FedDACS: 基于分布感知可控采样优化的个性化群体知识迁移

齐霄超，倪郑威
（浙江工商大学信息与电子工程学院，浙江 杭州 310018）

摘　要：个性化联邦学习作为分布式机器学习领域的前沿范式，通过为各客户端训练专属模型来应对数据异构性挑战。现有研究采用群体知识迁移技术，在服务器端部署共享预测器提升模型在基类上的泛化能力，在本地更新特征提取器提升模型个性化能力。但仍存在两方面显著缺陷：其一，全局共享的预测器难以适配高度异构的客户端分布，导致准确率下降；其二，知识迁移过程中易受分布差异干扰，出现负迁移效应。为此，提出了分布感知可控采样优化的个性化群体知识迁移框架（FedDACS）。该框架将服务器端传统单一共享预测器改为个性化预测器阵列，每个客户端对应专属预测器模块，与本地特征提取器构成个性化模型。此外，引入了分布感知可控采样（DACS）技术，通过实时分析各客户端数据分布，动态调整特征空间采样策略，控制个性化预测器的输入。该框架可有效抑制分布差异导致的负迁移效应。为了验证FedDACS的有效性，在CIFAR-10与CIFAR-100数据集上构建了7种非独立同分布（Non-IID）场景进行系统性验证。实验结果表明，相较于Fedgkt、FedProto等8种基准方法，FedDACS在CIFAR-10数据集上实现平均用户准确率提升幅度达6.85%，在CIFAR-100数据集上提升幅度达5.28%。验证了所提FedDACS方法能够有效提升个性化模型在不同数据异质性场景下的性能。

关键词：个性化联邦学习；可控采样；群体知识迁移；联邦知识蒸馏
中图分类号：TP181
文献标志码：A
doi: 10.11959/j.issn.1000−0801.DXKX250206

FedDACS: personalized knowledge transfer based on distribution-aware controlled sampling optimization

Qi Xiaochao, Ni Zhengwei
School of Information and Electronic Engineering, Zhejiang Gongshang University, Hangzhou 310018, China

Abstract: Personalized federated learning, as a cutting-edge paradigm in the field of distributed machine learning, is developed to address the challenges of data heterogeneity by training exclusive models for each client. Existing research employes population knowledge transfer techniques, where a shared predictor is deployed on the server to en-

收稿日期：2025−03−29；**修回日期：**2025−06−10
通信作者：倪郑威，zhengwei.ni@zjgsu.edu.cn
基金项目：桐乡市通用人工智能研究院项目（No.TAGI2-B-2024-0017）
Foundation Item: Tongxiang General Artificial Intelligence Research Institute Project (No.TAGI2-B-2024-0017)

hance the generalization ability of the model over the base class, while local feature extractors are updated to improve the personalization ability of the model. However, two significant drawbacks remain. Firstly, the globally shared predictor struggles to adapt to highly heterogeneous client distributions, resulting in decreased accuracy. Secondly, the knowledge transfer process is easily disturbed by distribution differences, leading to negative transfer effects. To address these issues, a distribution-aware controlled sampling optimization framework for personalized population knowledge transfer (FedDACS) was proposed. The traditional single shared predictor on the server was transformed into an array of personalized predictors, with each client corresponding to an exclusive predictor module, forming a personalized model alongside local feature extractors. Furthermore, a distribution-aware controlled sampling technique (DACS) was introduced, which dynamically adjusted the feature space sampling strategy by analyzing the data distributions of each client in real-time, thereby controlling the input of personalized predictors. The negative transfer effects caused by distribution differences were effectively mitigated. To validate the effectiveness of FedDACS, seven Non-IID scenarios were constructed on CIFAR-10 and CIFAR-100 datasets for systematic validation. Experimental results demonstrate that, compared to eight baseline methods such as Fedgkt and FedProto, FedDACS achieves an average user accuracy improvement of 6.85% on CIFAR-10 and an enhancement of 5.28% on CIFAR-100. These results indicate that the proposed FedDACS method can effectively improve the performance of personalized models in various data heterogeneity scenarios.

Key words: personalized federated learning, controlled sampling, group knowledge transfer, federated knowledge distillation

0　引言

在联邦学习（FL）网络中，个性化联邦学习（PFL）[1-2]已经成为一个关键的研究方向，它通过为每个客户端提供定制化模型来应对数据异质性，同时使得各客户端仍然能够从协作训练中获益。研究者基于不同的方向提出了多种个性化解决方案[1]，其技术路径可归纳为4类：（1）数据增强[3-4]，通过客户端数据重采样或生成对抗网络减少统计异质性；（2）模型微调[2,5]，基于全局模型进行本地自适应优化；（3）架构定制[6-7]，为不同客户端设计差异化模型结构；（4）相似性建模[8-9]，利用客户端关系网络传递知识。其中，基于架构的方式是目前主流的个性化方法。联邦知识蒸馏是其中的一种方法，因兼顾模型压缩与个性化适配的优势备受关注。

联邦知识蒸馏的核心思路是通过软标签或特征在全局–本地间知识迁移。现有的大部分研究工作[10-12]倾向于直接利用群体知识，聚合软标签以生成全局知识并进行迁移，如FedMD[6]、FedDF[10]、FedHKD[11]等算法；或服务器生成全局模型，客户端蒸馏全局模型的知识到本地，指导模型更新，如Fedgkt[12]。直接使用群体知识迁移的蒸馏方式计算高效，泛化性强，但也会因此带来一些个性化不足和负迁移影响。以Fedgkt为例分析发现：其一，全局模型无法精准捕捉到单个用户数据分布的特性，导致个性化不足；其二，知识迁移易受分布差异干扰，造成平均准确率下降。尽管后续的工作，如FedDKC [13]和FedICT[14]尝试解决知识差异问题，但它们往往依赖于正则化方法，而不是基于分布相似性明确选择相关知识源。其他一些方法调整知识分布[15-16]，可能依赖于公共数据集，或者过于关注全局对齐，而忽视了个性化需求。

为了解决在异质性联邦知识蒸馏（FKD）中，使用单一共享预测器导致个性化不足和负迁移的问题，提出了FedDACS：一种基于分布感知可控采样优化的个性化知识迁移框架。FedDACS引入了两个关键创新。

（1）个性化预测器阵列。与单一共享预测器

不同，FedDACS在服务器上维护一个个性化预测器阵列，每个客户端都有一个专门的预测器（对于客户端m是g_V^m）。这种结构使得每个预测器能够专注于与其对应客户端相关的知识，从而避免特征空间中来自无关数据分布的干扰。

（2）分布感知可控采样（DACS）。为了在不访问原始客户端数据或公共数据集的情况下有效地训练这些个性化预测器，FedDACS采用了DACS技术。DACS分析每个客户端上传的标签分布（ζ_m），并将其与服务器特征空间中的整体标签分布（ζ_g）进行比较。然后，基于这两种分布的加权组合，从特征空间中进行控制采样，为每个个性化预测器g_V^m创建量身定制的训练子集。通过动态调整源领域分布来生成知识，从而更好地匹配目标客户端的特征，增强知识传递的相关性并减轻负迁移。

FedDACS采用交替最小化策略，利用本地数据和来自服务器的蒸馏知识优化客户端的本地特征提取器（f_U^m），并通过DACS采样特征优化服务器端的个性化预测器（g_V^m）。本文详细介绍了FedDACS框架，并提供了广泛的实验验证，证明了其在各种非独立同分布（Non-IID）场景下，相较于先进的个性化联邦学习方法，能够有效提高个性化模型的准确性。

1 相关工作

1.1 联邦知识蒸馏

最近有研究提出了一些基于相似本地数据分布的客户端知识聚类、加权或过滤的方法[17-19]。Xie等[20]提出的PerAda算法，通过全局适配器规范每个客户端的个性化适配器以实现泛化，全局适配器利用知识蒸馏聚合来自所有客户端的泛化信息。然而，这些方法大多依赖公共数据集，但实际情况中，这类数据集常常难以获取。

在没有公共数据集的情况下，少数方法实现了联邦蒸馏。Wu等[21]设计了FedCache架构，利用知识缓存来捕获与每个私有样本相关的最新知识，并使用知识缓存驱动的个性化蒸馏进行模型训练。Tan等[22]提出的FedProto通过对齐全局原型与局部原型来解决数据异构问题。这些方法通常交换的信息较少，实际效用有限。最近一些研究考虑共性知识与个性化知识的平衡，Chen等[23]提出的MetaFed框架采用环形知识蒸馏，不需要中央服务器，通过自适应蒸馏进行知识传递。但该方法未考虑极端数据异构下顺序训练带来的灾难性遗忘问题。

Fedgkt是一种群体知识迁移训练算法，解决了边缘设备的资源限制问题，结合全局知识与个性化知识来优化模型。Wu等[13]提出的FedDKC算法，利用分布式知识一致性来解决本地模型的知识差异问题，并提出了基于核的知识求精（KKR）策略和基于搜索的知识求精（SKR）策略。此外，FedICT[14]结合了联邦知识蒸馏与多任务学习，支持多任务客户端。这些方法通过在蒸馏损失函数中加入正则化项来约束模型参数，但缺乏显式的知识选择机制，导致知识迁移效率和效果受到限制。

本文提出一种个性化的群体知识迁移策略，对知识来源进行采样，对来自不同客户端提供的知识进行选择性融合，提高知识迁移效率和个性化模型准确率，以更好地应对数据异质性和模型个性化。

1.2 调整知识分布

源域和目标域的相似性是迁移学习成功的关键[24]。一些方法在蒸馏过程中调整知识分布[15-16]。Itahara等[15]提出的半监督联邦学习框架，利用全局教师模型生成的软标签，以帮助本地模型克服Non-IID数据带来的偏差。但该方法依旧强调全局分布的重要性，忽略了个性化需求。Chang等[16]通过对抗训练强制服务器与客户端的模型输出分布尽可能接近，实现知识迁移。

但该方法依赖不切实际的公共数据集。

尽管群体知识迁移已得到大量研究，但仍面临以下挑战。

（1）在面对边缘客户端之间的高度数据异构分布时，如何设计个性化群体知识迁移框架，尽量减少负迁移效应，并带来个性化模型准确率的提升。

（2）在不依赖公共数据集的前提下，如何设计对知识来源的显式选择机制，调整知识分布，提高源域和目标域的相似性，平衡共识和个性化知识对模型的影响。

本文提出分布感知可控采样的个性化全体知识转移框架FedDACS，通过将全局预测器拆分为个性化预测器阵列，减少不相干知识来源干扰。同时感知本地数据分布，在特征空间结合全局分布与边缘数据分布进行可控采样，对知识来源进行选择。所提方法能够显著提升个性化模型的准确率，有效应对上述挑战。

2　问题描述

为了方便阐述，首先给出问题描述。假设客户端是M个资源受限设备，其中第m个客户端拥有本地数据集$\boldsymbol{\mathcal{D}}^m=\left\{\left(\boldsymbol{x}_i^m,\boldsymbol{y}_i^m\right)\right\}_{i=1}^{\left|\boldsymbol{\mathcal{D}}^m\right|}$。本文目标是在中央服务器的帮助下，运用全局数据集$\boldsymbol{\mathcal{D}}\triangleq\bigcup_{m\in[M]}\boldsymbol{\mathcal{D}}^m$进行协作学习，在无须交换原始数据条件下，为每个客户端m训练个性化模型$\phi_{\boldsymbol{W}}^m$。客户端本地模型$\varepsilon_{\boldsymbol{E}}^m$的特征提取器$f_{\boldsymbol{U}}^m$参与构成个性化模型的前半部分，服务器端的预测器$g_{\boldsymbol{V}}^m$作为构成个性化模型的后半部分。在优化过程中，服务器收集客户端上传的特征，利用服务器端特征空间的信息维护个性化预测器，预测器生成个性化知识返回客户端本地，帮助本地特征提取器优化。每个客户端的个性化目标函数为：

$$\phi_{\boldsymbol{W}}^m=\operatorname{argmin}l_{\text{local}}\left(\varepsilon_{\boldsymbol{E}}^m;\boldsymbol{\mathcal{D}}^m,\hat{\boldsymbol{y}}_{\boldsymbol{g}}^{\boldsymbol{m}}\right)+l_{\text{server}}\left(g_{\boldsymbol{V}}^m;\boldsymbol{\Phi}_{\text{sub}}^{\boldsymbol{m}}\right)\quad(1)$$

其中，l_{local}和l_{server}分别是客户端和服务器的损失函数，$\boldsymbol{\Phi}_{\text{sub}}^{\boldsymbol{m}}$是特征空间采样的信息，帮助预测器更新。考虑客户端的资源受限及隐私保护问题，采用交替最小化策略来解决这个双侧优化问题。

3　基于分布感知可控采样优化的个性化群体知识迁移方法设计

本节主要介绍FedDACS个性化群体知识迁移。关于符号表示，本文遵循标准的数学符号规范，其中$|\cdot|$表示1-范数。本文使用的数学符号及含义见表1。

3.1　FedDACS框架概述

本文研究大体上遵循Fedgkt中提出的无代理数据集的联邦学习框架，通过交替最小化方法优

表1　数学符号及含义

符号	定义
M，C	客户端数量，标签类别数量
$\boldsymbol{\mathcal{D}}^m$，$\boldsymbol{x}_i^m$，$\boldsymbol{y}_i^m$	客户端m的本地数据集，$\boldsymbol{\mathcal{D}}^m$中的第$i$个样本，第$i$个样本独热编码的真实标签
$\phi_{\boldsymbol{W}}^m$，$f_{\boldsymbol{U}}^m$，$g_{\boldsymbol{V}}^m$	客户端m参数为$\boldsymbol{W}$的个性化模型，客户端m参数为$\boldsymbol{U}$的特征提取器，客户端m参数为$\boldsymbol{V}$的个性化预测器
$\varepsilon_{\boldsymbol{E}}^m$，$\gamma_{\boldsymbol{O}}^m$	客户端m参数为$\boldsymbol{E}$的本地模型，客户端m参数为$\boldsymbol{O}$的本地简易分类器
$\boldsymbol{\Phi}$，$\boldsymbol{\Phi}^{\boldsymbol{m}}$，$\boldsymbol{\Phi}_{\text{sub}}^{\boldsymbol{m}}$	服务器端特征空间信息，客户端m的特征空间信息，客户端m在特征空间采样的信息
$\boldsymbol{h}^{\boldsymbol{m}}$	客户端m的特征信息
$\hat{\boldsymbol{y}}_{\boldsymbol{l}}^{\boldsymbol{m}}$，$\hat{\boldsymbol{y}}_{\boldsymbol{g}}^{\boldsymbol{m}}$	客户端m本地模型生成的预测标签，客户端m个性化预测器生成的预测标签
$\zeta_{\boldsymbol{m}}$，$\zeta_{\boldsymbol{g}}$	客户端m的本地数据分布，服务器端特征空间数据分布
T，t	通信总轮数；当前训练轮数
η，α	特征空间标签分布占采样分布比例，权衡知识相似性损失影响的超参数

化边缘设备与服务器端的模型参数，从而得到各客户端的个性化模型。传统无代理数据集的联邦学习框架在服务器端只设置一个共享预测器模型，即不同客户端的g_V^m是同一个模型。但FedDKC研究显示，共享预测器的知识有些是有害的，会对客户端本地模型产生负迁移影响。因此，在无代理数据集的联邦学习框架上进一步优化，提出个性化群体知识迁移框架FedDACS。FedDACS整体框架如图1所示，分为边缘和服务器两大部分，每个客户端都在服务器端有对应的个性化预测器模型g_V^m，即不同客户端的g_V^m不同，分布感知可控采样为其提供个性化的样本来源，以便为每个客户端从群体知识中获取属于自己的个性化部分进行知识迁移。

如图1（a）所示，客户端拥有本地数据集$\mathcal{D}^m$，$\boldsymbol{Y}$表示所有客户端的真实标签集合。为了便于本地特征提取器f_U^m优化，在边缘侧为其配备了一个简易的分类器γ_O^m。因此，客户端m的本地模型ε_E^m可以表示为$\varepsilon_E^m\left(f_U^m,\gamma_O^m\right)$。

对于客户端的本地模型ε_E^m，其中特征提取器f_U^m负责将输入样本$\boldsymbol{x}_i^m$映射至一个特征向量$\boldsymbol{h}_i^m$，即$\boldsymbol{h}_i^m=f_U^m\left(\boldsymbol{x}_i^m\right)$，$\boldsymbol{H}$表示所有客户端的特征向量集合。分类器$\gamma_O^m$对特征向量$\boldsymbol{h}_i^m$进行分析得到一个软标签，即预测向量$\hat{\boldsymbol{y}}_l^m$。$\hat{\boldsymbol{Y}}_l$表示所有客户端的预测向量集合（$\boldsymbol{Y}$、$\boldsymbol{H}$、$\hat{\boldsymbol{Y}}_g$、$\boldsymbol{Y}_{\text{sub}}$、$\hat{\boldsymbol{Y}}_{\text{sub}}$与之类似，是为形象描述框架结构抽象出的一个概括性表示，在实验中，并不将其作为一个集合进行运算或传输）。在本地更新时，计算本地预测向量$\hat{\boldsymbol{y}}_l^m$与真实标签$\boldsymbol{y}^m$之间的损失l_{CE}，本地预测向量$\hat{\boldsymbol{y}}_l^m$与服务器端传回的预测标签$\hat{\boldsymbol{y}}_g^m$之间的损失l_{KL}，将二者结合得到损失l_{local}指导本地模型更新。客户端在完成更新后，将特征向量$\boldsymbol{h}^m$、标签$\boldsymbol{y}^m$上传至服务器端的特征空间。

如图1（b）所示，服务器端由特征空间信息$\boldsymbol{\Phi}$、分布感知采样模块**DACS**和个性化预测器阵列$\{g_V^1,g_V^2,\cdots,g_V^m\}$组成。其中，个性化预测器$g_V^m$根据**DACS**采样出的相关样本$\boldsymbol{\Phi}_{\text{sub}}^m$进行模型更新，使得知识迁移的源域与目标域尽可能相似。预测器更新完成后，将其设为评估模式，输入各客户

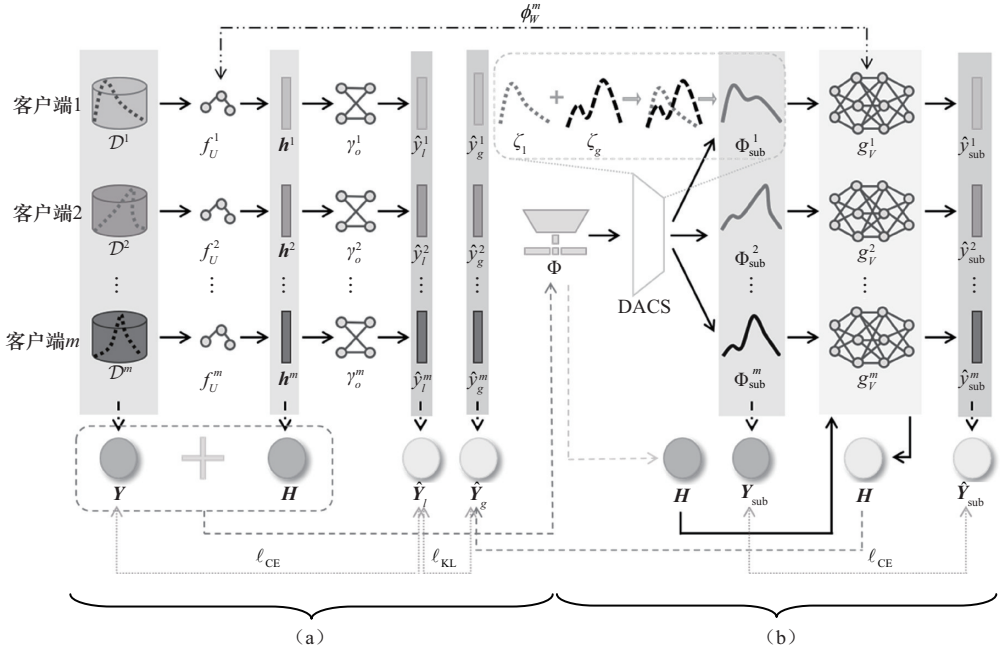

图1　FedDACS整体框架

端上传的信息，得到服务器端预测向量$\hat{\boldsymbol{y}}_{\boldsymbol{g}}^{\boldsymbol{m}}$。

3.2　分布感知可控采样 DACS

在联邦蒸馏过程中，数据分布的显著差异是一个亟须解决的重要问题，这一问题主要源自客户端数据的异质性。传统群体知识迁移框架 Fedgkt 采用双向蒸馏，即共享预测器从客户端的本地模型中吸收知识，客户端也从共享预测器中获取知识。但近期研究表明，在双向蒸馏框架中，局部数据的异质性对全局模型的性能会产生负面影响，而全局模型吸收客户端的知识会进一步加重这种影响[25-26]。这导致在互相学习知识的过程中，不一致的更新在所难免，并可能引发不稳定的收敛现象，进而形成客户端漂移[25-28]。

为了应对这一挑战，提出了分布感知可控采样策略 DACS，将互相学习知识的服务器端预测器与客户端模型之间的数据分布对齐，减少进行知识迁移的源域与目标域的差异，从而降低数据分布差异太大带来的知识负迁移影响。同时，改用单向知识蒸馏，避免表达能力较弱的本地知识对服务器端预测器的影响。在图 1（b）中，该策略如虚线框中所示。具体而言，包含两个主要步骤。

（1）分布感知

在收集到各客户端上传的信息$\boldsymbol{\Phi}$后，首先计算其本地标签分布$\boldsymbol{\zeta}_{\boldsymbol{m}}$。以客户端$m$为例，边缘侧上传的信息为$\boldsymbol{\Phi}^{\boldsymbol{m}}=\{\boldsymbol{h}^{\boldsymbol{m}},\boldsymbol{y}^{\boldsymbol{m}}\}$，通过计算其标签概率$p_c^m$，从而得到其本地标签分布向量$\boldsymbol{\zeta}_{\boldsymbol{m}}=(p_1^m,p_2^m,\cdots,p_C^m)$。其中$p_c^m$表示客户端$m$的样本中属于标签$c$的样本概率，计算式为：

$$p_c^m=\frac{\left|\left\{\left(\boldsymbol{h}_i^{\boldsymbol{m}},\boldsymbol{y}_i^{\boldsymbol{m}}\right)\middle|\left(\boldsymbol{h}_i^{\boldsymbol{m}},\boldsymbol{y}_i^{\boldsymbol{m}}\right)\in\boldsymbol{\Phi}^{\boldsymbol{m}},\boldsymbol{y}_i^{\boldsymbol{m}}=c\right\}\right|}{\left|\boldsymbol{\Phi}^{\boldsymbol{m}}\right|}\tag{2}$$

其中，$\left|\boldsymbol{\Phi}^{\boldsymbol{m}}\right|$表示上传样本信息数量。

个性化联邦学习是在追求全局模型泛化性能的基础上，为每个参与联邦学习的用户提供定制化模型的一种方法。本文提出的框架虽然不存在全局模型，但也应该考虑个性化模型在基类（在其他客户端上可见但在本地训练期间不可见的类）上的泛化能力。因此，将全局数据分布考虑进来，结合特征空间的全部客户端信息，计算全局标签分布$\boldsymbol{\zeta}_{\boldsymbol{g}}=(p_1^g,p_2^g,\cdots,p_C^g)$，其中$p_c^g$表示$\boldsymbol{\Phi}$中属于标签$c$的样本概率，计算过程与本地标签分布$\boldsymbol{\zeta}_{\boldsymbol{m}}$的类似。

（2）可控采样

为了控制个性化模型在基类上的泛化能力和个性化能力的平衡，在采样时，依照一定比例，将客户端标签分布$\boldsymbol{\zeta}_{\boldsymbol{m}}$和特征空间标签分布$\boldsymbol{\zeta}_{\boldsymbol{g}}$结合起来，对特征空间的信息进行可控采样，使得个性化模型在基类上也有一定的泛化能力。具体而言，样本集$\boldsymbol{\Phi}_{\mathbf{sub}}^{\boldsymbol{m}}$的形成可以表示为：

$$\boldsymbol{\Phi}_{\mathbf{sub}}^{\boldsymbol{m}}=\bigcup_{c=1}^{C}\mathcal{R}(\boldsymbol{\Phi},\eta\cdot p_c^g+\left(1-\eta\right)\cdot p_c^m)\tag{3}$$

其中，$\mathcal{R}(\boldsymbol{\Phi},p^m)$表示在$\boldsymbol{\Phi}$中以概率$p^m$对属于某个标签类别$c$的样本进行的随机采样。$\eta$是控制特征空间标签分布占采样分布比例的参数，取值介于 0 到 1，用于平衡标签c的全局样本概率p_c^g和本地样本概率p_c^m对采样概率p^m的影响，即平衡特征空间标签分布$\boldsymbol{\zeta}_{\boldsymbol{g}}$和本地标签分布$\boldsymbol{\zeta}_{\boldsymbol{m}}$对采集样本集$\boldsymbol{\Phi}_{\mathbf{sub}}^{\boldsymbol{m}}$的标签分布的影响，进而影响模型的基类泛化能力与个性化能力的平衡。

3.3　FedDACS 训练流程

设T为通信总轮数，t为当前训练轮数，$\boldsymbol{W}_t^m=\left\{\boldsymbol{U}_t^m,\boldsymbol{V}_t^m\right\}$为第$t$轮客户端$m$的个性化模型参数。第$t$轮客户端$m$的本地模型参数记为$\boldsymbol{E}_t^m=\left\{\boldsymbol{U}_t^m,\boldsymbol{O}_t^m\right\}$，个性化预测器参数记为$\boldsymbol{V}_t^m$。在训练初始阶段，即$t=0$时，所有客户端和服务器端个性化预测器阵列都随机初始化各自的模型参数$\boldsymbol{E}_0^m$和$\boldsymbol{V}_0^m$。第t轮训练过程如下，可以分为两部分。

（1）边缘侧优化

每个客户端m在其私有数据集$\boldsymbol{\mathcal{D}}^m$上，结合

接收到的个性化预测器知识$\hat{\boldsymbol{y}}_{\boldsymbol{g}}^{\boldsymbol{m}}=\cup_{i=1}^{\left|\mathcal{D}^m\right|}\hat{\boldsymbol{y}}_{\boldsymbol{g}}^{\boldsymbol{m}i}$，按下述式（4）更新其本地模型参数$\boldsymbol{E}_t^m=\left\{\boldsymbol{U}_t^m,\boldsymbol{O}_t^m\right\}$，从而更新个性化模型的特征提取器参数$\boldsymbol{U}_t^m$：

$$\left\{\boldsymbol{U}_{t+1}^m,\boldsymbol{O}_{t+1}^m\right\}=\boldsymbol{E}_{t+1}^m\leftarrow\boldsymbol{E}_t^m-\mathrm{lr}\nabla l_{\mathrm{local}}\left(\boldsymbol{E}_t^m;\mathcal{D}^m,\hat{\boldsymbol{y}}_{\boldsymbol{g}}^{\boldsymbol{m}}\right)\tag{4}$$

其中，lr是学习率，l_{local}是本地模型的训练损失，由交叉熵损失l_{CE}和知识相似性损失l_{sim}组合而成。l_{sim}通常采用KL散度损失。具体表达如下：

$$\begin{aligned}&l_{\mathrm{local}}\left(\boldsymbol{E}_t^m;\mathcal{D}^m,\hat{\boldsymbol{y}}_{\boldsymbol{g}}^{\boldsymbol{m}}\right)=l_{\mathrm{CE}}\left(\hat{\boldsymbol{y}}_{\boldsymbol{l}}^{\boldsymbol{m}},\boldsymbol{y}^{\boldsymbol{m}}\right)+\\&\alpha\cdot l_{\mathrm{sim}}\left(\hat{\boldsymbol{y}}_{\boldsymbol{l}}^{\boldsymbol{m}},\hat{\boldsymbol{y}}_{\boldsymbol{g}}^{\boldsymbol{m}}\right)=\\&\frac{\sum_{\left(\boldsymbol{x}_{\boldsymbol{i}}^{\boldsymbol{m}},\boldsymbol{y}_{\boldsymbol{i}}^{\boldsymbol{m}}\right)\in\mathcal{D}^m}\left\{l_{\mathrm{CE}}\left(\hat{\boldsymbol{y}}_{\boldsymbol{l}}^{\boldsymbol{m}i},\boldsymbol{y}_{\boldsymbol{i}}^{\boldsymbol{m}}\right)+\alpha\cdot l_{\mathrm{sim}}\left(\hat{\boldsymbol{y}}_{\boldsymbol{l}}^{\boldsymbol{m}i},\hat{\boldsymbol{y}}_{\boldsymbol{g}}^{\boldsymbol{m}i}\right)\right\}}{\left|\mathcal{D}^m\right|}\end{aligned}\tag{5}$$

其中，$\hat{\boldsymbol{y}}_{\boldsymbol{l}}^{\boldsymbol{m}i}$对应输入样本$\boldsymbol{x}_{\boldsymbol{i}}^{\boldsymbol{m}}$的预测向量（本地知识），即$\hat{\boldsymbol{y}}_{\boldsymbol{l}}^{\boldsymbol{m}i}=\varepsilon_{\boldsymbol{E}}^m\left(\boldsymbol{x}_{\boldsymbol{i}}^{\boldsymbol{m}}\right)=\gamma_{\boldsymbol{O}}^m\left(\boldsymbol{h}_{\boldsymbol{i}}^{\boldsymbol{m}}\right)$。$\alpha$是权衡知识相似性损失影响的超参数，取值范围是$(0,+\infty)$。本地训练完成之后，客户端将其特征提取器的输出$\boldsymbol{h}^{\boldsymbol{m}}$和本地标签$\boldsymbol{y}^{\boldsymbol{m}}$上传至服务器。此时，客户端$m$的个性化模型参数由这两部分组成$\left\{\boldsymbol{U}_{t+1}^m,\boldsymbol{V}_t^m\right\}$。

边缘优化流程可由算法1描述，具体如下。

算法1 边缘侧FedDACS

输入 客户端m的本地数据集$\mathcal{D}^{\boldsymbol{m}}$，通信总轮数$T$，学习率lr，超参数$\alpha$

for训练轮数$t=1,2,3,\cdots,T$

//服务器回传个性化预测器知识$\hat{\boldsymbol{y}}_{\boldsymbol{g}}^{\boldsymbol{m}}$

$\hat{\boldsymbol{y}}_{\boldsymbol{g}}^{\boldsymbol{m}}\leftarrow g_V^m\left(\boldsymbol{V}_t^m;\boldsymbol{h}^{\boldsymbol{m}}\right)$

//特征提取器参数$\boldsymbol{U}_t^m$更新

$\boldsymbol{U}_{t+1}^m\leftarrow\boldsymbol{E}_{t+1}^m\leftarrow\boldsymbol{E}_t^m-\mathrm{lr}\nabla l_{\mathrm{local}}\left(\boldsymbol{E}_t^m;\mathcal{D}^m,\hat{\boldsymbol{y}}_{\boldsymbol{g}}^{\boldsymbol{m}}\right)$

//提取特征

$\boldsymbol{h}^{\boldsymbol{m}}\leftarrow$空字典

$\boldsymbol{h}^{\boldsymbol{m}}[i]\leftarrow\boldsymbol{h}_{\boldsymbol{i}}^{\boldsymbol{m}}=f_U^m\left(\boldsymbol{x}_{\boldsymbol{i}}^{\boldsymbol{m}}\right)$

上传$\boldsymbol{\Phi}^{\boldsymbol{m}}=\{\boldsymbol{h}^{\boldsymbol{m}},\boldsymbol{y}^{\boldsymbol{m}}\}$到服务器

（2）服务器端优化

在服务器端接收到所有客户端的信息$\boldsymbol{\Phi}=\cup_{m=1}^M\boldsymbol{\Phi}^{\boldsymbol{m}}$后，进行分布感知的可控采样，为服务器端个性化预测器的更新提供采样信息$\boldsymbol{\Phi}_{\mathbf{sub}}^{\boldsymbol{m}}=\{\boldsymbol{h}_{\mathbf{sub}}^{\boldsymbol{m}},\boldsymbol{y}_{\mathbf{sub}}^{\boldsymbol{m}}\}$。服务器端个性化预测器参数$\boldsymbol{V}_t^m$在第$t$轮的更新过程如下：

$$\boldsymbol{V}_{t+1}^m\leftarrow\boldsymbol{V}_t^m-\mathrm{lr}\nabla l_{\mathrm{server}}\left(\boldsymbol{V}_t^m;\boldsymbol{\Phi}_{\mathbf{sub}}^{\boldsymbol{m}}\right)\tag{6}$$

其中，l_{server}采用简单的交叉熵损失，即：

$$\begin{aligned}&l_{\mathrm{server}}\left(\boldsymbol{V}_t^m;\ \boldsymbol{\Phi}_{\mathbf{sub}}^{\boldsymbol{m}}\right)=\\&\frac{1}{\left|\boldsymbol{\Phi}_{\mathbf{sub}}^{\boldsymbol{m}}\right|}\sum_{\left(\boldsymbol{h}_{\mathbf{sub}^i}^{\boldsymbol{m}},\boldsymbol{y}_{\mathbf{sub}^i}^{\boldsymbol{m}}\right)\in\boldsymbol{\Phi}_{\mathbf{sub}}^{\boldsymbol{m}}}\left\{l_{\mathrm{CE}}\left(\boldsymbol{y}_{\mathbf{sub}^i}^{\boldsymbol{m}},\hat{\boldsymbol{y}}_{\mathbf{sub}^i}^{\boldsymbol{m}}\right)\right\}\end{aligned}\tag{7}$$

在个性化预测器更新完成后，将模型设为评估模式，输入$\boldsymbol{h}^{\boldsymbol{m}}$，得到服务器端预测向量（服务器知识）$\hat{\boldsymbol{y}}_{\boldsymbol{g}}^{\boldsymbol{m}}$，将其传回给边缘侧，指导边缘优化。至此，客户端$m$的个性化模型参数完成第$t$轮更新，即$\boldsymbol{W}_{t+1}^m=\left\{\boldsymbol{U}_{t+1}^m,\boldsymbol{V}_{t+1}^m\right\}$。

FedDACS方法在服务器端的工作流程可由算法2描述，具体如下。

算法2 服务器端的FedDACS

输入 通信总轮数次T，学习率lr，超参数η

for训练轮数$t=1,2,3,\cdots,T$

for客户端$m=1,2,3,\cdots,M$

//服务器接收客户端m上传的信息$\boldsymbol{\Phi}^{\boldsymbol{m}}$

$\boldsymbol{\Phi}^{\boldsymbol{m}}\leftarrow f_U^m\left(\mathcal{D}^{\boldsymbol{m}}\right)$

$\boldsymbol{\Phi}=\cup_{m=1}^M\boldsymbol{\Phi}^{\boldsymbol{m}}$

计算$\zeta_{\boldsymbol{m}},\zeta_{\boldsymbol{g}}$

for客户端$m=1,2,3,\cdots,M$

//提取采样样本集$\boldsymbol{\Phi}_{\mathbf{sub}}^{\boldsymbol{m}}$

$\boldsymbol{\Phi}_{\mathbf{sub}}^{\boldsymbol{m}}=\cup_{c=1}^C\mathcal{R}\left(\boldsymbol{\Phi},\eta\cdot p_c^g+\left(1-\eta\right)\cdot p_c^m\right)$

//服务器端个性化预测器参数$\boldsymbol{V}_t^m$更新

$\boldsymbol{V}_{t+1}^m\leftarrow\boldsymbol{V}_t^m-\mathrm{lr}\nabla l_{\mathrm{server}}\left(\boldsymbol{V}_t^m;\boldsymbol{\Phi}_{\mathbf{sub}}^{\boldsymbol{m}}\right)$

//计算服务器端知识$\hat{\boldsymbol{y}}_{\boldsymbol{g}}^{\boldsymbol{m}}$

$\hat{\boldsymbol{y}}_{\boldsymbol{g}}^{\boldsymbol{m}} \leftarrow \boldsymbol{g}_{\boldsymbol{V}}^{\boldsymbol{m}}(\boldsymbol{V}_{t+1}^{m}; \boldsymbol{h}^{\boldsymbol{m}})$

回传服务器知识$\hat{\boldsymbol{y}}_{\boldsymbol{g}}^{\boldsymbol{m}}$给边缘侧

4　实验结果与分析

4.1　实验设置

本文使用 CIFAR-10 和 CIFAR-100 数据集。对于非独立同分布场景的设置和模拟，有研究[29]对目前方法进行了总结，在水平联邦学习中，针对非独立同分布的有效模拟通常分为 3 种类型：标签偏斜、特征偏斜和数量偏斜。考虑特征偏斜需要对数据集进行噪声添加或特征合成后再进行划分，本文主要模拟标签偏斜（$\#C=n$ 和 Dir (β) 对于 $\beta=0.5,1$）和数量偏斜（Qniid）这两类 Non-IID 场景。

实验配置：在 NVIDIA RTX 4090D 上使用 PyTorch 实现。客户端采用 ResNet-8[12]特征提取器；服务器采用 ResNet-18 作为预测器。FedDACS 采用学习率 0.005 进行边缘和服务器训练，批次大小为 128。优化使用随机梯度下降（SGD）算法，动量值为 0.9。边缘更新和服务器更新轮数均设置为 1。超参数 α 遵循一般设置，设为 1。$\eta=0.3$，基于验证（具体在第 4.2.4 节中讨论）进行设置。基线使用原始论文设置。

为了评估本文方法性能，将 FedDACS 与多种最先进的算法进行了对比。具体而言，所选对比算法包括两类。（1）传统联邦学习：FedAvg[1]、FedProx[5]、FedDCT[30]。（2）个性化联邦学习：Fedgkt、FedProto、GPFL[31]、FedRep[1]。为了衡量个性化模型性能，采用了平均用户模型准确率[32]（average user accuracy，Average UA）作为评估指标。UA 表示通过在本地测试数据集上验证的个性化模型的训练准确率，旨在反映个性化学习方法在实际应用中的效果。

4.2　实验结果

4.2.1　对比实验

在本文实验中，设置了 20 个客户端进行模拟，并在 2 个数据集上对提出的 FedDACS 方法与现有最先进的算法进行了性能对比。在 CIFAR-10 数据集上，对 4 种非独立同分布分割场景进行了实验，通信轮数设定为 200 轮；在 CIFAR-100 数据集上，则针对 3 种非独立同分布分割场景进行了实验，通信轮数为 600 轮。

平均用户模型准确率见表 2，展示了在 7 种数据异质性设置下，基于这 2 个数据集的测试准确度结果。通过对比分析，FedDACS 在不同 Non-IID 场景下均表现优异。具体地，根据实验结果，传统联邦学习的 3 种方法的性能都显著低于个性化联邦学习。与 Fedgkt 相比，FedAvg 的客户端平均准确率在 CIFAR-10 和 CIFAR-100 数据

表 2　平均用户模型准确率

方法	CIFAR－10				CIFAR－100		
	Qniid	Dir (0.5)	Dir (l)	$\#C=4$	Qniid	Dir (0.5)	Dir (l)
FedAvg	44.90%	44.54%	46.95%	45.52%	27.12%	27.69%	27.87%
FedProx	44.89%	43.37%	45.58%	42.54%	23.41%	24.31%	24.96%
FedDCT	49.69%	42.95%	51.78%	51.02%	40.64%	38.65%	39.55%
Fedgkt	74.29%	48.97%	75.38%	53.00%	60.19%	58.89%	59.58%
FedRep	69.47%	72.49%	67.40%	75.87%	36.86%	37.10%	33.68%
FedProto	56.62%	65.76%	56.50%	69.85%	21.63%	24.03%	19.07%
GPFL	78.03%	68.72%	65.76%	**83.38%**	37.72%	36.69%	35.14%
FedKD	57.45%	66.44%	57.39%	69.04%	23.33%	35.14%	20.69%
FedDACS	**86.34%**	**83.65%**	**84.37%**	82.33%	**64.30%**	**64.79%**	**65.43%**

集中分别低了17.43个百分点和31.99个百分点。这是由于客户端数据分布差异性导致的，说明了在联邦学习训练中构建客户个性化模型的必要性。在基于知识蒸馏的2种方法中，FedProto通过对齐类全局原型与类局部原型，即参照各客户端对该类样本的软标签的平均值，对本地模型进行更新。但仅交换类原型所带来的信息量过少，导致模型表现不尽如人意。而在这几种设置中，Fedgkt展现出较好的性能，特别是在4个特定场景下。然而，在数据异质性较强时，由于未对上传的信息进行处理，直接全部按序投入训练，而且未考虑各客户端本地的有害知识，模型的表现出现了显著下降。本文算法在其中7种设置下都实现了最优性能，在1种情况下达到了次优性能，可以看出，与现有方法相比，引入分布感知可控采样技术可以显著降低数据分布差异较大对知识迁移和模型建立带来的负面影响。

总体来看，与8个基准方法的最佳性能相比，FedDACS在CIFAR-10数据集上实现了平均用户准确率提高6.85%，在CIFAR-100数据集上则提高了5.28%。这些结果充分表明，所提出的FedDACS方法在不同数据异质性场景下，能够有效提升个性化联邦学习的性能。本文方法在面对数据异质性问题时，展现出了显著的优势，证明了其作为个性化联邦学习框架的有效性。

4.2.2 收敛速度对比

对比了FedDACS与现有个性化方法的收敛速度，具体体现在达到给定平均UA所需的通信轮数上。实验在CIFAR-10的Qniid设置下进行。

达到人定平均UA所需的通信轮数见表3，73(%)表示不同方法的分类准确率第1次达到73%所花费的通信轮数，数值越小表示收敛越快。FedDACS在达到所有给定平均UA时所需的通信轮数不超过Fedgkt的40%。由于本文方法加入了分布感知可控采样，拉近了源域与目标域的相似程度，加快了知识转移效率。

表3 达到给定平均UA所需的通信轮数

Average UA	73(%)	78(%)	82(%)
Fedgkt	161(base)	195(base)	N/A
FedRep	194(1.20×)	N/A	N/A
FedProto	N/A	N/A	N/A
GPFL	56(0.35×)	152(0.78×)	N/A
FedKD	N/A	N/A	N/A
FedDACS	45(0.28×)	66(0.34×)	108

注：符号N/A（不可用）表示该方法无法达到目标测试准确率

4.2.3 消融实验

本节探讨个性化预测器阵列与分布感知可控采样对个性化模型性能的独立影响。消融实验比较了以下3种情况：（1）FedDACS，包含个性化预测器阵列与分布感知可控采样；（2）only_personal，仅包含个性化预测器阵列而无采样的退化算法；（3）one_side_Fedgkt，既无采样也无预测器阵列，并且仅保留单侧知识迁移的Fedgkt算法。没有设置仅包含采样而无个性化预测器阵列的退化算法，理由是分布感知可控采样是为各客户端的个性化预测器提供不同分布的输入，在全局预测器上无法验证其有效性。

实验在4种不同的场景设置下进行，FedDACS及其2种退化算法在2个数据集上的平均用户模型准确率比较如图2所示。观察图2可知，FedDACS在2个数据集上的准确率显著优于其他两种算法，尤其在Dir (0.5)这种异质性较高的场景设置下优势明显。将FedDACS与基准算法only_personal比较可以发现：动态分布感知采样技术对个性化联邦性能起促进作用。例如，对于CIFAR-10数据集，FedDACS的平均用户准确率比only_personal提高近7%；在CIFAR-100数据集上，其性能甚至表现出超10%的提升。

将only_personal与基准算法one_side_Fedgkt比较，可以发现：only_personal算法相比于基准算法one_side_Fedgkt在CIFAR-10上的效果仅有小幅提升，但在CIFAR-100上的效果有所下降，这说明仅将共享预测器改为个性化预测器阵列无

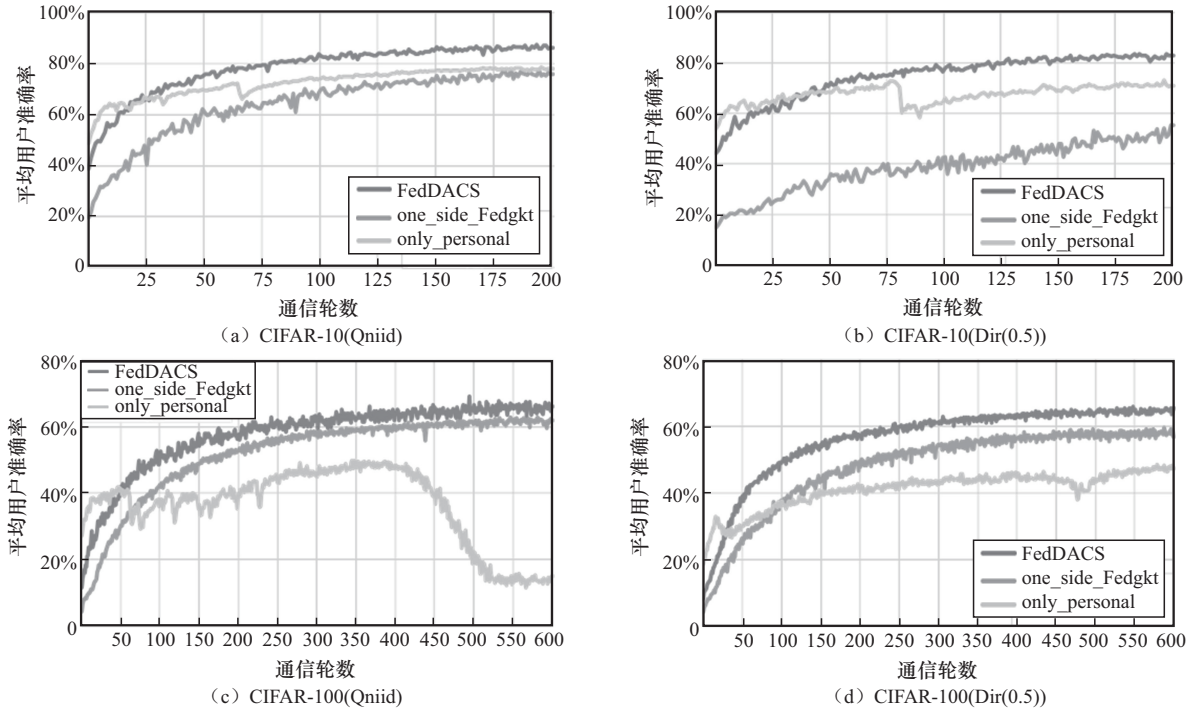

图2　FedDACS 及其 2 种退化算法在 2 个数据集上的平均用户模型准确率比较

法为联邦模型带来显著优势，甚至可能带来危害。然而，当个性化预测器阵列与分布感知动态采样机制结合使用时，即 FedDACS，其性能远超 one_side_Fedgkt。这表明，个性化联邦学习模型性能的提升主要依赖于个性化预测器阵列与分布感知动态采样机制协同作用。

4.2.4　超参数实验

η 是分布感知可控采样过程中的控制参数，取值介于 0 到 1。在采样过程中，随着参数 η 值的增加，特征空间标签分布 ζ_g 的影响权重增大。由于特征空间标签分布的影响，采样依据的分布会逐渐向特征空间标签分布靠拢，这意味着在采样时，更多数据点会被纳入考虑范围，导致采集样本量 $\left|\boldsymbol{\Phi}_{sub}^{m}\right|$ 增加。样本量的增加直接导致服务器端的计算开销上升。为了量化这种影响，统计了不同 η 值对应的采样比例（采样点数占全部数据点数的百分比），不同超参数设置下的采样比例和每个客户端平均每轮训练的计算时间见表 4。如在 $\eta=0.3$ 时，采样比例平均为 91.4%，而在 $\eta=1$ 时，采样比例增加到 99.8%。此外，通过记录每轮训练的平均时间观察到，随着采样比例的降低（如在 $\eta=0.3$ 的设置下），每个客户端平均每轮训练的计算时间减少了约 7%。这表明，合理调节 η 可以在保证模型性能的同时，有效降低计算资源的开销。

表 4　不同超参数设置下的采样比例和每个客户端平均每轮训练的计算时间

η	0	0.3	0.5	1
采样比例	50.67%	91.38%	95.85%	99.88%
计算时间/s	5.75	9.83	10.21	10.57

为了评估超参数 η 对个性化模型性能的影响，并寻求在最小采样量（即 η 尽可能小）的情况下，同时保证模型的个性化能力和基类泛化能力，基于 CIFAR-10 数据集的 Qniid 数据分割进行了实验。在本地类测试集和基类测试集上进行了详细的实验分析，其中，本地类测试集指的是符合各客户端原有数据分布的测试集，包含本地数据中拥有的类；基类测试集则是将所有客户端的本地

类测试数据汇总而成，涵盖所有标签类别。不同超参数设置下本地类测试集和基类测试集上的平均用户模型准确率见表5。

表5　不同超参数设置下本地类测试集和基类测试集上的平均用户模型准确率

η	0	0.3	0.5	1
本地类测试集	87.48%	86.34%	85.88%	86.10%
基类测试集	46.84%	83.23%	84.34%	84.21%

实验结果表明，当$\eta=0$时，FedDACS在本地类测试集上表现出最佳性能，体现了其卓越的个性化能力。然而，此时模型在基类测试集上的准确率低于50%，表明过度强调个性化可能显著牺牲基类泛化能力。随着η值增大至0.3，FedDACS在本地类测试集上的性能略有下降，但基类测试集上的准确率却显著提升。进一步地，当η逐渐增加到1时，本地类测试集与基类测试集上的准确率均趋于稳定。考虑本文目标是在尽可能减少数据采样量（即降低η值）的前提下，兼顾模型在本地类和基类上的良好表现，因此选择将η设置为0.3。

4.2.5　在大规模边缘设备上的实验

FedDACS是在边缘智能场景下改进的个性化群体知识迁移方法，为了解其可扩展性，评估了它在不同边缘客户端数量上的性能。实验在CIFAR-10的Qniid设置下进行，不同客户端数量下本地类测试集上的平均用户模型准确率见表6。从实验数据来看，添加更多的边缘客户端不会对精度产生负面影响。

表6　不同客户端数量下本地类测试集上的平均用户模型准确率

客户端数量	5	10	20	50	100	200
Average UA	79.50%	85.87%	86.34%	86.56%	85.88%	86.30%

4.2.6　公平性实验

本节实验基于CIFAR-10的Qniid设置展开，对不同客户端在FedDACS联邦架构下所获得的个性化模型性能展开实验分析。CIFAR-10-Qniid下各客户端数据分布如图3所示。

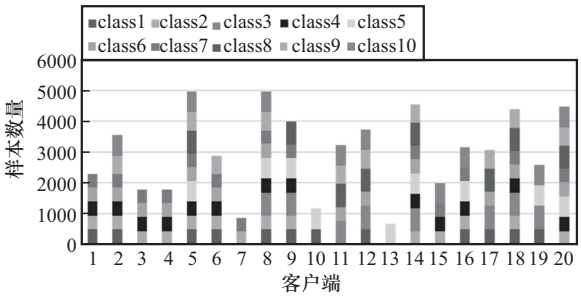

图3　CIFAR-10-Qniid下各客户端数据分布

公平性实验结果如图4所示，每个客户端的3列从左到右依次为：纯本地训练模型在本地类别上的表现；在FedDACS框架下训练的个性化模型在本地类上的表现；在FedDACS框架下训练的个性化模型在基类上的泛化表现。模型性能使用Top-1准确率指标进行评估。在数据类别极度单一的客户端（如客户端7、10、13，其本地类别

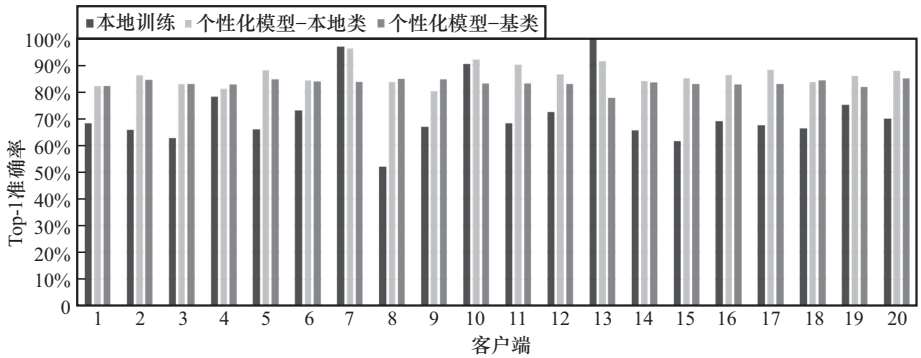

图4　公平性实验结果

不超过2类），参与FedDACS联邦训练后，本地类别准确率出现轻微下降（平均下降2.3%），但基类泛化能力达到群体平均水平（均值83.23%±1.7%），实现了基类泛化。其他客户端则呈现双重增益，在保持个性化模型在本地任务性能提升的同时，获得显著的基类适应性。

实验结果揭示了FedDACS内在的公平性机制，对于极端数据稀疏的客户端，该框架通过在局部性能上做出一些妥协，提升基类泛化能力；而对多数客户端，个性化联邦训练过程既优化了本地目标，同时也没有损害跨域的适应性。这种差异化策略很好地满足了个性化联邦学习在异构环境中的公平性需求——它通过动态权衡局部知识和全局共识，避免了弱势客户端陷入“过度个性化”的问题。

5　结束语

本文提出了一种新型的个性化群体知识迁移框架FedDACS。该方法将每个客户端的个性化模型分为两部分，分别部署在边缘设备和服务器端，并通过交替最小化来优化该模型。在优化过程中，本文引入了分布感知可控采样，以确保每个客户端能够学习到与其特定数据分布相关的知识。在不同数据集上进行了大量的实验，显示了FedDACS在个性化方面的优势。在未来的工作中，将致力于探索个性化联邦知识蒸馏在解决模型异构性问题方面的有效性，力求提高不同设备和客户端之间模型的兼容性和协同工作能力。

参考文献：

[1] Tan A Z, Yu H, Cui L Z, et al. Towards personalized federated learning[J]. IEEE Transactions on Neural Networks and Learning Systems, 2023, 34(12): 9587-9603.

[2] Fallah A, Mokhtari A, Ozdaglar A E. Personalized federated learning with theoretical guarantees: a model-agnostic meta-learning approach[C]//Proceedings of the 2020 Neural Information Processing Systems. Montreal: Neural Information Processing Systems, 2020.

[3] Zhao Y, Li M, Lai L Z, et al. Federated learning with non-IID data[EB]. 2018.

[4] Yang M, Wang X M, Zhu H B, et al. Federated learning with class imbalance reduction[C]//Proceedings of the 2021 29th European Signal Processing Conference (EUSIPCO). Piscataway: IEEE Press, 2021: 2174-2178.

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

[6] Li D L, Wang J P. FedMD: heterogenous federated learning via model distillation[EB]. 2019.

[7] Liang P P, Liu T, Liu Z Y, et al. Think locally, act globally: federated learning with local and global representations[EB]. 2020.

[8] SMITH V, CHIANG C K, SANJABI M, et al. Federated multi-task learning[C]//Proceedings of the 2025 Neural Information Processing Systems. Montreal: Neural Information Processing Systems, 2025.

[9] Deng Y Y, Kamani M M, Mahdavi M. Adaptive personalized federated learning[EB]. 2020.

[10] Lin T, Kong L J, Stich S U, et al. Ensemble distillation for robust model fusion in federated learning[EB]. 2020.

[11] Chen H C, Wang J, Haris V, et al. The best of both worlds: accurate global and personalized models through federated learning with data-free hyper-knowledge distillation[EB]. 2023.

[12] He C Y, Annavaram M, Avestimehr S. Group knowledge transfer: federated learning of large CNNs at the edge[EB]. 2020.

[13] Wu Z Y, Sun S, Wang Y W, et al. Exploring the distributed knowledge congruence in proxy-data-free federated distillation[J]. ACM Transactions on Intelligent Systems and Technology, 2024, 15(2): 1-34.

[14] Wu Z Y, Sun S, Wang Y W, et al. FedICT: federated multi-task distillation for multi-access edge computing[J]. IEEE Transactions on Parallel and Distributed Systems, 2024, 35(6): 1107-1121.

[15] Itahara S, Nishio T, Koda Y, et al. Distillation-based semi-supervised federated learning for communication-efficient collaborative training with non-IID private data[J]. IEEE Transactions on Mobile Computing, 2023, 22(1): 191-205.

[16] Chang H Y, Shejwalkar V, Shokri R, et al. Cronus: robust and heterogeneous collaborative learning with black-box knowl-

edge transfer[EB]. 2019.

[17] Cho Y J, Wang J Y, Chiruvolu T, et al. Personalized federated learning for heterogeneous clients with clustered knowledge transfer[EB]. 2021.

[18] Zhang J, Guo S, Ma X, et al. Parameterized knowledge transfer for personalized federated learning[C]//Proceedings of the 2021 Neural Information Processing Systems. Montreal: Neural Information Processing Systems, 2021, 34: 10092-10104.

[19] Cheng S J, Wu J W, Xiao Y H, et al. FedGEMS: federated learning of larger server models via selective knowledge fusion[EB]. 2021.

[20] Xie C L, Huang D A, Chu W D, et al. Perada: parameter-efficient federated learning personalization with generalization guarantees[C]//Proceedings of the 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2024: 23838-23848.

[21] Wu Z Y, Sun S, Wang Y W, et al. FedCache: a knowledge cache-driven federated learning architecture for personalized edge intelligence[J]. IEEE Transactions on Mobile Computing, 2024, 23(10): 9368-9382.

[22] Tan Y, Long G D, Liu L, et al. FedProto: federated prototype learning across heterogeneous clients[J]. Proceedings of the AAAI Conference on Artificial Intelligence. Menlo Park: AAAI Press, 2022, 36(8): 8432-8440.

[23] Chen Y Q, Lu W, Qin X, et al. MetaFed: federated learning among federations with cyclic knowledge distillation for personalized healthcare[J]. IEEE Transactions on Neural Networks and Learning Systems, 2024, 35(11): 16671-16682.

[24] 庄福振, 罗平, 何清, 等. 迁移学习研究进展[J]. 软件学报, 2015, 26(1): 26-39.
Zhuang F Z, Luo P, He Q, et al. Survey on transfer learning research[J]. Journal of Software, 2015, 26(1): 26-39.

[25] Lee G, Jeong M, Shin Y, et al. Preservation of the global knowledge by not-true distillation in federated learning[EB]. 2021.

[26] He Y T, Chen Y Q, Yang X D, et al. Class-wise adaptive self distillation for federated learning on Non-IID data (student abstract)[C]//Proceedings of the 2022 AAAI Conference on Artificial Intelligence. Menlo Park: AAAI Press, 2022, 36(11): 12967-12968.

[27] Karimireddy S P, Kale S, Mohri M, et al. Scaffold: stochastic controlled averaging for federated learning[C]//Proceedings of the 2020 Conference and Workshop on Neural Information Processing Systems. Montreal: Neural Information Processing Systems, 2020: 5132-5143.

[28] Yao D Z, Pan W N, Dai Y T, et al. Local-global knowledge distillation in heterogeneous federated learning with Non-IID data[EB]. 2021.

[29] Li Q B, Diao Y Q, Chen Q, et al. Federated learning on Non-IID data silos: an experimental study[C]//Proceedings of the 2022 IEEE 38th International Conference on Data Engineering (ICDE). Piscataway: IEEE Press, 2022: 965-978.

[30] Nguyen Q, Pham H H, Wong K S, et al. FedDCT: federated learning of large convolutional neural networks on resource-constrained devices using divide and collaborative training[J]. IEEE Transactions on Network and Service Management, 2024, 21(1): 418-436.

[31] Zhang J Q, Hua Y, Wang H, et al. GPFL: simultaneously learning global and personalized feature information for personalized federated learning[C]//Proceedings of the 2023 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway: IEEE Press, 2023: 5018-5028.

[32] Mills J, Hu J, Min G Y. Multi-task federated learning for personalised deep neural networks in edge computing[J]. IEEE Transactions on Parallel and Distributed Systems, 2022, 33(3): 630-641.

[作者简介]

齐霄超（2001– ），女，浙江工商大学信息与电子工程学院硕士生，主要研究方向为个性化联邦学习。

倪郑威（1989– ），男，博士，浙江工商大学信息与电子工程学院副研究员、硕士生导师，主要研究方向为人工智能、无线网络。