{\sum_{j=1}^{N}\boldsymbol{P}_j^c\left(\bar{u}\right)} \tag{22}$$

其中，p_k^c为服务器全局记忆库中类别 c 的原型，$\boldsymbol{P}_k^c(\bar{u})$为客户端 k，类别 c 的概率密度。

重新加权：

$$w(p_k^c)=\frac{1}{1+\exp\left(-\beta\left(\rho\left(p_k^c\right)-\tau\right)\right)} \tag{23}$$

其中，$w(p_k^c)$为原型更新权重，β控制遗忘强度，τ为遗忘阈值。

经相关性加权后的原型：

$$p_k^{c,new}=w\left(p_k^c\right)\cdot p_k^c \tag{24}$$

对于低相关性原型，直接从全局模型库中移除：

$$P_k^{forget}=\left\{p_k^c\in P_k\mid\rho\left(p_k^C\right)<\tau_{hard}\right\} \tag{25}$$

其中，τ_{hard}为硬阈值，低于该阈值的原型将从全局库中移除。

为了稳定表示空间，引入特征层蒸馏项：

$$L_{KD}=\frac{1}{\left|D_k^r\right|}\sum_{x\in D_k^t}\left\|f_a\left(x;\theta_k^t\right)-f_a\left(x;\theta_k^{t-1}\right)\right\|_2^2 \tag{26}$$

其中，L_{KD}为知识蒸馏损失，$f_a\left(x;\theta_k^t\right)$和$f_a\left(x;\theta_k^{t-1}\right)$分别为当前和上一轮模型的特征编码函数。

服务器根据各客户端上传的信息更新全局模型，并利用准确遗忘机制筛除跨任务的错误记忆，从而优化全局知识。基于上述内容，整体算法见算法 1。

算法 1 Server

输入： Initial global model parameters θ^0; total rounds T; forgetting threshold τ and intensity factor β; each client k's local data sequence $\{D_k^t\}_{t=1}$ and prototype memory M_k

输出： Final global model parameters and global prototype memory M.

1. Initialize: Set global prototype memory
2. for each communication round t = 1 to T do
3. Server samples a subset X_b of clients for training;

4. for each client $k \in X_b$ (in parallel) do
5. Server sends current global model θ^{t-1} to client *k*;
6. Client k initializes local model $\theta_k^t \leftarrow \theta_k^{t-1}$;
7. Local dataset of task *r* at round *t* on client *k*;
8. Client *k* uploads updated parameters θ_k^t and local prototype set to server;
9. end for (client-side training phase)
10. Server aggregates θ_k^t: $K \in X_b$ into global model // e.g., FedAvg weighted by $\left|D_k^t\right|$.
11. Server updates global prototype memory $M_k \leftarrow M_k \cup \left\{\left(x_i, y_i\right)\right\}, with\ prob. \frac{\left|M_k\right|}{t}.$ // merge collected client prototypes
12. // Server-side accurate forgetting mechanism (NF-based correlation analysis)
13. for each prototype $p_k^c \in M$ do
14. Compute correlation score $\rho\left(p_k^c\right)$ via the normalizing flow (NF) model;
15. if $\rho(p_k^c) < \tau$ then
16. Remove p_k^c from *M* // forget low-confidence prototype
17. else
18. Reduce the influence weight of p_k^c // suppress less relevant feature
19. end if
20. end for
21. Server broadcast updated global model and filtered prototype memory θ^t and filtered prototype memory M' to all clients;
22. end for

3.4 客户端-服务器端原型交互流程

为明确客户端与服务器之间在原型更新与筛选过程中的交互逻辑，本文对每一轮联邦通信中的信息流进行描述。在第 *t* 轮通信开始时，服务器首先向选中的客户端广播当前全局模型参数以及上一轮维护的全局原型记忆库。客户端在接收全局模型后，基于当前任务数据与本地记忆库执行在线类增量训练，并在训练过程中动态构建当前批次的类别原型。同时，客户端结合上一轮服务器返回的原型信息，对本地特征表示进行约束，以保证局部表示与全局知识的一致性。

（1）客户端本地训练与原型提取。客户端 *k* 在接收到服务器广播的全局模型参数后，利用当前任务数据与本地记忆库进行本地训练。训练完成后，客户端根据当前任务数据计算各类别的在线原型均值$\mu_{k,c}^r$，并将以下两项信息上传至服务器：1)更新后的本地模型参数θ_k^t；2)本地类别原型集合$P_k^r = \{\mu_{k,c}^r\}_{c \in C_k^r}$，以及每类原型的样本数量$n_{k,c}^r$，其中$C_k^r$为第 *r* 任务可见类别集合。

（2）服务器聚合与原型合并。服务器收集所有参与客户端的模型参数与原型信息后，服务器端接收到各客户端原型后，激活基于 NF 的评估器。该评估器将$\mu_{k,c}^r$映射至潜空间，计算其概率密度$\boldsymbol{P}(\mu_{k,c}^r)$。若$\rho(p_k^c) < \tau$，则判定该原型为受异构噪声干扰的冲突原型并予以剔除。然后执行两步聚合：1)对模型参数采用加权联邦平均，权重正比于客户端参与训练的样本量；2)对原型进行类级别合并：对于每个类别 *c*，收集所有拥有该类样本的客户端上传的原型与样本量，计算加权全局原型$p_k^c = \frac{\sum_k n_{k,c}^r \mu_{k,c}^r}{\sum_k n_{k,c}^r}$，并将新产生的全局原型并入全局原型记忆库$M$中。

（3）服务器端准确遗忘筛选。服务器利用归一化流模型对全局原型记忆库中的每个原型进行概率密度估计，计算各原型在当前特征分布下的相关性得分。根据相关性评分$\rho(p_k^c)$将原型划分为三类：1)高相关性原型——保留并赋予完整权重，作为可靠的全局知识；2)中等相关性原型——降权保留，抑制其对后续训练的潜在干扰；3)低相关性离群原型——直接从记忆库中移除，防止异构噪声污染全局知识库。筛选后的全局原型记忆库M'即为经过净化的知识集合。

（4）净化知识广播与客户端继承。服务器将聚合后的全局模型参数与净化后的全局原型记忆库M'一并广播至所有客户端。客户端在下一轮本地训练中，除使用本地记忆库进行回放外，还可参考全局原型作为额外的知识约束，从而在本地形成更完整的类别分布视图，缓解 Non-IID 导致的局部偏差。

通过上述交互机制，在联邦训练过程中形成了本地表征学习—全局原型筛选—知识反馈约束的闭环优化路径，使得跨客户端的知识在多轮通信中不断被筛选、强化与重构，从而有效缓解异构环境下的知识冲突与错误记忆累积问题。

在 FCL 中，异构客户端上传的本地原型可视

为对全局类别分布的局部观测。由于 Non-IID 的影响，这些局部观测存在系统性偏差，若不加甄别地聚合，等效于在全局知识库中引入非平稳噪声。本文通过客户端净化—服务器筛选的信息质量控制框架来阻断噪声传播。

为进一步说明 OPE、APF 和 NF 三个模块之间的内在关联，本文从信息处理流程角度分析：OPE 负责客户端侧原型结构优化，APF 进一步增强类别判别信息，而 NF 在服务器端完成全局知识一致性筛选，该过程是由原型质量提升—判别信息强化—异常知识过滤组成的逐级优化过程。

首先，客户端接收到的新任务数据通常存在类别不均衡和分布偏移问题，导致原始原型包含大量不稳定局部信息。OPE 通过在线原型平衡机制调整原型分布，使上传前的知识表示具有更高一致性。然而，原型结构优化无法保证所有类别之间具有充分判别性，因此仍需要进一步通过 APF 增强容易混淆类别之间的边界信息。

其次，APF 基于 OPE 输出的稳定原型进行类别相关性分析，通过自适应反馈策略强化具有判别价值的类别特征。如果缺少 OPE，APF 需要直接处理未校正的局部原型，容易将数据偏移引入的噪声误认为有效类别信息；反之，仅使用 OPE 只能改善原型分布，无法充分优化类别间边界。

最后，由于联邦环境中不同客户端的数据分布存在不可避免的异构性，仅依赖客户端优化仍可能产生少量异常知识。因此，服务器端 NF 利用全局原型分布建模能力，对上传原型进行概率一致性评估，并过滤低概率异常样本。与客户端侧方法相比，NF 能够利用全局视角发现跨客户端异常模式；但其无法修正客户端生成阶段已经产生的低质量表示。因此，OPE、APF 和 NF 分别针对 FCL 中的原型可靠性、类别判别性以及全局一致性三个关键问题，形成由局部到全局、由表示优化到知识筛选的递进式信息处理链条。

4 理论分析

4.1 统一优化目标与联邦持续学习训练形式

在 FCL 中，系统由 N 个客户端组成。第 k 个客户端在第 t 轮通信中使用当前任务 r 对应的数据集 D_k^r，并维护一个本地记忆缓冲区 M_k 存储少量历史样本用于回放。服务器维护全局模型参数 θ^t，客户端本地模型参数记为 θ_k^r。在第 t 轮的本地训练过程中，客户端 k 的训练小批次由当前任务样本和回放样本组成，记为：

$$B_k^t = X \cup X^b \tag{27}$$

其中，$X \subset D_k^r$ 为当前任务样本，$X^b \subset M_k$ 为从记忆库中回放的样本，B_k^t 表示客户端 k 在第 t 轮通信中构造的训练批次。在该设定下，客户端本地的优化目标不再是单一的监督分类风险最小化，而是同时兼顾新任务学习和旧知识保持的复合目标。

为此，本文在客户端侧采用分类监督原型层面对比和实例层面对比的联合损失，形式化定义为：

$$L_k^t(\theta) = L_{cls}(\theta; B_k^t) + \lambda_{SupCon} L_{SupCon}(\theta; B_k^t) + \lambda_{OPE} L_{OPE}(\theta; B_k^t) + \lambda_{Diff} L_{Diff}(\theta; B_k^t)$$

其中，$\lambda_{SupCon}, \lambda_{OPE}, \lambda_{Diff}$ 分别为监督对比损失、OPE 损失和蒸馏损失的权重系数。

完成本地训练后，客户端上传更新后的参数 θ_k^t。服务器在第 t 轮对采样客户端集合 S_t 进行加权聚合：

$$\theta^t = \sum_{k \in S_t} \left(\frac{n_k^t}{\sum_{j \in S_t} n_j^t} \right) \cdot \theta_k^t \tag{29}$$

其中，n_k^t 表示客户端 k 在第 t 轮的参与训练样本量，$S_t \subseteq \{1, \ldots, N\}$ 表示第 t 轮实际参与本地训练与模型聚合的客户端集合。该聚合确保数据量大的客户端对全局更新贡献更大，从而符合联邦优化中的标准加权原则。

4.2 联邦优化的稳定性分析：统计异质性与梯度漂移缓解

在 Non-IID 的联邦环境中，不同客户端的局部目标函数存在明显差异。记客户端 k 在第 t 轮的本地目标函数为：

$$F_k^t(\theta) = \mathrm{E}_{(x,y) \sim D_k^r} \left[\ell(\theta; x, y) \right] \tag{30}$$

全局目标可抽象为加权求和：

$$F^t(\theta) = \sum_{k=1}^{N} \left(\frac{n_k^t}{n^t} \right) F_k^t(\theta), n^t = \sum_{k=1}^{N} n_k^t \tag{31}$$

经典联邦学习的不稳定来源通常归结为梯度漂移，即不同客户端在本地更新时沿着各自的梯度方向优化，从而造成聚合后的更新方向偏离全局最优下降方向。该现象在持续学习中会进一步加剧，因为每一轮 t 的目标函数本身随任务变化而不断变化。

在每轮本地训练过程中，客户端通常使用随机小批次梯度估计真实梯度，为进行理论讨论，本文采用联邦非凸优化中常见的标准条件。

在联邦非凸优化中，利普希茨光滑性（L-smoothness）、梯度方差有界性以及客户端梯度偏差有界性是收敛性分析的三个标准条件，在大量研究中得到广泛应用与验证[5,39,40]。本文在此基础上，结合 FCL 的任务更迭特性，引入以下假设：

假设 1：全局目标 $F^t(\theta)$ 在参数空间内 L 利普希茨光滑（L-Lipschitz smooth），每个客户端本地的损失函数的梯度是 L 利普希茨连续的。

$$\left\|\nabla F(\theta_1)-\nabla F(\theta_2)\right\|_2 \le L\left\|\theta_1-\theta_2\right\|_2 \quad (32)$$

其中，θ_1,θ_2 为任意两组模型参数，L 为光滑常数，当模型参数发生小幅变化时，目标函数的梯度不会出现剧烈震荡，从而使得基于梯度下降的优化过程具有一定的稳定性。

假设 2：本地随机梯度方差有界；记 $g_k^t(\theta)$ 为客户端 k 在参数θ处的随机梯度估计，$\nabla F_k^t(\theta)$ 表示本地目标函数的精确梯度，由于客户端训练使用小批量数据进行随机梯度下降，因此 $g_k^t(\theta)$ 与 $\nabla F_k^t(\theta)$ 之间存在采样噪声。本文假设这种随机梯度噪声的二阶矩是有界的：

$$\mathrm{E}\left[\left\|g_k^t(\theta)-\nabla F_k^t(\theta)\right\|_2\right] \le \sigma^2 \quad (33)$$

其中，σ^2 为梯度方差上界，期望 $\mathrm{E}[\cdot]$ 对应于随机抽样过程的随机性，该假设意味着：尽管随机梯度会引入不确定扰动，但其波动幅度受到控制，不会无限增大。换而言之，客户端的局部更新不会因为极端的梯度噪声而变得完全不可预测，从而保证了训练过程在统计意义上的可分析性与可重复性。

假设 3：在 Non-IID 数据条件下，不同客户端的本地目标函数 $F_k^t(\theta)$ 对应的梯度方向可能存在显著差异。这种差异会导致经典 FedAvg 出现客户端漂移（client drift）现象，即各客户端在本地训练后更新方向偏离全局优化方向，从而降低聚合后边的有效下降程度。为刻画这种统计异质性强度[39]，本文进一步假设客户端梯度与全局梯度之间的偏差存在期望意义下有界：

$$\mathrm{E}\left[\left\|\nabla F_k^t(\theta)-\nabla F^t(\theta)\right\|_2\right] \le \delta^2 \quad (34)$$

其中，δ^2 表示客户端梯度异质性，衡量了客户端间的异质性程度。当 δ^2 较小，说明各客户端的目标函数相近，本地梯度方向与全局梯度一致性较高，此时联邦聚合更稳定；反之，当 δ^2 较大，则说明客户端数据分布差异显著，全局模型更容易受到冲突更新影响，从而出现收敛速度下降或性能震荡等问题。

在上述假设条件下，本文进一步推导 PAF-FCL 框架在异构 FCL 场景中的收敛特性。记第t轮通信后全局模型为θ^t，全局目标函数为$F(\theta)=\sum_k p_k F_k(\theta)$，其中$p_k=n_k/\sum_k n_k$为客户端$k$的权重。设定每轮本地更新次数为$E$，学习率为$\eta$，则有以下结论：

定理 1（收敛界）在假设 1-3 成立且学习率满足$\eta \le 1/L$的条件下，PAF-FCL 经过t轮通信后，全局模型梯度的期望范数满足：

$$\frac{1}{T}\sum_{t=0}^{T-1}\mathrm{E}\|\nabla F^t(\theta^t)\|_2 \le \mathrm{O}\left(\frac{\sigma^2}{\sqrt{T}}+\frac{\delta^2}{T}+\min(\delta_{\mathrm{NF}}^2,\delta^2)\right)$$

其中，σ^2为梯度方差上界，δ^2为统计异质性上界，δ_{NF}^2为NF准确遗忘机制对冲突原型筛选后引入的额外方差项，满足$\delta_{\mathrm{NF}}^2 \le \delta^2$，NF 筛选导致的梯度偏差：$\epsilon_{NF}^t=\nabla F^t(\theta^t)-\nabla \tilde{F}^t(\theta^t)$，并且$\mathbb{E}\|\epsilon_{NF}^t\|_2 \le \delta_{\mathrm{NF}}^2$。

在 Non-IID 条件下，PAF-FCL 的梯度范数期望平均值随通信轮次R增大而收敛，其收敛速率为$\mathcal{O}(1/\sqrt{T})$，这表明引入 OPE、APF 及 NF 模块不会破坏联邦优化的收敛性质。由于 NF 准确遗忘机制能够筛除低相关性原型，从而有效压缩了异质性导致的恒定误差项。

证明：

设第t轮 NF 筛选后的有效全局梯度估计为$\tilde{g}^t$，客户端k上传的本地随机梯度为g_k^t。

由假设 1 可知：

$$F^t(\theta^{(t+1)}) \le F^t(\theta^t)-\eta\langle\nabla F^t(\theta^t),\tilde{g}^t\rangle+\frac{L\eta^2}{2}\|\tilde{g}^t\|_2$$

将$\tilde{g}^t$拆解为全局精确梯度与误差项之和：$\tilde{g}^t=\nabla F(\theta^t)+\epsilon^t+\epsilon_{NF}^t$，其中$\epsilon^t$表示由随机采样误差，$\epsilon_{NF}^t$表示 NF 筛选造成的残余知识偏差。

则：$\mathbb{E}[F^t(\theta^{t+1})] \le F^t(\theta^t)-\eta\|\nabla F^t(\theta^t)\|_2+\eta\mathbb{E}\|\epsilon^t+\epsilon_{NF}^t\|_2+\frac{L\eta^2}{2}\mathbb{E}\|\tilde{g}^t\|_2$

根据假设 2：$\mathbb{E}\|\epsilon^t\|_2 \le \sigma^2$，根据 NF 筛选机制 $\mathbb{E}\|\epsilon_{NF}^t\|_2 \le \delta_{\mathrm{NF}}^2$，结合假设 3 可得：$\mathbb{E}\|\epsilon^t+\epsilon_{NF}^t\|_2 \le \sigma^2+\min(\delta_{\mathrm{NF}}^2,\delta^2)$

代入得：

$$\mathrm{E}[F^t(\theta^{(t+1)})] \le F^t(\theta^t)-\frac{\eta}{2}\|\nabla F^t(\theta^t)\|_2+\frac{\eta}{2}\left(\sigma^2+\min(\delta_{\mathrm{NF}}^2,\delta^2)\right) \quad (37)$$

对 $t=0,\ldots,T-1$ 求和，利用伸缩和 $\frac{\eta}{2}\sum_{t=0}^{T-1}\mathbb{E}\|\nabla F^t(\theta^t)\|_2 \le F^0-F^*+\frac{\eta T}{2}(\sigma^2+\min(\delta_{\mathrm{NF}}^2,\delta^2))$，两边除以为$\frac{\eta T}{2}$：

$$\frac{1}{T}\sum_{t=0}^{T-1}\mathrm{E}\|\nabla F^t(\theta^t)\|_2 \le \frac{2(F^0-F^*)}{\eta T}+\sigma^2+\min(\delta_{\mathrm{NF}}^2,\delta^2) \tag{38}$$

令$\eta=\mathcal{O}(\frac{1}{\sqrt{T}})$

得：

$$\frac{1}{T}\sum_{t=0}^{T-1}\mathrm{E}\|\nabla F^t(\theta^t)\|_2 \le \mathrm{O}\left(\frac{\sigma^2}{\sqrt{T}}+\frac{\delta^2}{T}+\min(\delta_{\mathrm{NF}}^2,\delta^2)\right)$$

当 NF 筛选足够有效、剔除大部分离群原型时，$\delta_{\mathrm{NF}}^2 \ll \delta^2$，收敛界中的恒定误差项由$\delta^2$显著降至$\delta_{\mathrm{NF}}^2$。

在 FCL 中，δ^2随任务步r的推进呈阶段性增长，而基于 NF 的准确遗忘机制在每一轮通信中动态更新原型分布估计并剔除离群特征，使得δ_{NF}^2的增长速度显著低于δ^2，从而在多任务序列中维持相对稳定的收敛状态。

5 实验

5.1 实验数据

本文在 FCL 的场景下评估所提出模型的有效性。实验主要采用 CIFAR-100 作为主数据集，并使用 CIFAR-10 进行小规模验证。为构造类增量学习任务流，在实验设置中，CIFAR-100 的 100 个类别被划分为 10 个连续任务，其中每个任务包含 10 个不同且互不重叠的类别；CIFAR-10 则划分为 5 个任务，每个任务包含 2 个类别。Tiny-ImageNet 的 200 个类别划分为 10 个任务，每个任务包含 20 个类别。联邦系统包含 N 个客户端（N=10），各数据集的信息汇总见表 1，客户端之间的数据分布为 Non-IID。具体而言，采用 label-skew 与 quantity-skew 的组合来模拟统计异质性：每个客户端仅覆盖全局类别集合中的一部分，且不同客户端持有样本数量不均衡，从而形成类别覆盖与数据规模的双重偏斜。

在每轮通信中，服务器以采样比例 ρ=0.5 随机选取部分客户端参与训练。参与的客户端利用其当前可见数据执行本地模型更新，并将模型参数上传至服务器进行聚合。聚合完成后，服务器再将更新后的全局模型广播至客户端，以启动下一轮训练。

表 1 实验数据划分

数据集	总类别数	任务数	每任务类别数
CIFAR-10	10	5	2
CIFAR-100	100	10	10
Tiny-ImageNet	200	10	20

5.2 实验设置

模型结构方面，客户端采用特征提取器、投影头、分类器的判别架构：特征提取器使用 ResNet-18 作为骨干网络，投影头为轻量线性映射层，分类器采用单头（single-head）设置以满足 CIL 评估要求，即在测试时不提供任务标识，模型需要在所有已见类别上进行统一预测。训练过程中，客户端不仅使用交叉熵损失学习当前任务的判别边界，同时引入原型相关的表示约束以提升类内紧凑性与类间可分性，从而降低在 Non-IID 与持续学习叠加场景下对快捷方式特征的依赖。

为缓解遗忘，客户端维护一个固定容量的本地记忆库，用于从历史样本中抽取回放小批次与当前任务小批次混合训练；记忆库采用在线更新策略，即持续从当前数据流中抽取样本写入，当容量已满时随机替换旧样本，以保证在不保存完整历史数据的前提下仍能保留有限的历史信息。

为保证不同数据集下训练预算一致，本文对所有数据集统一采用相同训练：在每个任务阶段执行 T=100 轮通信；每轮中每个参与客户端进行 E=5 次本地更新，数据增强采用常用的图像增强方法，用于提升表示学习的稳健性。在优化与联邦训练流程上，客户端使用 Adam 优化器进行本地更新，学习率初始设置为 0.01，全局通信服务器采用加权平均方式对参与客户端上传的模型进行聚合，并在每轮通信后广播最新全局模型；在 CIFAR-10 和 CIFAR-100 实验中设置固定的通信轮数与本地迭代预算，以保证不同方法在相同通信约束下进行公平对比。

本文选取的基准算法包括联邦学习经典算法 FedAvg[5]和 FedProx[39]，以及联邦持续学习算法 FedWeIT[8]、GLFC[11]、FedProTIP[40]、FedTA[41]和 FEAT[42]。

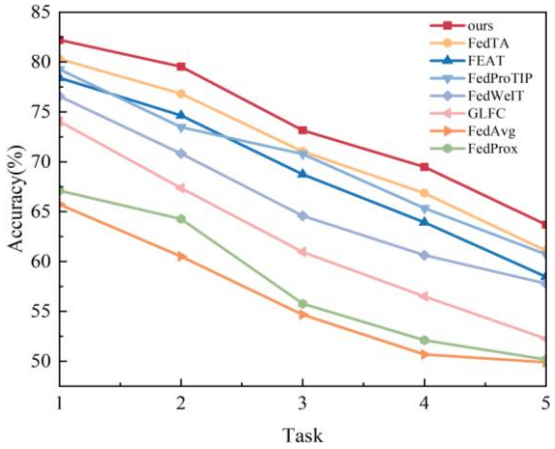
实验评价指标包括分类准确率（Accuracy, ACC）和平均遗忘率（Forgetting Rate, FT）。平均遗忘是在连续学习过程中评估后向迁移的程度，量化为每项任务的峰值准确率和结束准确率之间的差距。

5.3 实验分析

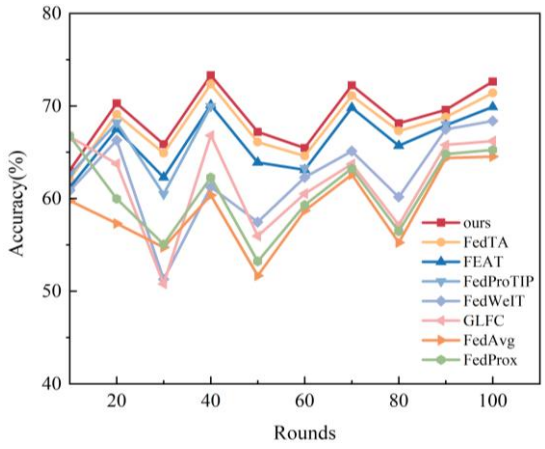
5.3.1 准确率对比

为了验证 PAF-FCL 的性能,设计实验将其与基线模型进行了对比。图 1~3 展示了在 CIFAR-10 数据集上，设置$\alpha=\{0.1,0.3,0.5\}$三种不同异构场景，α越小，表示数据异构程度越强烈。所提出的 PAF-FCL 框架与各基准方法在不同任务阶段和通信轮次下的准确率对比。整体而言，PAF-FCL 在平均准确率和训练稳定性方面均表现出显著优

势，验证了原型引导与准确遗忘机制在异构联邦环境下的有效性。



（a）不同任务下各模型准确率对比



（b）不同轮次下各模型准确率对比

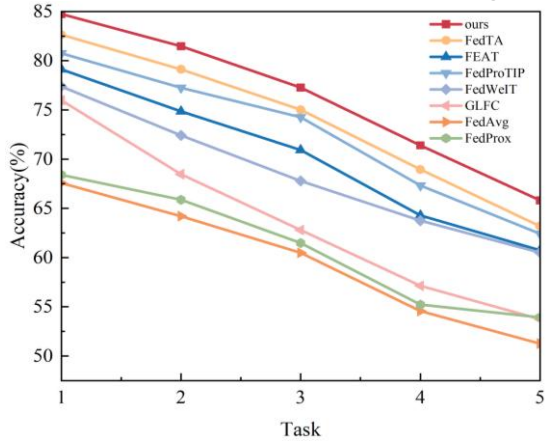
图 1　在$\alpha = 0.1$、CIFAR-10 数据集下的准确率

在较强异质性条件下，各模型整体表现相对稳定，但随着任务数的增加，均呈现不同程度的性能下降趋势。结果如图 1 所示，在任务由 1 逐步增加至 5 的过程中，所有方法的分类准确率均呈现单调下降趋势，这反映出持续学习过程中典型的“灾难性遗忘”现象。

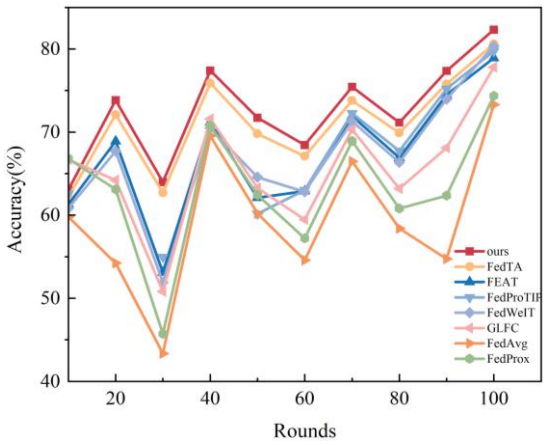
相比之下，本文方法在整个任务序列中始终保持最高准确率，且下降幅度相对较小。FedTA 与 FEAT 在任务初期具有较强竞争力；Task 1 时准确率分别达到 79%与 78%，略高于部分传统 FCL 方法。但随着任务序列的推进，其在知识保持方面出现更明显的衰减，例如，在第 5 个任务时，本文方法仍保持领先优势，而 FedAvg 与 FedProx 下降最为明显。

此外，FedProTIP、FedWeIT 与 GLFC 等改进方法在一定程度上缓解了遗忘问题，但其准确率曲线仍明显低于本文方法。这表明在低异质性场景下，本文方法在知识保持与新任务适应之间取得了更优的平衡。如图 1（b）所示，在不同通信轮次下，各模型准确率均随训练轮次增加呈波动上升趋势。

本文方法在整个训练过程中始终位于性能曲线的上方，并在后期轮次保持最优表现。相比之下，FedAvg 与 FedProx 在早期轮次波动较大，收敛速度相对较慢。



（a）不同任务下各模型准确率对比

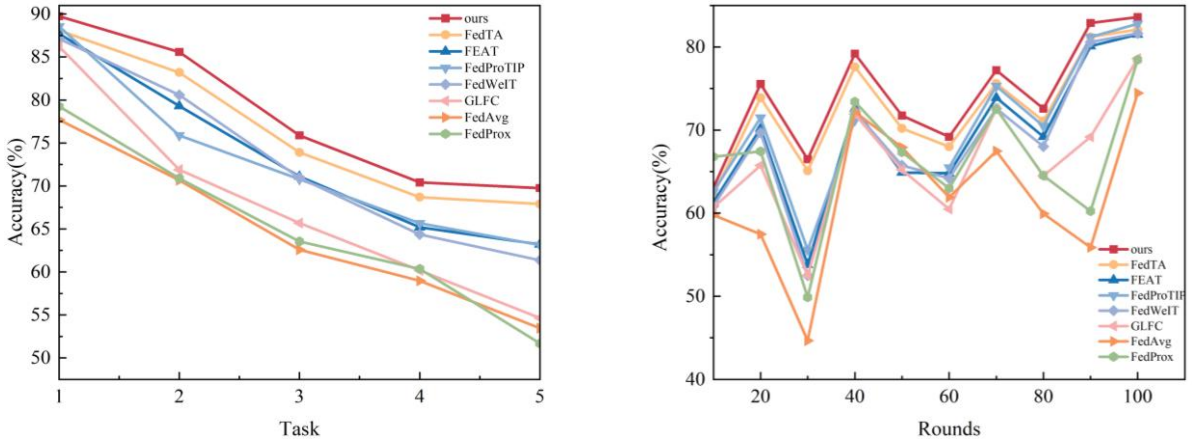


（b）不同轮次下各模型准确率对比

图 2　在$\alpha = 0.3$、CIFAR-10 数据集下的准确率

当α增大至 0.3 时，客户端间的数据分布差异有所缓和，但仍属于 Non-IID 环境。实验结果如图 2 所示，可以明显观察到任务递增带来的性能衰减趋势仍然存在，但下降幅度较$\alpha = 0.1$时有所缓解。

与强异构场景相比，各模型之间的性能差距在中等异构条件下更加清晰。本文方法在每一个任务阶段均取得最高准确率，且在后期任务阶段仍能保持较高性能水平。FedTA 与 FEAT 在中等异构条件下表现稳健，整体准确率曲线位于传统 FCL 方法之上，但在知识保持与新任务适应的平衡上不及 PAF-FCL。这说明该方法在处理中等程度分布偏移时，能够有效保持旧任务知识并快速适应新任务，从图 2(b)通信轮次变化趋势来看，各模型在早期训练阶段波动幅度依然较大，但随着轮次增加逐渐趋于稳定。

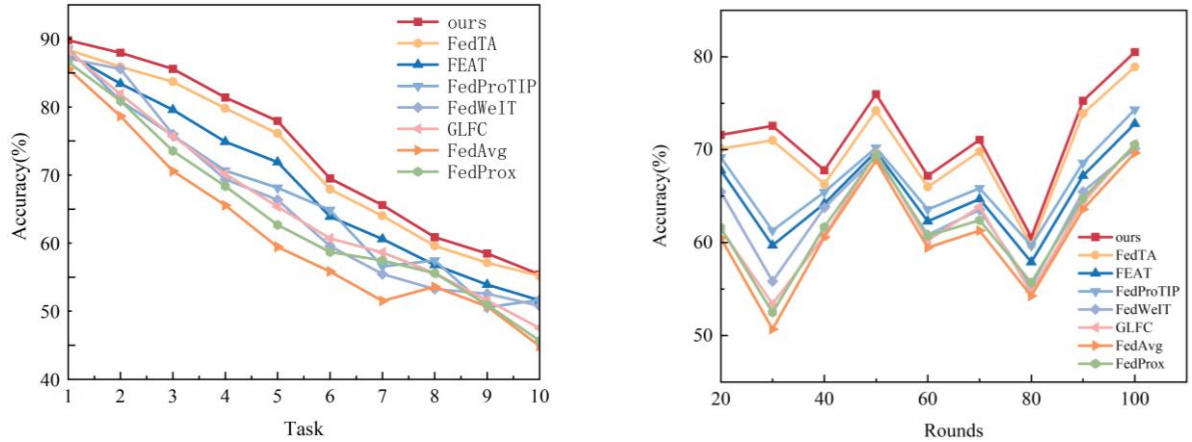


（a）不同任务下各模型准确率对比　　（b）不同轮次下各模型准确率对比

图 3 在$\alpha = 0.5$、CIFAR-10 数据集下的准确率

根据图 3 展示的实验结果，本文提出的 PAF-FCL 框架在 CIFAR-10 数据集上展现出了显著优于对比基准的性能。如图 3（a）所示，随着连续任务数从 1 增加到 5，所有方法均表现出不同程度的灾难性遗忘，即平均准确率呈下降趋势，但 PAF-FCL 的下降幅度最小且全程保持领先。这种抗遗忘能力的优势主要归因于 PAF-FCL 在客户端与服务器端的双重优化策略：在客户端，引入的 OPE 与 APF 机制使得本地模型能够提取出类内紧凑、类间分离的特征表示，这种高判别性的特征有效抑制了快捷方式学习，增强了模型对旧知识的记忆稳健性；与此同时，在服务器端，针对异构联邦环境下的特征偏差问题，基于 NF 的准确遗忘机制发挥了关键作用。通过对上传的原型进行概率建模与相关性评估，PAF-FCL 成功识别并剔除了低置信度的噪声特征，从而避免了跨任务过程中的错误记忆累积，确保了全局模型在任务序列延伸时的性能稳定性。

进一步观察图 3（b）中随通信轮次变化的准确率曲线，可以发现模型性能呈现出与任务切换周期相对应的波动特征。在每个新任务引入的初期，由于数据分布的突变，所有模型的准确率均会出现短暂下跌，FedTA 的收敛曲线在局部轮次接近本文方法，但波动幅度较大，与其最优全局原型选择策略在每轮参与客户端采样不同时产生的选择不稳定性有关。FEAT 的收敛过程较为平滑，但上升斜率平缓，其等角紧帧的强几何约束虽抑制了震荡，却也在一定程度上限制了模型对新任务分布的快速适应能力。但 PAF-FCL 均能实现最迅速的性能回升并达到更高的收敛峰值。这种快速适应能力得益于 APF 机制对易混淆类别的针对性增强，使模型能够在有限的通信轮次内迅速修正决策边界。



（a）不同任务下各模型准确率对比　　（b）不同轮次下各模型准确率对比

图 4　在$\alpha = 0.5$、CIFAR-100 数据集下的准确率

图 4 进一步展示了在更复杂的 CIFAR-100 数据集上，各模型在包含 10 个连续任务的长序列场景中的性能表现。与 CIFAR-10 相比，CIFAR-100 不仅类别总数增加了十倍，任务划分也更加细粒度，这对模型在特征空间的判别能力与抗遗忘能力提出了严峻挑战。

然而，PAF-FCL 在任务后期表现出了显著的鲁棒性优势。在任务 1 到任务 5 的初期阶段，PAF-FCL 与其他基准方法保持了相近的性能，但在任务 6 之后，随着累积的历史知识变得愈发庞杂，FedTA 在 Task 1 至 Task 3 的初期阶段与本文方法差距微小，但自 Task 4 起衰减加速，其原因是 Tail Anchor 依赖全局最优原型选择来确定特征混合的参考锚点，CIFAR-100 拥有 100 个细粒度类别，类别间语义重叠显著增加，Tail Anchor 将基于混淆的锚点进行特征混合，导致旧类决策边界在多任务推进中被渐进侵蚀。

传统方法如 FedAvg 和 GLFC 因无法有效处理新旧任务间的特征干扰，准确率出现加速衰退，最终跌至 50%以下。图 4（b）描绘了 CIFAR-100 训练过程中随通信轮次变化的准确率曲线，该曲线不仅反映了模型的收敛速度，还揭示了其在 Non-IID 环境下的适应稳定性。

表 2 展示了在中度异构环境下，各模型在 CIFAR-100 数据集 10 个连续任务上的准确率演变过程。实验数据表明，随着任务数量从 Task 1 增加到 Task10，所有对比算法的准确率均出现了明显的下滑，其中 FedAvg 从初始的 63.69%跌至最终的 34.98%，降幅高达 45%，凸显了长序列 FCL 中灾难性遗忘的严重性。

然而，本文提出的 PAF-FCL 在所有阶段均保持了最优性能，特别是在任务末期，仍能维持 46.61%的准确率，CIFAR-100 的细粒度类别特征使得传统方法极易产生类间干扰，而 PAF-FCL 则通过 OPE 维持特征流形的均衡结构，配合 APF 在线感知混淆并定向增强，在任务切换后实现更迅速的性能回升。

表 2 CIFAR-100、$\alpha = 0.1$数据集上的准确率（%）对比

Method	Task									
	1	2	3	4	5	6	7	8	9	10
FedAvg	63.69	59.53	51.49	50.12	45.70	43.21	40.20	39.87	36.47	34.98
FedProx	67.48	63.29	56.10	51.52	46.22	42.99	40.67	39.31	37.96	35.35
GLFC	65.78	61.82	57.02	54.26	46.70	43.19	41.16	39.60	39.54	38.51
FedWeIT	72.19	63.40	61.34	56.93	51.71	50.13	44.57	40.10	39.50	37.06
FedProTIP	75.49	64.25	56.70	53.67	52.12	47.25	45.54	43.79	38.46	37.40
FEAT	74.14	68.06	64.22	60.16	55.43	52.87	49.32	46.7	44.68	43.13
FedTA	76.81	71.52	67.97	63.48	59.82	57.70	54.37	51.85	49.74	48.22
Ours	79.21	66.80	63.65	57.37	55.16	52.46	50.44	48.28	47.36	46.61

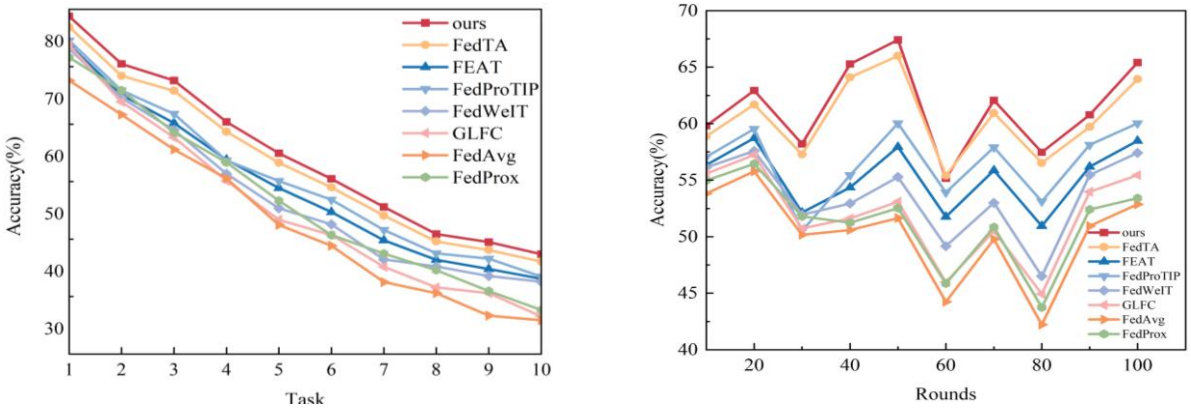
表 3 CIFAR-100、$\alpha = 0.3$数据集上的准确率（%）对比

Method	Task									
	1	2	3	4	5	6	7	8	9	10
FedAvg	65.20	61.24	57.17	55.42	50.20	44.34	41.72	38.17	37.98	36.04
FedProx	68.10	65.34	58.36	55.82	51.22	48.79	42.78	40.36	38.58	37.31
GLFC	67.31	64.40	60.12	57.48	50.79	49.26	45.54	41.81	40.93	37.29
FedWeIT	75.16	70.03	64.47	60.83	57.02	55.53	49.44	48.39	40.73	39.16
FEAT	76.48	72.68	68.76	64.37	60.82	57.95	54.87	52.65	49.85	47.96
FedTA	78.61	74.94	71.25	66.90	63.55	60.87	57.43	54.93	52.67	51.40
FedProTIP	77.20	75.16	69.41	64.39	59.63	56.60	50.37	49.24	44.33	43.76
Ours	80.21	77.52	75.94	69.37	65.50	63.71	59.34	55.73	51.13	50.32

在数据异构性较低的环境下，各模型的表现如表 3 所示。相比于表 2，各方法的整体准确率有所提升，但 PAF-FCL 的领先优势依然显著。从数据趋势看，PAF-FCL 从 Task 1 的 80.21%到 Task 10 的 50.32%，其准确率衰减曲线最为平缓。值得注意的是，在 Task 3 到 Task 7 的中期阶段，PAF-FCL 的领先幅度进一步扩大，这主要得益于 APF 机制。在异构联邦环境中，由于各客户端可见类别不均，APF 能够实时监测本地模型对特定类别的认知偏差，并通过原型混合增强技术进行定向补偿。这种针对性的特征修复机制，使得全局模型在聚合时能够获得更完整的类别分布视图，有效地抑制了因 Non-IID 导致的模型漂移。

由于 CIFAR-100 的任务切换频率较高，曲线

呈现出密集的震荡形态。观察可知，PAF-FCL 在每个新任务引入后的准确率恢复阶段均表现出更陡峭的上升斜率，这归功于客户端引入 OPE 与 APF 机制。面对 CIFAR-100 中大量细粒度且易混淆的类别，APF 机制能够针对性地增强困难类别的特征边界，从而加速了本地模型对新知识的同化速度。



（a）不同任务下各模型准确率对比　　（b）不同轮次下各模型准确率对比

图 5　在$\alpha = 0.5$、Tiny-ImageNet 数据集下的准确率

为了进一步验证 PAF-FCL 框架在处理高维图像数据与大规模类别增量场景下的可扩展性，本文在 Tiny-ImageNet 数据集上进行了 10 个连续任务的评估。如图 5（a）所示，Tiny-ImageNet 包含 200 个类别且图像分辨率更高，这对特征提取器的泛化能力与模型的记忆容量提出了更为严苛的要求。实验结果显示，随着任务序列的推进，所有对比方法的性能衰减幅度均明显大于 CIFAR 数据集，这反映了任务复杂度的提升加剧了灾难性遗忘现象。然而，PAF-FCL 依然展现出了最优的渐进性能。

图 5（b）展示了在 Tiny-ImageNet 训练过程中测试集准确率随通信轮次的变化情况。由于每个任务包含 20 个新类别，在任务切换节点，模型面临着剧烈的知识重组压力，导致曲线呈现出显著的锯齿状波动。观察 PAF-FCL 的训练轨迹可以发现，其在每个新任务阶段的准确率回升速度与收敛峰值均显著优于 FedProTIP 和 FedWeIT 等先进基线方法。例如在第 40 轮至 50 轮的区间内，PAF-FCL 迅速适应了新任务分布并达到了约 65% 的局部峰值，而其他方法则表现出明显的滞后性。这表明客户端引入的 OPE 与 APF 机制在复杂图像特征提取中发挥了关键作用，通过在本地增强易混淆类别的特征表示，模型能够以更少的通信轮次缓解冷启动问题。

表 4　Tiny-ImageNet、$\alpha = 0.1$数据集上的准确率（%）对比

Method	Task									
	1	2	3	4	5	6	7	8	9	10
FedAvg	60.40	52.55	46.38	44.82	39.34	35.92	32.82	30.53	32.64	32.35
FedProx	63.16	54.74	55.41	49.20	45.46	40.05	38.38	35.82	34.39	34.41
GLFC	63.78	55.37	54.22	48.38	45.33	41.07	39.69	36.93	35.17	35.85
FedWeIT	67.50	56.13	51.70	48.94	45.42	42.29	41.54	40.61	38.11	37.45
FedProTIP	69.26	59.27	57.32	51.80	47.02	46.75	43.15	41.86	38.33	37.64
FEAT	68.96	64.53	60.32	55.54	51.49	47.66	45.12	43.26	40.95	39.53
FedTA	69.93	66.37	63.80	58.46	54.15	50.87	48.23	46.06	43.82	42.65
Ours	70.59	65.79	59.51	55.38	49.21	47.40	45.05	43.36	40.93	44.58

Tiny-ImageNet 的 200 个类别划分出的 10 个任务流对模型的全局协调能力提出了极高要求。表 4 的数据显示，当任务进行到中后期，大多数现有算法的性能下降斜率开始变陡，这表明它们在处理多客户端异构特征时，无法有效过滤掉过时的、或是与当前全局分布相冲突的冗余信息。而 PAF-FCL 引入的知识筛选机制，实质上在服务器端构建了一道置信度防线，它能够基于概率统计特征对各客户端上传的更新进行精细化评估。

从数据中可以看出，各基准方法在任务 1 至任务 10 的演进中均出现了严重的性能滑坡。例如，经典的 FedAvg 在任务中后期的准确率已跌至 30% 左右，陷入了深度遗忘的瓶颈。与此形成鲜明对比的是，本文提出的 PAF-FCL 在整个学习生命周期内展现出了极强的韧性。尤其在 Task 10 结束时，本文方法仍能实现 44.58%的分类精度。

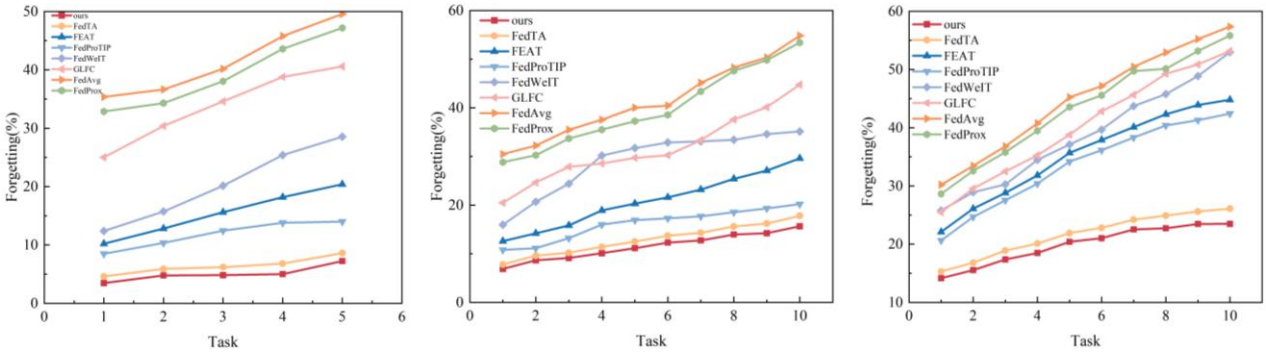
表 5 Tiny-ImageNet、$\alpha = 0.3$数据集上的准确率（%）对比

Method	Task									
	1	2	3	4	5	6	7	8	9	10
FedAvg	61.79	53.16	49.65	47.75	44.99	40.65	35.53	32.63	34.35	35.15
FedProx	64.86	58.44	54.59	50.45	46.67	41.57	36.06	35.79	36.81	36.81
GLFC	63.77	57.81	55.57	51.72	45.41	42.62	40.80	37.51	36.03	37.65
FedWeIT	70.06	60.26	58.89	53.09	49.77	47.28	43.12	40.45	39.34	38.86
FEAT	72.43	66.38	61.84	57.45	53.96	50.84	47.63	44.97	42.85	41.58
FedTA	74.61	69.85	65.92	61.25	57.63	54.28	50.95	47.83	45.65	44.12
FedProTIP	71.65	61.47	59.66	51.80	48.03	46.67	44.88	42.10	40.08	38.41
Ours	76.84	70.78	66.54	62.57	58.73	59.73	52.85	47.68	47.95	45.97

表 5 展示了在优化配置下，各模型在 Tiny-ImageNet 连续 10 个任务中的平均准确率变化情况。实验结果进一步验证了本文提出的 PAF-FCL 框架在大规模类别增量场景下的优越性能。在任务 1 阶段，PAF-FCL 即取得了 76.84% 的高起始准确率，显著领先于 FedAvg 和 FedProTIP，这表明本文方法在初始阶段就能构建出更为稳健的特征判别基准。随着任务序列的推进，虽然所有算法都面临着不断累积的参数更新压力，但 PAF-FCL 的准确率下降斜率最为平缓。在长达 10 个任务的持续学习后，本文方法依然能够保持 45.97%的分类精度，相比于其他基准方法普遍处于 35%-38%之间的水平，展现出了长效记忆维持能力。

5.3.2 遗忘率对比

在持续学习评估体系中，平均准确率虽然能够反映模型在某一时刻的综合性能，但却难以直观地量化模型对历史知识的保护程度。为了深入探究各方法在克服灾难性遗忘方面的有效性，本节引入遗忘率（Forgetting Rate, FT）作为核心评估指标。遗忘率用于衡量模型对历史任务知识的遗忘程度，其定义为各任务在训练过程中达到的最高准确率与完成全部任务学习后该任务最终准确率之间差值的平均值。在 FCL 的环境下，遗忘不仅源于任务间的干扰，还受到客户端之间统计异质性引发的模型漂移的影响。因此，一个稳健的 FCL 框架必须在获取新任务增量知识的同时，最大程度地降低旧任务决策边界的退化。



（a）CIFAR-10 （b）CIFAR-100 （c）Tiny-ImageNet

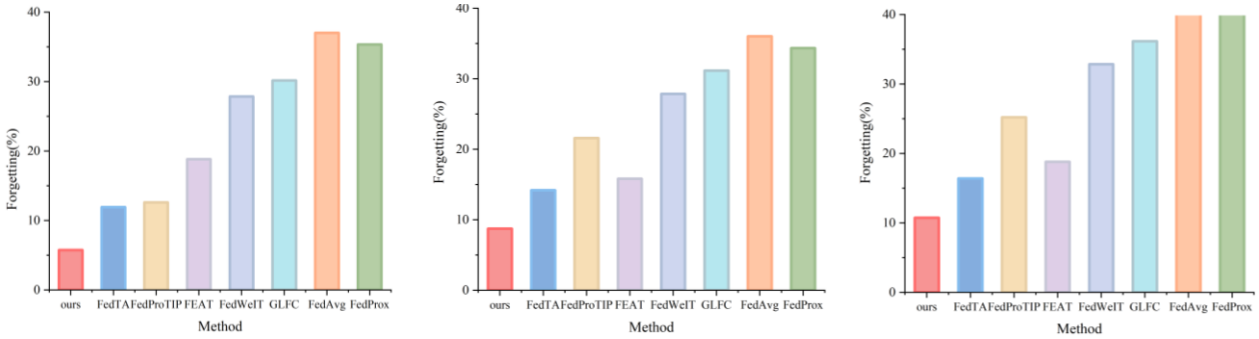
图 6 各类方法在不同任务下的遗忘率对比

随着数据集图像分辨率的提高和类别划分的细化，所有模型的遗忘率总体呈逐步上升趋势。实验结果如图 6 所示，图中给出了各对比方法在 CIFAR-10、CIFAR-100 和 Tiny-ImageNet 数据集上完成全部序列任务后的平均遗忘率。该结果表明，随着特征维度和类别数量的增加，数据分布的异质性以及类别间的潜在混淆程度进一步加剧，新任务引入的分布偏移对旧任务决策边界的干扰更加明显。FedTA 的遗忘率整体受控但上升斜率略陡。FedTA 的 Tail Anchor 机制通过将新类特征与冻结的旧类输出混合来固化旧知识，其遗忘抑制效果取决于全局原型选择的准确性。CIFAR-10 仅有 10 个类别，语义重叠有限，原型选择偏差较小，因此遗忘率绝对值不高。FedAvg 和 FedProx 在三个数据集上均暴露出最高的遗忘率。原因在于这类传统联邦算法仅执行简单的参数平均，缺乏对历史知识的显式保护；在 Non-IID 环境下，异构客户端在新任务上产生的局部梯度极易直接覆写全局模型中属于旧任务的关键权重，客户端漂移加剧了灾难性遗忘。

GLFC、FedWeIT 以及 FedProTIP 等方法由于引入了知识蒸馏、参数隔离或原型对齐等机制，其遗忘率较 FedAvg 有明显下降。然而，在 CIFAR-100 和 Tiny-ImageNet 的复杂任务流中，FEAT 的遗忘率在 Task 1 时为 21.5%，至 Task 10 升至 42.0%，FEAT 的 ETF 机制在复杂类别结构

下负面效应的分析——类别数越多、类别间语义差异越不均匀，等距假设与真实分布的偏离越严重，遗忘抑制的代价越大。由于这些方法难以精准甄别客户端上传特征中的噪声与冲突，导致其构建的全局记忆库仍会随任务推移而被部分“污染”，遗忘率依然保持在较高水平。

本文提出的模型在三个数据集上均实现了相对最低的遗忘率。PAF-FCL 并非绝对杜绝了知识遗忘，而是将其控制在了更为稳健的范围内。产生这一结果的核心原因在于模型在本地与全局实施的双重约束机制：一方面，在客户端侧，通过 OPE 与 APF，模型学习到了更加紧致且边界清晰的原型特征，这种高判别性的表示在后续任务更新时具有较强的抗干扰能力；另一方面，在服务器侧，面对不可避免的异构特征冲突，基于 NF 的准确遗忘机制发挥了屏障作用。



（a）CIFAR-10　　（b）CIFAR-100　　（c）Tiny-ImageNet

图 7　各类方法在$\alpha = 0.1$、不同数据集下的遗忘率对比

为了全面评估不同 FCL 算法在长序列任务中的知识保持能力，本文在 CIFAR-10、CIFAR-100 以及 Tiny-ImageNet 三个复杂度递增的数据集上进行了平均遗忘率的对比实验。实验结果如图 7 所示。在对比分析中，传统 FL 方法 FedAvg 和 FedProx 表现出明显的性能脆弱性。由于缺乏有效的知识锚定机制，这两个模型在处理后续任务的本地更新时，往往会产生剧烈的参数偏移，导致其遗忘率在所有数据集中均处于最高位。尽管 GLFC 和 FedWeIT 等基线通过引入外部特征参考或参数解耦策略在一定程度上缓解了知识丢失，但在面对 CIFAR-100 这种任务跨度较大的场景时，其内部仍然由于无法有效辨别客户端上传信息的置信度，而导致旧任务的表征被潜在的噪声数据干扰。

PAF-FCL 在三组对比实验中均展现了较强的记忆留存表现，其遗忘率曲线的上升幅度较为平缓。这一优势的本质在于模型改变了被动保存的逻辑，转而通过对特征质量的主动管控来降低遗忘风险。具体的实验统计结果如表 6 所示。

表 6　各方法在不同数据集和α上的准确率和遗忘率（%）对比

Method	CIFAR-10				CIFAR-100				Tiny-ImageNet			
	$\alpha = 0.5$		$\alpha = 0.3$		$\alpha = 0.5$		$\alpha = 0.3$		$\alpha = 0.5$		$\alpha = 0.3$	
	ACC	FT	ACC	FT	ACC	FT	ACC	FT	ACC	FT	ACC	FT
FedAvg	63.75	37.16	59.30	33.97	62.80	32.10	57.59	31.23	55.34	34.31	52.66	35.62
FedProx	67.53	39.07	62.74	35.26	63.81	34.13	59.67	33.10	58.25	35.19	53.22	31.37
GLFC	65.21	44.76	63.28	41.73	61.74	31.14	61.20	30.12	59.77	29.75	61.04	27.72
FedWeIT	72.10	37.23	70.36	34.58	67.25	30.53	62.37	27.50	60.62	25.62	61.70	23.11
FEAT	76.24	24.12	72.85	22.92	69.46	24.62	66.87	22.37	63.76	21.24	61.98	20.10
FedTA	82.15	18.40	79.65	17.27	74.85	20.52	72.18	18.92	68.44	17.65	66.81	16.28
FedProTIP	79.42	21.03	75.46	20.02	70.31	25.87	72.97	20.34	69.08	19.03	65.05	15.41
w/o OPE	83.46	10.92	81.52	8.28	76.91	10.34	75.63	9.42	72.03	8.64	70.84	8.51
w/o APF	84.37	11.84	82.47	9.63	77.68	11.62	76.24	10.73	73.02	9.83	71.73	7.76
w/o NF	85.71	8.95	83.96	7.42	79.12	9.58	77.86	8.21	74.21	7.12	72.88	6.21
w/o OA	82.18	13.75	79.17	10.09	75.22	12.10	74.31	11.47	71.16	11.01	69.29	9.36
Ours	87.33	6.76	85.26	6.04	80.62	8.97	79.48	7.08	75.98	5.79	74.49	5.03

为进一步分析各模块对模型性能的独立贡献，本文对 PAF-FCL 中的关键组件进行了系统性的消融实验，分别移除在线原型平衡（w/o OPE）、自适应原型反馈（w/o APF）、服务器端准确遗忘机制（w/o NF）以及移除 OPE 和 APF 模块（w/o OA），并在三个数据集上进行评估，结果如表 6 所示。对比单独移除 OPE 与单独移除 APF 的结果，二者性能接近，说明 OPE 与 APF 在特征学习中存在互补关系——OPE 维护全局结构稳定，APF 在此基础上有针对性地修复局部混淆，单独保留任一模块均可提供基础判别能力，但均不完整。

单独移除 OPE 和 APF 后，模型准确率分别降至 75.63%和 76.24%，相较完整模型 79.48%分别下降 3.85 和 3.24 个百分点；同时移除 OPE 与 APF 后，准确率进一步降至 74.31%，下降 5.17 个百分点。与此同时，w/o OPE 和 w/o APF 的遗忘率分别为 9.42%和 10.73%，而同时移除 OPE 与 APF 后遗忘率上升至 11.47%，说明客户端侧两个机制同时缺失时，模型的知识保持能力进一步下降。上述结果表明，OPE 主要通过优化客户端原型结构、缓解局部数据分布偏移来提高表示的稳定性，APF 则进一步针对类别间混淆增强判别信息，二者分别从原型结构优化和类别判别增强两个层面发挥作用。

进一步地，由表 6 可知，移除 NF 后，模型准确率由 79.48%下降至 77.86%，下降 1.62 个百分点，同时遗忘率由 7.08%上升至 8.21%，表明服务器端 NF 通过全局原型一致性筛选能够有效降低异常知识对模型的干扰。综合上述消融结果，OPE、APF 和 NF 分别作用于原型结构、类别判别以及跨客户端异常知识过滤三个层面。客户端侧的 OPE 与 APF 共同改善原型表示质量和类别判别能力，服务器端 NF 进一步从全局视角筛除低可靠性原型，三者具有明确的功能互补关系，共同提升了模型在联邦持续学习过程中的准确率和知识保持能力。

从实验结果的整体分布来看，PAF-FCL 及其消融版本在各项指标上均表现出显著的优势。在平均准确率方面，PAF-FCL 在 Tiny-ImageNet 具备高维特征空间的挑战性任务中，依然能保持 74.49%的准确率，优于 FedAvg 和 FedProTIP。这种跨数据集的稳健表现说明，通过端云协同的优化，模型能够更有效地捕捉跨客户端的共性特征。与此同时，PAF-FCL 在遗忘率的控制上表现卓越，在多个场景下将 FT 指标压缩至 10%以内，最低甚至达到了 5.03%。这一客观数据证明，模型在吸纳新任务知识时，对历史决策边界的保护程度达到了较高水平，而非通过牺牲旧知识来获取新性能。

5.3.3 通信开销

在 FCL 的实际部署场景中，通信开销是评估系统可行性的关键指标，尤其是在网络带宽受限的边缘计算环境下。本节从通信数据量与达到目标精度所需的总通信轮次两个维度对 PAF-FCL 框架的效率进行了深入探讨。

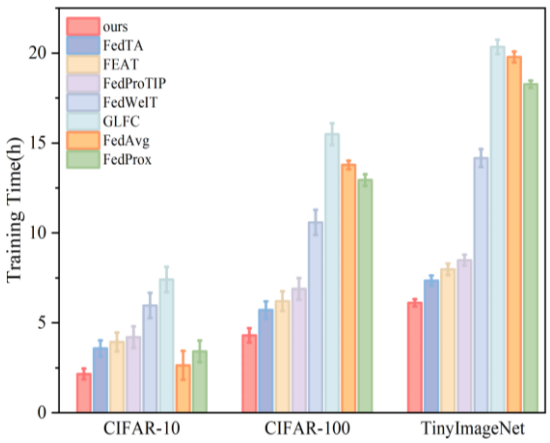
从理论层面分析，联邦学习的通信负担主要来源于模型参数的周期性同步。在本文提出的 PAF-FCL 框架中，除了基础的 ResNet-18 模型参数外，客户端还需向服务器上传各类的原型信息。然而，由于原型实质上是特征空间中的高维均值向量，且数量受限于当前任务的类别数，这使得单轮通信中增加的数据量相比于模型参数量可以忽略不计。

表 7 在 CIFAR-100 数据集下，各方法通信开销对比

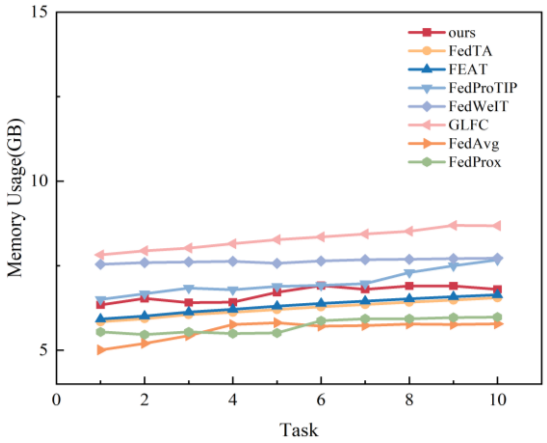
Method	额外载荷（MB）	准确率（%）	轮次	训练时间（h）	累计开销（GB）
FedAvg	-	55.34	150	8.5	6.48
FedProx	-	58.25	145	8.48	4.91
GLFC	4.25	61.04	130	10.49	5.07
FedWeIT	1.78	61.70	115	9.74	4.65
FEAT	1.62	63.48	102	8.65	3.98
FedTA	1.34	67.92	88	7.21	3.31
FedProTIP	1.56	65.05	90	7.74	3.07
Ours	0.64	75.98	70	4.36	2.69

为了评估 PAF-FCL 在资源受限环境下的部署可行性，本文在表 7 中对比了各方法的通信载荷与系统效率。实验结果表明，尽管 PAF-FCL 引入了微量的额外载荷，但其在整体通信效率上具有优势。FEAT 与 FedTA 进一步考虑了异构场景下的任务适应能力，因此在准确率和遗忘控制方面优于 FedWeIT 等传统持续学习方法。其中，FedTA 通过尾部锚点机制增强了全局特征对齐能力，在复杂异构场景下表现出较好的稳定性，因此整体性能接近 FedProTIP。然而，这类方法仍需要频繁同步额外特征信息或锚点参数，导致训练时间与累计通信开销仍然较高。

与 FedWeIT 这种通过牺牲带宽换取性能的方法不同，PAF-FCL 侧重于通过高质量的原型反馈加速全局收敛。得益于 OA 模块对本地特征空间的精准构建，模型仅需 70 轮通信即可达到最优性能，使得累计通信开销较基准方法 FedAvg 降低了约 53%。此外，由于 NF 模块运行于服务器端，并未增加客户端的计算负担，PAF-FCL 在训练时间上依然保持了较好的优势。



（a）各方法在不同数据集下训练时间对比



（b）各方法在 Tiny-ImageNet 数据集下 GPU 内存使用情况

图 8　各方法训练效率对比

为了评估算法的可行性，本文进一步对比了各方法在训练时间与内存占用方面的表现。实验结果如图 8 所示。从图 8(a)可以观察到，PAF-FCL 在三个数据集上的训练耗时均保持在较低水平。虽然 FedAvg 和 FedProx 不涉及任何额外的抗遗忘机制，其训练速度最快，但 PAF-FCL 的耗时仅略高于这两个基础算法，展现出了较高的计算效率。相比之下，GLFC 和 FedWeIT 的训练时间显著增加，尤其在 Tiny-ImageNet 数据集上，GLFC 的耗时最长。这是因为 GLFC 涉及复杂的全局知识蒸馏计算，而 FedWeIT 则在本地训练中包含了繁琐的参数分解与掩码搜索过程。

图 8（b）展示了不同方法在 Tiny-ImageNet 数据集上的 GPU 内存使用情况。从整体趋势可以看出，随着任务数量的增加，各方法的内存占用均呈现缓慢上升趋势。这主要是由于持续学习过程中需要保留历史任务信息或中间表示，从而导致额外的存储开销。FedAvg 与 FedProx 的内存占用最低，由于这类方法仅维护单一模型参数，不涉及额外的历史信息存储或结构扩展，因此内存开销较小。然而，这类方法在持续学习场景下容易出现灾难性遗忘问题。相比之下，FedProTIP、FEAT 以及 FedTA 等方法由于引入原型或特征对齐机制，需要额外存储辅助信息，因此内存开销有所增加。

FedWeIT 与 GLFC 的内存开销相对较高，这是因为它们需要维护多子网络结构或额外的约束信息，从而显著增加了模型规模与存储需求。

本文方法的内存占用介于轻量级方法与复杂结构方法之间。虽然引入了原型增强与特征建模模块，但通过自适应筛选机制有效控制了无效信息的存储规模，从而避免了内存开销的快速增长。因此，本文方法在保证较高性能的同时，实现了较为平衡的内存使用效率。

6　总结

本文针对 FCL 中严峻的灾难性遗忘问题以及异构环境下的模型漂移挑战，提出了端云协同优化方法 PAF-FCL。该方法通过在客户端和服务器端实施差异化且互补的知识管理策略，平衡了模型对新任务的适应性与对旧知识的保持力。具体而言，本文方法的核心贡献在于揭示并利用了 FCL 中表示质量-混淆程度-知识可信度三者之间的因果关联：客户端 OPE 与 APF 构成双层表征优化，从源头提升原型质量并抑制混淆；服务器 NF 构建概率密度防线，对聚合后的全局知识进行置信度筛选。二者并非孤立的功能模块，而是通过高质量原型输入、低噪声聚合、精准筛选反馈、迭代强化表征的闭环，在统一的信息质量控制准则下协同运作。这种端侧净化、云侧甄别的协同设计，使得模型在不断吸纳增量知识的过程中能够始终维持核心历史知识的完整性。

在实验验证方面，本文在 CIFAR-10、CIFAR-100 以及 Tiny-ImageNet 等多维度数据集上进行了详尽的对比分析。定量分析结果表明，PAF-FCL 不仅在分类准确率上优于现有主流 FCL 算法，更在长序列任务中展现出了遗忘抑制能力，其遗忘率指标在复杂异构场景下依然能维持在较低水平。

在未来研究中，我们将提升系统的通用性与长期演进能力作为重要研究方向。一方面，拟将现有的知识过滤与特征增强机制进一步拓展至跨模态持续学习任务，以适应更加复杂的多源异构数据环境；另一方面，拟结合动态神经架构搜索技术，实现模型容量的自适应扩展与结构动态调整，从而增强联邦持续学习系统对长期任务演化和复杂动态环境的适应能力，并进一步拓展其在真实场景中的应用价值。

参 考 文 献

[1] Myeongho J, Hyoje L, Yedarm S, Myungjoo K. Learning without prejudices: Continual unbiased learning via benign and malignant forgetting // Proceedings of the Eleventh International Conference on Learning Representations, ICLR 2023. Kigali, Rwanda, 2023

[2] Abhishek K, Hal Daume III. Learning task grouping and overlap in multi-task learning. arXiv preprint, arXiv:1206.6417, 2012

[3] Sebastian T. A lifelong learning perspective for mobile robot control //Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS 1994,.Munich, Germany, 1994: 23-30

[4] Zhi zhong L, Derek H. Learning without forgetting //Proceedings of the Computer Vision - ECCV2016 - 14th European Conference, Amsterdam, The Netherlands, 2016: 614-629

[5] McMahan B, Moore E, Ramage D, et al. Communication-efficient learning of deep networks from decentralized data //Proceedings of the 20th International Conference on Artificial Intelligence and Statistics. Fort Lauderdale, USA, 2017: 1273-1282

[6] Qiang Y, Yang L, Chen T J, and Tong T X. Federated machine learning: Concept and applications. ACM Transactions on Intelligent Systems and Technology (TIST), 2019, 10(2):1-19

[7] Qu L Q, Zhou Y Y, Paul P L, Xia Y D, Wang F F, Ehsan Adeli, Li Fei-Fei, and Daniel Rubin. Rethinking architecture design for tackling data heterogeneity in federated learning//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022: 10061-10071

[8] Yoon J, Jeong W, Lee G, et al. Federated continual learning with weighted inter-client transfer //Proceedings of the International Conference on Machine Learning, Virtual, 2021: 12073-12086

[9] Qi D Q, Zhao H D, Li S. Better generative replay for continual federated learning //Proceedings of the Eleventh International Conference on Learning Representations, ICLR. Kigali, Rwanda, May 1-5, 2023

[10] Wang, Z., Zhang, Y., Xu, X., et al. Federated probability memory recall for federated continual learning. Information Sciences, 2023, 629: 551-565

[11] Dong J, Wang L, Fang Z, et al. Federated class-incremental learning //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. New Orleans, USA,2022: 10164-10173

[12] Liu C H, Qu X Y, et al. Fedet: A communication-efficient federated class-incremental learning framework based on enhanced transformer //Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence, IJCAI. Macao, China, 2023: 3984-3992

[13] Wuerkaixi, A., Cui, S., et al. Accurate forgetting for heterogeneous federated continual learning. arXiv preprint arXiv:2502.14205, 2025

[14] Conor D, Artur B, et al. Neural spline flows //Proceedings of the Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems 2019, NeurIPS. Vancouver, Canada, 2019: 7509-7520

[15] Christina W, Daniel E. et al. Learning likelihoods with conditional normalizing flows. arXiv preprint arXiv:1912.00042, 2019

[16] Geirhos R, Jacobsen J H, Michaelis C, et al. Shortcut learning in deep neural networks. Nature Machine Intelligence, 2020, 2(11): 665-673

[17] Wei Y, Ye J, Huang Z, et al. Online prototype learning for online continual learning //Proceedings of the IEEE/CVF International Conference on Computer Vision, 2023: 18764-18774

[18] Wu Mei-Jun, Yang Xin, Pan Chao-Fan, et al. Autoencoders combined with continual learning: Current status, challenges and prospects. Journal of Computer Science, 2025, 48(02): 317-357 (in Chinese)
(吴美君,杨新,潘超凡,等.自编码器结合持续学习:现状、挑战与展望[J].计算机学报,2025,48(02):317-357)

[19] Jiang Wei-Jin, Cui Xin-Yu, Liu Zhi-Hua, et al. Analytical Federated Learning Method Based on Learnable Aggregation Weights. Journal of Computer Science, 2026, 49(01): 84-108 (in Chinese)
(蒋伟进,崔新雨,刘志华,等.基于可学习聚合权重的解析性联邦学习方法[J].计算机学报,2026,49(01):84-108)

[20] Gao Sheng, Yuan Li-Ping, Zhu Jian-Ming, et al. A blockchain-based privacy-preserving asynchronous federated learning method. Scientia Sinica Informationis, 2021, 51: 1755-1774 (in Chinese)
(高胜, 袁丽萍, 朱建明, 等. 一种基于区块链的隐私保护异步联邦学习. 中国科学: 信息科学, 2021, 51: 1755–1774)

[21] Xiong Z, Cheng Z, Lin X, et al. Facing small and biased data dilemma in drug discovery with enhanced federated learning approaches. Science China Life Sciences, 2022, 65(3): 529-539

[22] Peng B, Chi M M, Liu C, et al. Non-IID federated learning via random exchange of local feature maps for textile IIoT secure computing. Science China Information Sciences, 2022, 65(7): 170302

[23] Wang H Y, Mikhail Y, Sun Y K, et al. Federated learning with matched averaging //Proceedings of the 8th International Conference on Learning Representations, ICLR, Addis Ababa, Ethiopia, April, 2020: 26-30

[24] Mohri M, Sivek G, Suresh A T. Agnostic Federated Learning //Proceedings of the 36th International Conference on Machine Learning, PMLR, Long Beach, USA, 2019, 97:4615-462

[25] Zhu Z D, Hong J Y ,Zhou J Y . Data-free knowledge distillation for heterogeneous federated learning //Proceedings of the 38th International Conference on Machine Learning, ICML 2021, Virtual, 2021, 139: 12878-12889

[26] Li Z, Hoiem D. Learning without forgetting. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 40(12): 2935-2947

[27] Kirkpatrick, J., Pascanu, R., Rabinowitz, et al. Overcoming catastrophic forgetting in neural networks. Proceedings of the National Academy of Sciences, 2017, 114(13): 3521-3526

[28] Rusu, A.A., Rabinowitz, N.C, et al. Progressive neural networks. arXiv preprint arXiv:1606.04671, 2016

[29] Aljundi R., Belilovsky E, Tuytelaars T, et al. Online continual learning with maximal interfered retrieval. Advances in Neural Information Processing Systems, 2019, 32

[30] Shim D, Mai Z, Jeong J, et al. Online class incremental continual learning with adversarial Shapley value //Proceedings of the AAAI Conference on Artificial Intelligence, Vancouver, Canada, 2021, 35(11): 9630-9638

[31] Wu C, Herranz L, Liu X, et al. Memory replay gans: Learning to generate new categories without forgetting //Proceedings of the Neural Information Processing Systems 31: Annual Conference on Neural Information Processing Systems 2018, NeurIPS, Montréal, Canada, 2018: 5966-5976

[32] Rebuffi S-A, Kolesnikov A, Sperl G, Lampert C H. iCaRL: Incremental classifier and representation learning //Proceedings of the IEEE conference on Computer Vision and Pattern Recognition. Honolulu, USA, 2017: 2001-2010

[33] Liu Hai-Jun, Wang Hao-Long, Liu Ya-Hui, et al. A personalized federated learning method based on prototype clustering mechanism. Journal of Communications, 2025, 46(S1): 92–101 (in Chinese)
(刘海军,王浩龙,刘雅辉,等.基于原型聚类机制的个性化联邦学习方法[J].通信学报, 2025, 46(S1): 92-101)

[34] Zhang Hao-Dong, Yang Liu, Yu Jian, et al. A federated continual learning method based on prototype learning. Scientia Sinica Informationis, 2024, 54(10): 2428-2442 (in Chinese)
(张浩东,杨柳,于剑,等.基于原型学习的联邦持续学习方法[J].中国科学:信息科学,2024,54(10): 2428-2442)

[35] Li Y, Xu W, Wang H, et al. Personalized federated domain-incremental learning based on adaptive knowledge matching. In European Conference on Computer Vision, 2025: 127-144. Springer

[36] Tran MT, Le T, Le XM, et al. Text-enhanced data-free approach for federated class-incremental learning //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA, 2024: 23870-23880

[37] Zhang H, Cisse M, Dauphin YN, Lopez-Paz D. mixup: Beyond empirical risk minimization. arXiv preprint arXiv:1710.09412, 2017

[38] Khosla P, Teterwak P, Wang C, et al. Supervised contrastive learning. Advances in Neural Information Processing Systems, 2020, 33: 18661-18673

[39] Li T, Sahu A K, Zaheer M, et al. Federated optimization in heterogeneous networks //Proceedings of Machine Learning and Systems, Austin, USA, 2020, 2: 429-450.

[40] Cha S., Chen H, Vikalo H. Task-Agnostic Federated Continual Learning via Replay-Free Gradient Projection. arXiv preprint arXiv:2509.21606, 2025.

[41] Yu H, Yang X, Zhang L, et al. Handling spatial-temporal data heterogeneity for federated continual learning via tail anchor //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2025: 4874-4883.

[42] Qi Z, Tang Y.P, Meng L, et al. From Selection to Scheduling: Federated Geometry-Aware Correction Makes Exemplar Replay Work Better under Continual Dynamic Heterogeneity. arXiv preprint arXiv:2604.08617, 2026.

JIANG Wei-Jin, Ph.D., Professor. His main research interests include edge computing, social computing, and cyberspace security.

CUI Xin-Yu, M.S. candidate, Her main research interests include federated learning and privacy protection.

Liu Zhi-Hua, M.S. candidate, His main research interests include federated learning and privacy protection.

CHEN Shen-You, M.S. candidate, His main research interests include federated learning and privacy protection.

HU Jia-Long, M.S. candidate, His main research interests include dynamic pricing of data and federated learning.

CHEN Xiao-Hong, Ph.D., Professor, Academician of Chinese Academy of Engineering. Her main research interests include edge computing, social computing, and cyberspace security.

Background

This study focuses on the challenges of federated continual learning (FCL) in heterogeneous federated environments. In many real-world applications such as healthcare, finance, and edge intelligence, data are generated continuously across distributed clients. These data streams naturally form sequential tasks, requiring models to continuously learn new knowledge while retaining previously acquired information. However, the decentralized nature of federated learning and the non-independent and identically distributed (Non-IID) characteristics of client data make continual learning significantly more challenging.

A major issue in FCL is the catastrophic forgetting problem, where learning new tasks leads to the degradation of performance on previously learned tasks. This issue becomes more severe in federated settings due to limited local storage, privacy constraints, and heterogeneous data distributions among clients. Most existing

approaches attempt to mitigate forgetting by storing historical data or maintaining additional model parameters. However, these strategies either violate privacy requirements or impose significant storage and computation overhead on resource-constrained clients.

Another emerging requirement in practical applications is accurate forgetting, which aims to selectively remove specific knowledge from a trained model while preserving the utility of the remaining information. Accurate forgetting is particularly important in privacy-sensitive scenarios where certain data must be removed due to regulatory or user requirements. Nevertheless, existing forgetting approaches often rely on retraining or approximate unlearning strategies, which can introduce substantial computational costs and degrade model performance.

To address these challenges, this paper proposes a prototype-guided accurate forgetting method for heterogeneous federated environments. The proposed approach leverages class prototypes to capture essential semantic representations of learned knowledge while avoiding direct access to raw data. By introducing a prototype purification mechanism, the method effectively separates retained knowledge from information that needs to be forgotten, enabling precise knowledge removal while maintaining model stability.

The proposed framework integrates prototype guidance with federated continual learning, enabling efficient knowledge preservation and targeted forgetting without violating data privacy constraints. This design not only mitigates catastrophic forgetting in sequential federated tasks but also improves the controllability of knowledge removal in distributed learning systems.

The significance of this research lies in providing a new paradigm for privacy-preserving continual learning and controllable knowledge forgetting in federated environments.