doi:10.14132/j.cnki.1673-5439.2025.05.013

面向Non-IID场景的通信高效个性化联邦学习算法

黄宇虹,陈思光

(南京邮电大学 物联网学院,江苏 南京 210003)

摘要:在客户端数据非独立同分布(Non-Independent and Identically Distributed, Non-IID)的场景中,为了向客户端提供个性化且通信高效的解决方案,提出了一种面向Non-IID场景的通信高效个性化联邦学习算法。具体地,为充分利用相似客户端之间的知识提升模型性能,同时保留本地客户端的个性化信息,提出一种融合模型分层思想与聚类思想的个性化联邦学习算法。为进一步解决通信开销高的问题,设计选择性模型聚合策略,在中心服务器通过最大均值差异评估客户端数据分布与全局数据分布的相似性,并基于相似性计算各客户端优先级分数,选择优先级较高的客户端进行通信。该策略可有效减少客户端与中心服务器的累计通信次数,提高通信效率并加速模型收敛。最后,仿真实验结果表明,相较于其他联邦学习算法,所提算法能够在保证高准确率的前提下,将累计通信次数减少至少50%。

关键词:个性化联邦学习;聚类;选择性聚合;非独立同分布

中图分类号:TP393 **文献标志码**:A **文章编号**:1673-5439(2025)05-0111-08

Communication-efficient personalized federated learning for Non-IID scenarios

HUANG Yuhong, CHEN Siguang

(School of Internet of Things, Nanjing University of Posts and Telecommunications, Nanjing 210003, China)

Abstract: In scenarios where client data are non-independent and identically distributed (Non-IID), this paper proposes a communication-efficient personalized federated learning algorithm based on Non-IID data to provide personalized and efficient communication solutions for clients. Specifically, to leverage the knowledge among similar clients and retain personalized information of local clients, we develop a personalized federated learning algorithm that combines hierarchical modeling with clustering ideas. Furthermore, to address the issue of high communication overhead, we design a selective model aggregation strategy. The central server evaluates the similarity between the client data distribution and the global data distribution using maximum mean discrepancy. Based on this similarity, the server computes a priority score for each client and selects those with higher scores for communication. This strategy effectively reduces the cumulative communication rounds between clients and the central server, thereby improving communication efficiency and accelerating model convergence. Experimental results demonstrate that compared with existing representative works, the proposed algorithm reduces the cumulative commu-

收稿日期:2024-11-09;**修回日期**:2025-03-06 **本刊网址**:http://nyzr.njupt.edu.cn

基金项目:国家自然科学基金(61971235)、江苏省"333高层次人才培养工程"和江苏省研究生科研与实践创新计划项目(KYCX24_1205)资助项目

作者简介:黄宇虹,女,硕士研究生;陈思光(通信作者),男,博士,教授,sgchen@njupt.edu.cn

引用本文:黄宇虹,陈思光.面向Non-IID场景的通信高效个性化联邦学习算法[J].南京邮电大学学报(自然科学版),2025,45(5):111-118.

Citation: HUANG Yuhong, CHEN Siguang. Communication-efficient personalized federated learning for Non-IID scenarios[J]. Journal of Nanjing University of Posts and Telecommunications(Natural Science Edition), 2025, 45(5): 111-118.

nication rounds by over 50% while maintaining high accuracy.

Keywords: personalized federated learning; clustering; selective aggregation; non-independent and identically distributed

在传统的机器学习模型训练过程中,中心服务器需要收集所有客户端数据来训练模型,这种方法虽然简便,却容易暴露数据隐私,致使许多有价值的数据由于其敏感性而无法被访问和利用[1]。在此背景下,联邦学习(Federated Learning, FL)应运而生,其解决了数据共享与隐私保护之间的矛盾,协调了隐私保护与多方协同训练之间的平衡性,并且在移动设备中有很好的鲁棒性,可以提高模型训练的灵活性和效率[2-3]。

然而,联邦学习主要依靠中心服务器聚合客户端本地模型以获得高质量的全局模型,这通常难以捕捉每个设备的个性化信息,对于一些需要精细个性化的任务是一个挑战。此外,设备异构性、数据异构性以及模型异构性的存在也会对全局模型推理或分类性能造成一定影响[4-5]。为了更好地利用每个设备的本地数据,往往在联邦学习的基础上应用个性化策略,以提升建模效果,为用户提供定制化服务。通过对各个客户端应用不同的学习网络和提供针对性的模型参数,个性化联邦学习能够灵活地适应不同客户端数据,从而更好地解决数据非独立同分布(Non-Independent and Identically Distributed, Non-IID)等异构性问题[6]。

在联邦学习框架中,可以采取多种方法实现个性化[7]。部分研究采用"联邦学习+本地适应"策略,首先训练一个强泛化能力的全局模型,然后通过本地适应为每个客户端实现个性化。这类个性化联邦学习方法旨在提高异构数据下全局模型的性能,以进一步提升模型在本地数据上的个性化性能。Hanzely等[8]提出了无循环梯度下降(Loopless Local Gradient Descent, L2GD)算法,并通过优化本地模型与全局模型的混合优化目标来实现个性化联邦学习。Chen等[9]采用迁移学习的方式为各客户端构建个性化模型。Wu等[10]提出FedHome算法,一种用于家庭健康监测的云边协同的联邦学习算法,设置了一种生成式卷积自动编码器网络,在云端学习所有用户的粗粒度特征,在边缘侧使用基于用户个人数据的重构类平衡数据集进行模型重训练,实现准确和个性化的健康监测。上述方法依赖于全局模型的泛化能力,在训练的过程中倾向于训练一个泛化能力强的全局模型,再进行本地适应,这会导致许多与用户行为相关的宝贵信息丢失。

为了提供更具个性化的解决方案,部分研究通过建模客户关系来实现个性化,数据分布差异较大的客户端之间不会相互干扰,而数据分布相似的客户端之间可以相互增强。文献[11]提出了聚类联邦学习算法,借助基于余弦相似性的二分法将客户端划分至不同的聚类簇,簇内训练单独的全局模型,从而为该聚类簇中的客户端提供定制模型。文献[12]基于客户端的优化方向相似性对客户端进行分组,以实现高模型性能。除此之外,部分研究着眼于模型结构优化,通过设置部分个性化参数层达到模型个性化的目的。文献[13]指出使用小部分个性化参数也能达到全模型个性化的效果,在深度学习中只对模型的特定组件进行个性化可以提高性能并减少内存占用,同时降低灾难性遗忘的风险。文献[14]中提出的FedPer算法是典型的部分模型个性化,该算法提出将训练模型分层,将模型参数视为共享层与个性化层两组参数,共享层进行全局训练,而个性化层仅在本地训练,实验表明个性化层可以克服异构数据带来的不良影响。进一步地,文献[15]提出了FedRep算法,该算法丰富了FedPer算法的理论依据,并在其基础上改变了共享层与个性化层的更新时间节点,交替进行共享层与个性化层的更新,提升个性化能力。

上述方案往往聚焦于联邦学习模型个性化方法而忽视了联邦学习中的另一个挑战,即联邦学习中通信开销大的问题。为了减少联邦学习训练过程的通信开销,文献[16]对联邦学习中的卷积神经网络模型提出了针对性的模型压缩方案,采用参数近似和参数选择的算法,减少了上行和下行数据传输量。文献[17]提出了一种计算分层的联邦学习框架来实现轻量级消息传输。文献[18]结合了对客户端模型的动态采样和对客户端模型神经参数的选择性遮蔽这两种策略来提高传统联邦平均学习的通信效率。文献[19]引入了最大均值差异(Maximum Mean Discrepancy, MMD),通过最小化全局模型和本地模型输出之间的MMD损失,本地模型能够从全局模型学习到更多知识,从而加速训练过程的收敛,减少通信轮次。文献[20]从参与训练的设备中排除不利于全局模型收敛的本地模型,利

用剩余的本地模型完成全局模型的聚合更新，增加系统的训练准确率。

上述研究方案分别阐述了当前联邦学习的个性化解决方案和通信开销解决方案，但在个性化联邦学习中对高效通信的研究与分析较少，同时在通信高效类联邦学习中也缺乏对个性化方案的研究。因此，为应对个性化联邦学习的高通信开销问题并进一步缓解非独立同分布数据带来的负面影响，本文提出了一种面向Non-IID场景的通信高效个性化联邦学习算法，主要贡献总结如下：

（1）构建了一个面向Non-IID场景的通信高效个性化联邦学习。中心服务器采用基于K-means的聚类方法将客户端分组，使客户端在数据异构的情况下也能充分学习相似客户端的知识。训练模型分为共享层与个性化层，在客户端与中心服务器通信时仅传输共享层的参数。

（2）提出了一种选择性模型聚合策略。中心服务器计算客户端数据分布与全局分布的相似性，并根据该相似性计算客户端优先级分数，选择优先级较高的客户端参与通信。这一策略可有效减少客户端与中心服务器的累计通信次数，提高通信效率，加速模型的收敛。

（3）大量的仿真结果表明本文所提的面向Non-IID场景的通信高效个性化联邦学习算法降低了数据异构带来的负面影响，减少了模型训练时的累计通信次数，提高了通信效率，在保证准确率的前提下，可将累计通信次数减少至少50%。

1　系统模型

本文构建了一个面向Non-IID场景的通信高效个性化联邦学习模型，如图1所示，该网络模型由用户层与中心服务器层两部分组成，具体每层的功能定义与关联如下：

（1）用户层。用户层具有N个客户端，各客户端都拥有各自的本地数据集D_i，各客户端之间的本地数据非独立同分布，例如地理位置差异、数据采集方式差异、隐私和法律限制等因素造成的数据间分布差异。用户层具有以下功能：(a)模型训练。在客户端中，训练模型由共享层与个性化层组成。客户端从中心服务器下载共享层模型参数，并将其与本地的个性化层参数进行融合，生成更新后的本地模型，该模型既包含了全局知识，又包含了来自个性化层的本地知识。接着使用本地数据对模型进行训练更新，本地训练是对完整模型的一次或多次训练。(b)特征中心计算。对训练过程中产生的样本嵌入向量进行统计，计算本地数据类别中心，为中心服务器的聚类和选择性模型聚合策略提供支持。(c)参数上传。本地训练结束后，客户端就会上传共享层模型参数与本地类别中心至中心服务器。

（2）中心服务器层。中心服务器层包含一个中心服务器，中心服务器具有低延迟、高吞吐以及高可靠性等特点，能够处理大量数据，承担了联邦学习算法的客户端聚类、参数聚合、模型更新等重要任务。中心服务器层具有以下功能：(a)聚类。基于客户端特征中心，采用K-means方法对客户端进行聚类，以便更有效地识别客户端之间的主要差异，并优化知识共享的效率。(b)选择性模型聚合。计算客户端特征中心与全局特征中心的相似性，并根据该相似性计算客户端优先级分数，选择优先级较高的客户端参与通信。接着，将接收到的共享层模型进行聚合，并将更新后的全局共享层模型下发至各客户端。

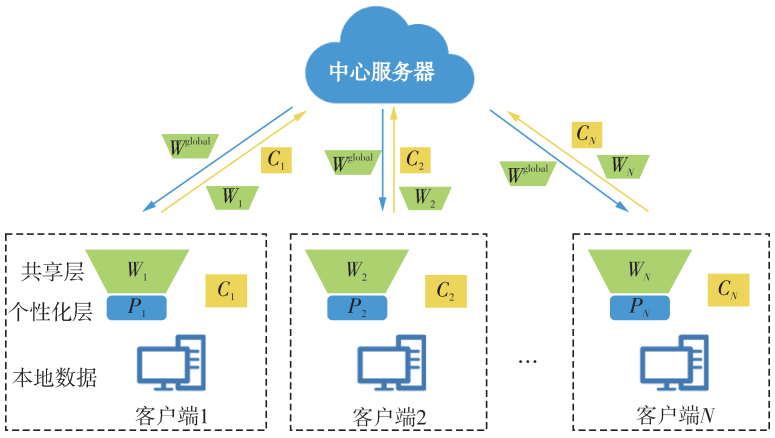

图1　网络模型

2 本文算法

2.1 客户端层模型训练与特征中心计算

本文客户端模型采用“共享层+个性化层”的分层构建方法。常见的任务模型通常由特征提取器与分类器组成,其中特征提取器负责从输入数据中提取通用特征,视为共享层;而分类器则根据每个客户端的数据特征进行个性化的分类,因此视为个性化层。在与中心服务器通信时,仅上传共享层模型。

模型训练 客户端i利用本地训练数据集$(x,y)\in D_i$更新本地模型(W_i,P_i),其中W_i为共享层模型,P_i为个性化层模型。在接收到中心服务器下发的全局共享层W_{t-1}^{global}后,客户端i通过本地梯度更新来优化个性化层模型参数。客户端i的共享层与个性化层模型更新如下

$$(W_{t,i},P_{t,i})=\text{SGD}(W_{t-1}^{\text{global}},P_{t-1,i}) \tag{1}$$

其中,$(W_{t,i},P_{t,i})$表示融合后的模型,SGD表示随机下降梯度(Stochastic Gradient Descent, SGD)[21]算法。

特征中心计算 在模型训练的过程中,客户端i会存储每个训练样本x在共享层模型下的嵌入向量$\boldsymbol{f}_i(W_i,x)$,该嵌入向量是样本的低维稠密表示,能够有效捕捉数据的高层次语义特征。随后,对本地各类别k的样本嵌入向量求平均,计算类别中心$\boldsymbol{C}_{i,k}$,即

$$\boldsymbol{C}_{i,k}=\frac{1}{|D_{i,k}|}\sum_{(x,y)\in D_{i,k}}\boldsymbol{f}_i(W_i,x) \tag{2}$$

计算得到的类别中心$\boldsymbol{C}_{i,k}$代表了类别k的语义特征。接着,客户端将包含本地所有类别的类别中心集合$\boldsymbol{C}_i=\{\boldsymbol{C}_{i,k}\}_{k=1}^{K}$、本地平均训练损失$l_{t,i}$与$W_i$一同上传至服务器进行全局更新。

2.2 中心服务器层的聚类与选择性模型聚合

聚类 聚类操作促使数据分布相似的客户端协同训练,从而优化模型训练过程。本文基于客户端本地数据集的类别中心相似度进行聚类,并在训练损失稳定后执行聚类,以确保模型性能持续提升。

(1)损失监测。本文采用滑动窗口的方法检测损失的变化趋势。在每一轮全局训练中,服务器将客户端的平均训练损失值记录到一个固定长度为M的滑动窗口中,形成损失序列。接着对滑动窗口内的损失值进行线性回归,拟合出损失值关于训练轮次的变化直线,即

$$L_t=a\cdot t+b \tag{3}$$

其中,L_t表示第t轮的损失值,a为直线的斜率,反映损失的变化趋势。当训练损失趋于平稳,服务器进入聚类状态。

(2)聚类操作。在聚类阶段,本文为每个构建客户端i构建类别中心向量$\boldsymbol{v}_i$,该向量由类别中心集合$\boldsymbol{C}_i$中的向量拼接组成。但在数据异构的场景中,某些类别k在客户端i上可能没有样本数据。为减少类别缺失对聚类的影响,本文使用全局统计信息$\boldsymbol{C}^{\text{global}}=\{\boldsymbol{\mu}_1,\boldsymbol{\mu}_2,\cdots,\boldsymbol{\mu}_k\}$对缺失的类别进行填充,其中,$\boldsymbol{\mu}_k$是类别$k$的全局平均中心向量,定义为

$$\boldsymbol{\mu}_k=\frac{1}{|N_k^{\text{clients}}|}\sum_{i\in N_k^{\text{clients}}}\boldsymbol{C}_{i,k} \tag{4}$$

其中,N_k^{clients}表示所有拥有类别k的客户端子集。将所有客户端特征向量组成矩阵$\boldsymbol{F}$,即

$$\boldsymbol{F}=\begin{bmatrix}\boldsymbol{v}_1\\ \vdots\\ \boldsymbol{v}_N\end{bmatrix} \tag{5}$$

接着,使用K-Means算法对矩阵$\boldsymbol{F}$进行聚类

$$\text{K-Means}(\boldsymbol{F},num)\rightarrow cluster \tag{6}$$

其中,num为聚类数,$cluster$为聚类结果。聚类完成后,服务器对每个聚类簇分别执行联邦学习训练,并将各簇的全局模型融合,形成增强后的全局模型。随后,聚类簇结构被释放,服务器恢复对所有客户端的统一联邦学习训练。

选择性模型聚合 在传统的联邦学习中,服务器在每轮通信中通常随机选择一定比例的客户端参与聚合,这种方式在数据分布独立同分布(Independent and Identically Distributed, IID)场景中简单有效,但在Non-IID场景中未充分考虑客户端数据异构性对模型收敛的影响。因此,为了稳定提升模型的收敛速度并减少通信轮次,本文提出了一种选择性模型聚合策略。该策略首先为每个客户端计算一个优先级分数,考虑了客户端本地数据的特征分布与全局数据分布的相似性以及客户端的历史参与情况。根据计算得到的优先级分数,服务器选择优先级较高的客户端进行通信,从而加速模型的收敛,并减少通信轮次。

(1)相似性度量。本文采用MMD来量化$\boldsymbol{C}_i$与$\boldsymbol{C}^{\text{global}}$的相似性。MMD是一种用于比较两个概率分布$p$和$q$相似性的统计指标,其核心思想是在一个函数集$\mathcal{F}$中寻找到一个函数$f$,使得两个分布中样本在该函数下的均值差异达到最大值,该最大值就是MMD值。MMD越小,则两个分布越相似,可表达为

$$\text{MMD}(\mathcal{F},p,q)=\sup_{f\in\mathcal{F}}(\mathbb{E}_{a\sim p}[f(a)]-\mathbb{E}_{b\sim q}[f(b)]) \tag{7}$$

给定观测数据 $A=\{a_1,a_2,\cdots,a_m\}$ 和 $B=\{b_1,b_2,\cdots,b_n\}$,其中 A 和 B 分别是概率分布 p 和 q 的独立同分布样本,则可用样本上计算的经验均值代替总体期望。为了使计算可行,通常引入核函数 $k(a,b)$ 作为非线性映射。当函数集被限定为再生核希尔伯特空间上的单位球时,根据文献[22]可以计算出样本空间上MMD值的近似估计,即

$$\mathrm{MMD}[F,A,B]=\left(\frac{1}{m^2}\sum_{i,j=1}^{m}k(a_i,a_j)-\frac{2}{mn}\sum_{i,j=1}^{m,n}k(a_i,b_j)+\frac{1}{n^2}\sum_{i,j=1}^{n}k(b_i,b_j)\right)^{\frac{1}{2}} \tag{8}$$

然而,由于联邦学习中客户端数据不出本地,直接访问 A 与 B 是不可行的。因此,本文利用特征中心 $\boldsymbol{C}_i$ 代表数据分布,并采用高斯核函数 $k(a,b)=\exp(-\gamma\|a-b\|^2)$,$\gamma$ 为超参数。最终将本地类别中心 $\boldsymbol{C}_i$ 与全局类别中心 $\boldsymbol{C}^{\mathrm{global}}$ 代入式(8),即可计算MMD值。

(2)客户端选择。为解决部分客户端长期未被选中导致其训练贡献不足的问题,本文引入时间衰减项 $1-\mathrm{e}^{-(t-t_i')}$,其中 t 是当前轮次,t_i' 是客户端 i 上次参与训练的全局轮次,用于提升这些客户端的优先级。上传优先级分数计算公式可表示为

$$\mathrm{score}_{t,i}^{\mathrm{priority}}=\alpha\cdot\mathrm{MMD}(\boldsymbol{C}_i,\boldsymbol{C}^{\mathrm{global}})+\beta\cdot\left(1-\mathrm{e}^{-(t-t_i')}\right) \tag{9}$$

为了避免客户端优先级频繁变动导致模型收敛不稳定,本文采用指数平滑对当前优先级进行平滑处理,最终得出每个客户端的最终优先级分数,即

$$\mathrm{score}_{t,i}^{\mathrm{priority}}=\lambda\cdot\mathrm{score}_{t,i}^{\mathrm{priority}}+(1-\lambda)\cdot\mathrm{score}_{t-1,i}^{\mathrm{priority}} \tag{10}$$

其中,λ 为平滑因子。通过上述方法,综合考虑客户端本地数据分布和历史参与情况,动态调整客户端优先级,从而平衡各客户端的训练贡献,并提高模型性能和收敛效率。服务器计算得到 $\mathrm{score}_t^{\mathrm{priority}}$ 后,选择排名前一定比例的客户端参与本轮通信。这一过程与传统的联邦学习方法类似,但更加注重客户端的个性化数据分布和历史参与情况。最后,服务器对收到的本地模型进行聚合,得到全局模型,即

$$W_t^{\mathrm{global}}=\frac{1}{|I_t|}\sum_{i\in I_t}W_{t,i} \tag{11}$$

其中,I_t 为服务器接收到的客户端子集,然后服务器将全局模型下发给各服务器进行下一轮的全局训练。

为更好地理解本文提出的面向Non-IID场景的通信高效个性化联邦学习算法,将上述训练过程归纳为算法1。

算法1　面向Non-IID场景的通信高效个性化联邦学习算法

输入:

训练数据集 D_i,迭代轮次 T,参与率 r。

输出:

客户端 i 的个性化联邦学习模型 (W_i,P_i),更新后的全局模型 W^{global}。

初始化客户端模型;
for t **in** $(1,\cdots,T)$ **do**
　for i **in** I_t **do**
　　客户端 i 进行以下操作:
　　$(W_{t,i},P_{t,i})\leftarrow(W_{t-1}^{\mathrm{global}},P_{t-1,i})$;
　　根据式(1)训练更新本地模型得到 $(W_{t,i},P_{t,i})$;
　　根据式(2)计算类别中心 $\boldsymbol{C}_{i,k}$;
　　将 $\boldsymbol{C}_i=\{\boldsymbol{C}_{i,k}\}_{k=1}^{K}$、本地平均训练损失 $l_{t,i}$ 以及 $W_{t,i}$ 发送到服务器;
　end for
　中心服务器进行以下操作:
　根据式(11)聚合模型,得到 W_t^{global};
　根据式(3)检测损失变化趋势,得到斜率 a;
　if $a\geqslant 0$ **then**
　　根据式(6)聚类得到 *cluster*,对每个聚类簇分别执行一轮联邦学习训练;
　　将各簇的全局模型融合,形成增强后的全局模型 W_t^{global};
　end if
　根据式(10)计算各客户端上传优先级分数,选择排名前 $r\cdot N$ 数量的客户端 I_t 进行通信;
end for

3　仿真实验与性能评估

3.1　实验设置

数据集　本文采用了CIFAR-10和CIFAR-100数据集[23]。CIFAR-10包含60 000张32×32像素的RGB彩色图片,涵盖10个类别,每个类别包含6 000张图像,其中50 000张用于训练,10 000张用于测试。CIFAR-100数据集是CIFAR-10的扩展,包含100个类别,每个类别有600张图像,共计60 000张32×32像素的RGB图像。训练数据以非独立同分布方式分发给 N 个用户,模拟客户端数据分布不均的实际场景。

基准方案　实验对比了3种经典算法,联邦平均算法(Federated Averaging, FedAvg)[24]、联邦表示学习算法(Federated Representation Learning, FedRep)[15]和主体聚合与更新的联邦平均算法(Federated Averaging with Body Aggregation and Body Update, FedBABU)[25]。FedAvg是最典型的联邦学习算法,其使用中心服务器对客户端上传的参数进行平均再下发。FedRep引入了一种新颖的联邦表示

学习框架,在数据异构性环境中学习共享的低维表示,同时为每个客户端构建个性化的本地分类器。在FedBABU中,模型的头部部分保持不变,而仅更新模型的主体,从而增强模型在数据异构情况下的个性化能力。

模型与参数设置 实验中使用的CNN模型由4个基础卷积单元和一层全连接层组成,前者用于共享的特征提取,后者作为个性化层,根据客户端本地数据进行分类调整。实验中,局部迭代次数设置为5次,全局迭代次数设置为100次,batch大小为32,学习率为0.01。实验环境包括1个中心服务器和20个本地客户端,客户端参与率设置为0.5。

3.2 仿真实验

本节通过仿真实验来验证面向Non-IID场景的通信高效个性化联邦学习算法的有效性,并与其他经典联邦学习方案作对比,显示本文所提算法的优越性能。为了进一步评判算法的有效性,本文采用平均准确率与累计通信次数作为评判算法有效性的标准。设True_i是客户端正确预测的样本数量,S_i是客户端i的测试样本总数,平均准确率的计算为

$$\mathrm{Accuracy}=\frac{1}{N}\sum_{i=1}^{N}\frac{\mathrm{True}_i}{S_i} \tag{12}$$

图2和图3分别展示了学习率(Learning Rate,LR)设置为0.1、0.01和0.005时产生的准确率变化和损失变化。从图中可以看到,随着全局迭代轮次的增加,平均准确率在不同学习率设置下均逐步提升并最终趋于稳定,损失值也随着训练过程的进行逐渐降低并趋于收敛,表明本文算法设计的合理性。

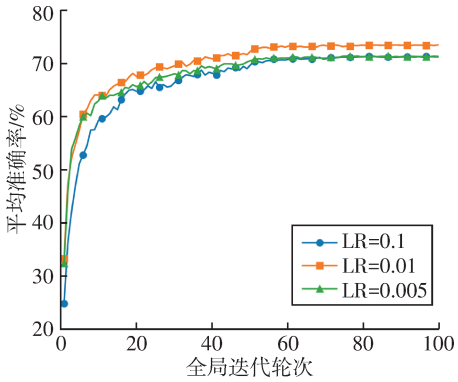

图2 不同学习率下的平均准确率变化

为了显示本文算法应对异构性数据的有效性以及通信高效性,与现有的3种联邦学习算法进行对比。

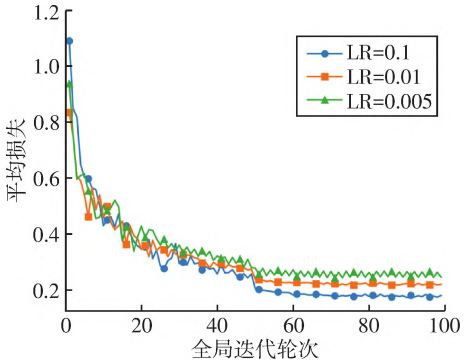

图3 不同学习率下的平均损失

图4展示了在N=20个客户端、LR=0.01的设置下,本地类别数量变化时各算法的平均准确率变化。本地类别数量越少,客户端之间的异构程度越高。从图4中可以看出,本文算法准确性均高于其他3种算法,体现了本文算法的优越性。这是因为通过采用分层策略,可以在共享层上学习更通用的特征,同时在个性化层中保留客户端本身特性,从而缓解了数据异构性带来的影响。此外,本文提出的选择性模型聚合策略通过筛选与全局原型相似度较高的客户端进行通信,减少了异构性对模型聚合的干扰,提升了共享层模型的全局特征表征能力,使得总体准确率优于FedRep和FedBABU两种个性化算法。

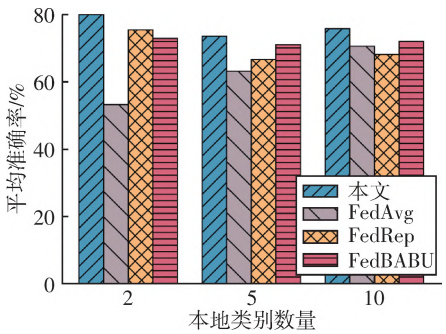

图4 客户端类别数变化时的平均准确率

图5展示了本地类别数量为5类、LR=0.01的设置下,各算法在不同客户端数量下的平均准确率变化。从图中可以看到这些算法都能够很好地适应大规模客户端参与的场景。其中,本文提出的算法始终保持最高的平均准确率,突显了其在处理数据异构性的优势。尤其在客户端数量较多时,本文算法能够充分挖掘不同客户端的通用信息,通过高效的选择性模型聚合策略和聚类操作,进一步提升了模型的性能。

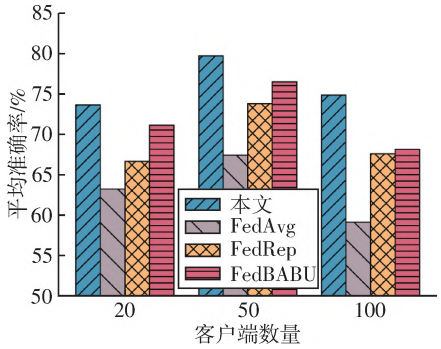

图5　不同客户端数量下的平均准确率

图6展示了N=20个客户端在不同参与率下的平均准确率变化。从图中可以看出，当客户端参与率为0.1时，由于每轮参与聚合的客户端数量较少，全局模型接收到的更新信息较为有限，这在一定程度上影响了模型的收敛速度和最终性能，但本文算法依然取得了72.08%的准确率，显著优于FedAvg和FedRep，展现了其在低通信开销场景下的鲁棒性和高效性。

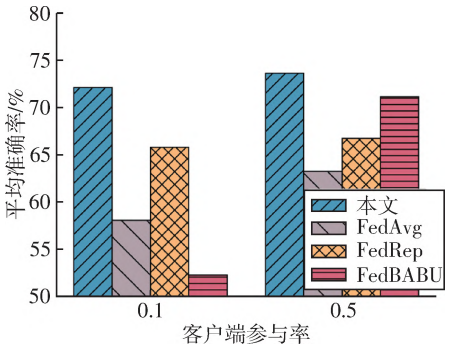

图6　不同客户端参与率下的平均准确率

为了进一步展示复杂场景中本文算法的有效性与优越性，图7展示了N=20个客户端在LR=0.01的设置下，各算法在CIFAR-100数据集下的平均准确率变化。可以观察到，由于CIFAR-100相较于CIFAR-10任务复杂度显著提升，各算法的模型学习效果均有所下降。然而，本文提出的算法在这一复杂场景中仍展现出明显的优势，其平均准确率均高于其他对比算法，充分体现了处理高复杂度数据的鲁棒性和适应性。这是因为本文算法的个性化模型分层设计，能够让客户端在高效学习全局知识的同时适应本地数据分布。此外，选择性模型聚合策略通过优化共享层的特征提取能力，配合聚类策略进一步提升了模型的性能，确保在数据异构和任务复杂度增加的情况下仍能保持卓越的表现。

在联邦学习中，通信开销是影响算法效率的重要因素，因此表1统计了在N=20个客户端，参与率为0.5下，各算法在CIFAR-10任务下达到目标准确率时所需的累计通信次数，以评估不同算法的通信效率。从表1数据可以看到，由于本文采用了个性化模型设计以及选择性通信策略，因此相较于FedAvg算法与FedRep算法，本文算法能够在保证准确率的同时将累计通信轮次减少至少50%，极大降低了通信开销。

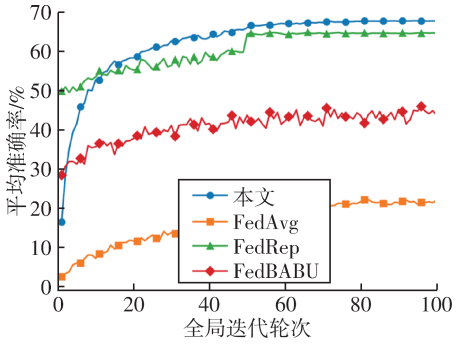

图7　各算法在CIFAR-100任务下的平均准确率

表1　各算法在CIFAR-10任务下达到目标准确率时所需的累计通信次数

算法	目标准确率		
	40%	50%	60%
本文	20	30	60
FedAvg	80	220	660
FedRep		10	290
FedBABU	30	70	230

4　结束语

本文提出了一种面向Non-IID场景的通信高效个性化联邦学习算法。该算法通过模型分层思想实现个性化建模，在此基础上，服务器利用客户端本地数据特征中心，通过K-means聚类加速模型收敛。同时，采用最大均值差异度量本地数据与全局数据分布的相似性，并结合客户端历史参与情况计算优先级，选择优先级较高的客户端进行通信，从而加快全局模型的收敛速度并减少通信开销。实验结果表明，与现有算法相比，本文所提算法能够在保证高准确率的前提下，显著降低通信开销。未来研究将聚焦于聚类算法优化，实现更精细的客户端分组，从而更好地适应复杂的Non-IID场景。

参考文献：

[1] MIAO Y F, CHEN S G. Efficient privacy-preserving federated learning against inference attacks for IoT[C]// IEEE Wireless Communications and Networking Conference (WCNC). 2023: 1-6.

[2] NGUYEN D C, DING M, PATHIRANA P N, et al. Fed-

erated learning for Internet of Things: a comprehensive survey[J]. IEEE Communications Surveys & Tutorials, 2021, 23(3): 1622-1658.

[3] 陈晶, 彭长根, 谭伟杰. 基于联邦学习的多源数据用户画像设计方案[J]. 南京邮电大学学报(自然科学版), 2023, 43(5): 83-91.
CHEN Jing, PENG Changgen, TAN Weijie. An multi-source data user portrait design scheme based on federated learning[J]. Journal of Nanjing University of Posts and Telecommunications (Natural Science Edition), 2023, 43(5): 83-91. (in Chinese)

[4] WU Q, HE K W, CHEN X. Personalized federated learning for intelligent IoT applications: a cloud-edge based framework[J]. IEEE Open Journal of the Computer Society, 2020, 1: 35-44.

[5] 陈飞扬, 周晖, 张一迪. FCAT-FL: 基于Non-IID数据的高效联邦学习算法[J]. 南京邮电大学学报(自然科学版), 2022, 42(3): 90-99.
CHEN Feiyang, ZHOU Hui, ZHANG Yidi. FCAT-FL: an efficient federated learning algorithm based on Non-IID data[J]. Journal of Nanjing University of Posts and Telecommunications (Natural Science Edition), 2022, 42(3): 90-99. (in Chinese)

[6] VAHIDIAN S, MORAFAH M, LIN B. Personalized federated learning by structured and unstructured pruning under data heterogeneity[C]//IEEE 41st International Conference on Distributed Computing Systems Workshops (ICDCSW). 2021: 27-34.

[7] 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.

[8] HANZELY F, RICHTARIK P. Federated learning of a mixture of global and local models[EB/OL]. [2024-03-20]. https://arxiv.org/abs/2002.05516v3.

[9] CHEN Y Q, QIN X, WANG J D, et al. FedHealth: a federated transfer learning framework for wearable healthcare[J]. IEEE Intelligent Systems, 2020, 35(4): 83-93.

[10] WU Q, CHEN X, ZHOU Z, et al. FedHome: cloud-edge based personalized federated learning for in-home health monitoring[J]. IEEE Transactions on Mobile Computing, 2022, 21(8): 2818-2832.

[11] SATTLER F, MULLER K R, SAMEK W. Clustered federated learning: model-agnostic distributed multitask optimization under privacy constraints[J]. IEEE Transactions on Neural Networks and Learning Systems, 2021, 32(8): 3710-3722.

[12] DUAN M M, LIU D, JI X Y, et al. FedGroup: efficient federated learning via decomposed similarity-based clustering[C]//IEEE International Conference on Big Data and Cloud Computing (BdCloud). 2021: 228-237.

[13] PILLUTLA K, MALIK K, MOHAMED A, et al. Federated learning with partial model personalization[C]// Proceedings of International Conference on Machine Learning. 2022: 17716-17758.

[14] ARIVAZHAGAN M G, AGGARWAL V, SINGH A K, et al. Federated learning with personalization layers[EB/OL]. [2024-04-10]. http://arxiv.org/abs/1912.00818.

[15] COLLINS L, HASSANI H, MOKHTARI A, et al. Exploiting shared representations for personalized federated learning[C]// Proceedings of International Conference on Machine Learning. 2021: 2089-2099.

[16] HUANG A, CHEN Y, LIU Y, et al. RPN: a residual pooling network for efficient federated learning[C]// Proceedings of 24th European Conference on Artificial Intelligence. 2020: 1-7.

[17] SHI Y H, LI X, CHEN S G. Toward smart and efficient service systems: computational layered federated learning framework[J]. IEEE Network, 2023, 37(6): 264-271.

[18] JI S X, JIANG W Q, WALID A, et al. Dynamic sampling and selective masking for communication-efficient federated learning[J]. IEEE Intelligent Systems, 2022, 37(2): 27-34.

[19] YAO X, HUANG C F, SUN L F. Two-stream federated learning: reduce the communication costs[C]//IEEE Visual Communications and Image Processing (VCIP). 2018: 1-4.

[20] WU H D, WANG P. Node selection toward faster convergence for federated learning on non-IID data[J]. IEEE Transactions on Network Science and Engineering, 2022, 9(5): 3099-3111.

[21] ROBBINS H, MONRO S. A stochastic approximation method[J]. The Annals of Mathematical Statistics, 1951, 22(3): 400-407.

[22] GRETTON A, BORGWARDT K M, RASCH M J, et al. A kernel two-sample test[J]. Journal of Machine Learning Research, 2012, 13(1): 723-773.

[23] KRIZHEVSKY A, HINTON G. Learning multiple layers of features from tiny images[EB/OL]. [2024-05-10]. https://www.cs.toronto.edu/~kriz/cifar.html.

[24] MCMAHAN B, MOORE E, RAMAGE D, et al. Communication-efficient learning of deep networks from decentralized data[C]//Proceedings of the 20th International Conference on Artificial Intelligence and Statistics. 2017: 1273-1282.

[25] OH J, KIM S, YUN S Y. FedBABU: toward enhanced representation for federated image classification[EB/OL]. [2024-05-20]. https://arxiv.org/abs/2106.06042.

(责任编辑:李小溪)