

面向6G网络的联邦学习高效训练优化方法

苏瑞¹, 朱亚丽², 朱晓荣¹

(1. 南京邮电大学通信与信息工程学院, 江苏 南京 210003; 2. 江苏商贸职业学院, 江苏 南通 226000)

摘要: 第六代移动通信技术(6G)预计将充分利用大量敏感数据并有望实现一个完全互联的世界, 联邦学习作为一种新兴的分布式计算范式广泛应用于各领域。在传统的联邦学习方法中, 收敛前的大量通信轮次导致了巨大的通信成本。此外, 在每一轮训练中, 客户端在上行通信阶段向参数服务器上传大量的模型参数, 增加了通信开销。针对6G网络提出了一种高效的联邦学习训练优化方法(CEFED), 采用定制化的局部训练策略, 通过约束项在更少的通信轮数内实现快速收敛。考虑到上行链路拥塞, 引入了部分客户端参与规则, 以避免多个客户端同时上传参数更新。同时针对数据异质性导致的局部模型之间的差异, 提出了一种基于数据质量的加权聚合策略, 通过为不同本地参数分配合理权重以提升模型精度。实验结果表明, 该算法在多种联邦学习设置下, 均超越了经典的FedAVG算法。与基准算法相比, CEFED算法不仅在提升测试精度的基础上加快了模型的收敛速度, 而且有效节省了通信开销。

关键词: 联邦学习; 数据异质性; 收敛速率; 通信开销

中图分类号: TN929.5

文献标志码: A

doi: 10.11959/j.issn.2096-3750.XXXX.

An efficient training optimization method for federated learning in 6G networks

SU Rui¹, ZHU Yali², ZHU Xiaorong¹

1. School of Communications and Information Engineering, Nanjing University of Posts and Telecommunications, Nanjing 210003, China

2. Jiangsu Vocational College of Business, Nantong 226000, China

Abstract: The sixth generation of mobile communication technology (6G) is expected to make full use of a vast amount of sensitive data and is anticipated to achieve a completely interconnected world. Federated learning, as an emerging distributed computing paradigm, is widely applied in various fields. In traditional federated learning methods, a large number of communication rounds before convergence lead to significant communication costs. Moreover, during each training round, clients upload a large amount of model parameters to the parameter server during the uplink phase, increasing communication overhead. An efficient federated learning training optimization method (CEFED) has been proposed for 6G networks, which adopts a customized local training strategy to achieve rapid convergence within fewer communication rounds through a constraint term. Considering uplink congestion, a partial client participation rule is introduced to prevent multiple clients from uploading parameter updates simultaneously. At the same time, in response to the differences between local models caused by data heterogeneity, a weighted aggregation strategy based on data value is proposed, which assigns reasonable weights to different local parameters to improve model accuracy. Experimental results show that the algorithm outperforms the classic FedAVG algorithm in various federated learning settings. Compared with the benchmark

收稿日期: XXXX-XX-XX; 修回日期: XXXX-XX-XX

通信作者: 朱晓荣, xrzhu@njupt.edu.cn

基金项目: 国家自然科学基金(No.92367102); 南通市科技计划项目(No.MS2023062)

Foundation Items: The National Natural Science Foundation of China(No.92367102), The Science and Technology Planning Project of Nantong(No.MS2023062)

algorithm, the CEFED algorithm not only accelerates the model's convergence rate while improving test accuracy, but also effectively saves on communication costs.

Key words: federated learning, data heterogeneity, convergence rate, communication overhead

0 引言

在全球范围内,随着对第六代移动通信技术(6G, the sixth generation mobile communication technology)的探索不断深入,人工智能(AI, artificial intelligence)技术在日常生活和工业场景中都得到了广泛应用,6G网络预计将充分利用大量敏感数据,实现一个完全互联的世界。这一愿景不仅为人们带来了前所未有的便利,也将促进6G与AI内生智能的融合^[1-3]。与此同时,6G网络的高效率和高可靠性为分布式计算提供了强大的支持,在6G网络中,数据通常存储在海量边缘设备中,数据分散且数据量庞大。传统的集中式数据处理方式不仅会暴露用户隐私,还可能导致数据在传输过程中的安全风险。因此,如何在保护用户隐私的前提下,实现数据的高效处理和利用,成为了6G网络发展中必须面临的挑战^[4]。

联邦学习作为一种新兴分布式机器学习方法,其设计初衷是解决大规模分布式训练中的隐私保护问题。传统的机器学习模型训练通常需要将数据集集中到单个服务器进行处理,这种方式不仅面临数据隐私泄露的风险,还可能导致数据传输和存储的成本高昂。谷歌等科技公司提出的联邦学习框架^[5],通过在客户端本地进行模型训练,并将训练结果(而非原始数据)上传至服务器进行聚合,有效保护了用户数据的隐私安全,并降低了数据传输和存储的负担。随着6G网络愿景的逐渐清晰,其超高速率、超大连接数、超低时延等特性为联邦学习提供了新的应用场景和可能性。研究人员开始关注如何将联邦学习与6G网络相结合,以提升模型的训练效率、降低通信开销,并满足6G网络中对于数据隐私和安全性的严格要求。

联邦学习虽然在通信效率和隐私保护方面具有很大的优势,但传统联邦学习方式仍面临着通信开销太大、收敛速度过慢等主要问题。这些问题源于以下三个方面:

首先,在联邦学习的模型训练方面,网络中各个客户端设备产生或收集到的数据各不相同,这些

数据的统计特征服从非独立同分布(Non-IID, non independent identically distributed),这意味着不同客户端 i, j 之间的局部数据分布 $P_i \neq P_j$ 。这种非独立同分布会降低典型的基于随机梯度下降(SGD, stochastic gradient descent)的算法^[6-7]的训练效率,同时严重影响全局模型的收敛速率^[8]。

其次,传统的联邦学习系统并未设立严格的参与方筛选机制,一般采用随机策略选择客户端参与聚合。然而,由于客户端的异质性、不可控性以及分布式等特征^[9],采用随机策略接收客户端聚合全局模型,可能会导致劣质客户端涌入,从而破坏联邦学习系统的整体性能^[10]。劣质客户端由于数据质量较差,模型参数受到异构数据的影响偏离全局模型的更新方向^[11],这类不可靠的客户端提供的模型参数也可能导致模型性能下降^[12]。

最后,联邦学习中现有的聚合过程也会显著影响整体通信开销。传统的聚合方法,如简单的平均策略^[5],虽然在实现上相对直接,但往往未能充分考虑各个客户端的数据特征。这种简单的聚合策略可能导致全局模型的性能受限,无法有效利用各客户端提供的数据,尤其是在数据分布不均或客户端数量较多的情况下。此外,简单平均也可能忽视某些客户端在训练过程中所贡献的重要信息,进而影响模型的泛化能力。

因此,本文提出了一种面向6G网络的高效联邦学习训练优化方法(CEFED, An efficient federated learning training optimization method),旨在从上述三个方面对联邦学习的训练过程进行优化,实现加速模型收敛的同时降低通信开销。这种方法的主要贡献在于以下三个方面:

1)引入了一种定制化的本地训练策略,通过在本地客户端的目标函数中增加约束项来提高收敛速度,以减少通信总轮数。

2)提出了一种部分客户端选择方法,选择计算效率较高的客户端同时上传其训练结果,并采用 ϵ -greedy策略平衡探索与利用,从而降低每个轮次的通信成本。

3)设计了一种更合理的权值计算方法,在非独

立同分布设置中,通过分析客户端数据量、梯度相似度和数据多样性三个关键因素,构建数据质量评估模型并调整聚合权值,以加速模型收敛。

实验结果表明,与传统的联邦学习算法相比,我们的方法在高数据异质性的情况下,不仅加快了模型的收敛速度,而且在测试精度上取得了显著提高。

本文其余部分组织如下:第1章主要介绍了相关工作。第2章主要介绍系统模型和问题描述。第3章具体描述了基于联邦学习的高效训练优化算法。第4章进行了仿真测试和结果分析。第5章对全文进行了总结。

1 相关工作

在本章中,我们首先介绍了与本文所提方法相关的三类现有研究:已有缓解异质性效应的方案、其他部分客户端选择方法和加权平均的不同变体,然后分析了6G网络中联邦学习的应用。

1.1 缓解异质性效应的方案

联邦学习作为一种新兴的机器学习范式,展现出了解决隐私保护和大规模数据训练需求之间矛盾的巨大潜力。当前,联邦学习的研究重点包括优化聚合过程、减少通信开销以及加强隐私保护^[13]。同时,如何处理不同客户端间的数据差异性,也是联邦学习领域需要克服的重要难题。

FedProx^[14]旨在通过在其优化目标中引入正则项来规范训练过程,防止局部训练过拟合,从而减轻统计异质性。MOON^[15]利用模型表示之间的相似性来纠正个体的局部训练,以提高异构分布下的模型性能。SCAFFOLD^[16]使用控制变量来纠正其本地更新中的“客户端漂移”。PerAvg^[17]引入了元学习的思想,通过对初始共享模型进行微调,使得新用户可以快速地将其应用到自己的数据集上。

然而,上述工作主要依赖于正则化项、控制变量等方法来减轻异质性影响,虽然在一定程度上有效,但它们均未能在减少通信开销和有效处理异质性之间实现良好的平衡。特别是在6G网络的复杂环境下,如何结合客户端选择方法以及加权聚合策略形成一个完整的联邦学习训练优化方法来提高模型训练的效率和性能,仍值得深入研究。

1.2 部分客户端选择方法

在分布式机器学习领域,尤其是联邦学习中,

客户端选择策略对于模型训练的效率和性能至关重要。研究者们提出了多种客户端选择方法,旨在优化训练过程并提高学习效率。

POWER-OF-CHOICE^[18]选择本地损失较高的客户端以提高收敛速度。文献[19]提出一种基于强化学习的客户端选择方法来抵消非独立同分布数据带来的偏差。FedCS^[20]开发了一种贪婪策略,通过最大化参与每一轮通信的客户端数量来加速训练的收敛。文献[21]提出了一种层结构的联邦学习系统TiFL,根据客户端每轮的训练时间将所有客户端划分到不同的层,并在每轮训练中选择同一层内的客户端以缓解资源异构场景内的“落后者”问题。文献[22]提出了Oort框架,结合客户端的模型精度和训练速度提出了一种新的评价客户端训练效用的指标,并根据该指标选出有利于提升训练性能的客户端优先参与训练。

虽然上述方法考虑了资源和数据分布异构性,但缺少对数据量异构性在内的数据特征的综合考虑,导致在选择参与训练的客户端时可能忽略了对全局模型性能有重要影响的客户端,或者过度依赖于特定指标(如本地损失、训练时间或模型精度),从而限制了训练效率和模型精度的进一步提升。

1.3 加权平均的变体

在联邦学习中,模型聚合是实现全局模型更新的关键步骤,它涉及到如何将来自不同客户端的本地模型有效地合并。加权平均聚合方法因其在提高模型性能和处理数据异质性方面的潜力而受到广泛关注。

FedAT^[23]使用了一种离散感知的加权聚合启发式方法来引导和平衡训练,以进一步提高准确性。FedAdp^[24]采用加权平均方法,基于局部梯度和全局梯度之间的夹角来加速收敛,仅适用于纵向联邦学习的情况。文献[25]中,作者采用基于聚类结果的熵作为模型聚合的权重,同时要求所有客户端的参与和熵信息上传到中央服务器,这可能会引发隐私问题。CFL^[26]考虑在每个训练迭代中根据本地梯度的余弦相似度自适应地将参与设备分组,并对每组进行加权模型聚合。

上述研究从一定程度上优化了全局模型的加权聚合,但它们没能从多角度充分利用本地模型以及数据特征评估数据质量,并且需要额外的信息上传,这可能会引发隐私泄露的问题,本文接下来提

出基于数据质量的加权聚合方法，正是针对实际异构场景下这一问题进行的有益探索和尝试。

1.4 6G网络中联邦学习的应用

随着6G网络的快速发展，海量边缘设备的接入为网络智能化提供了丰富的数据资源，联邦学习作为一种新兴的分布式机器学习范式，因其能够在保护数据隐私的同时实现多源数据的联合学习，被视为6G网络中的关键技术之一。近年来，研究人员对6G网络中联邦学习的应用进行了广泛探索。

文献[27]针对6G网络中的能耗挑战，提出了几种基于联邦学习的绿色节能6G网络架构。文献[28]提出了一种面向6G物联网（IoT, Internet of Things）的联邦仿生学习框架，在6G网络中实现了异构IoT资源的有效分配。文献[29]针对时间敏感的6G网络场景，提出了一种基于联邦强化学习的移动边缘计算系统来最小化平均信息年龄。文献[30]对联邦学习如何有效实现6G场景中的参与方个性化定制进行了深入探索。

尽管上述研究表明联邦学习在6G网络中具有

广泛的应用前景，但现有研究仍面临诸多挑战。特别是在实际场景中，联邦学习的训练过程往往受到通信开销、异构数据、计算资源等因素的影响，导致训练效率低下、模型收敛速度慢等问题，因此设计一种面向6G网络的高效联邦学习训练优化方法具有重要的研究价值和实际应用意义。

2 系统模型和问题描述

在本节中，我们首先总结了典型的联邦学习过程，然后介绍了系统模型的构建，最后分析了联邦学习的通信成本。为便于阅读，参数说明总结在表1中。

2.1 经典联邦学习基本步骤

本文聚焦于传统中心化联邦学习，在这种架构下，所有客户端在初始阶段拥有相同的神经网络模型。这些分布式设备用它们的私有数据迭代训练本地模型固定的次数，完整的一轮训练包括以下阶段：

表1 参数说明

符号	说明	符号	说明
k	客户端索引	$\tilde{M}$	通信成本上界
t	训练轮次索引	θ_k^t	第 t 轮训练客户端 k 的本地迭代次数
K	总客户端数	E	本地迭代次数上限
T	总训练轮次	ϕ	约束项的惩罚因子
$n(t)$	参与第 t 轮训练的客户端数量	e_k	本地计算效率
S_t	参与第 t 轮训练的客户端子集	ζ_k	本地训练时间
D_k	客户端 k 的本地数据集	p_k	每个客户端的选中概率
D	总数据集	ε	探索率
X_k	客户端 k 的输入特征集	λ	探索率的衰减系数
Y_k	客户端 k 的标签集	φ_k	客户端 k 的数据量
ω_k^t	第 t 轮训练客户端 k 的模型参数	γ_k	客户端 k 的梯度相似度
ω^0	初始化模型参数	u_k^t	第 t 轮训练客户端 k 的本地模型梯度
ω^t	第 t 轮训练聚合后的全局模型参数	u^t	第 t 轮训练的全局梯度向量
ω^*	模型参数大小	$iter$	累计训练轮次
$f_i(\omega_k)$	数据样本 i 的损失函数	G_k	客户端 k 的格拉姆矩阵
$F_k(\omega_k)$	客户端 k 的局部损失函数	σ_k	客户端 k 的数据多样性
$g_k(\omega_k; \omega^t)$	加入正则化项后客户端 k 的局部损失函数	C_k	客户端 k 的数据质量评估模型
$F(\omega)$	全局模型损失函数	β	迪利克雷分布的参数
M	总通信成本	ψ	病态异构数据类别占比

(1) 下行链路通信阶段：联邦学习过程从下行链路通信阶段开始，该阶段将全局模型参数 ω^t 从参数服务器下载到第 t 轮的 K 个客户端当中。

(2) 本地训练阶段：每个客户端根据收集到的

数据进行模型更新。在传统的联邦学习中，所有客户端的迭代次数都是相同的。即使不同的客户端有不同的训练数据，在训练阶段仍然执行相同的迭代次数。

(3) 上行链路通信阶段：所有客户端通过上行链路信道同时将更新的模型参数传输到参数服务器。这一阶段直接决定了每轮通信开销的大小。下行链路和上行链路通信阶段构成完整的通信回合，以上行链路通信轮为结束。

(4) 聚合阶段：每个客户端将最新的模型参数传输到参数服务器。在传输的消息到达参数服务器后，服务器将接收到的本地模型参数进行加权聚合，得到全局模型并分发回客户端。

总之，每个客户端重复上述一系列步骤，直到全局模型达到收敛，联邦学习过程停止。

2.2 系统模型

在6G网络愿景中，工业物联网（IIoT, Industrial Internet of Things）作为一种典型应用场景，旨在为未来海量IoT设备提供分布式智能服务，并满足日益复杂且多元化的需求。随着工业4.0及后续版本的推进，工业物联网已逐渐成为智能制造、智能电网等领域的基石。在此基础上，结合6G网络的特性，以及联邦学习分布式训练的优势，本文针对6G工业物联网场景进行联邦学习系统模型的构建。该系统模型架构不仅强化了数据隐私保护，还通过高效的模型训练提升了6G工业物联网的智能化水平。具体而言，该架构包含以下几个关键组成部分：

(1) 6G通信网络：作为数据传输与模型更新的基础设施，6G通信网络提供了高速率、低时延、广覆盖的通信能力，确保了联邦学习过程中的数据传输效率与实时性。

(2) 数字孪生技术：通过数字孪生技术将实体工业物联网设备映射为移动端设备、工业机械设备、制造机器设备三个不同类别的IoT虚拟设备，实现了对物理世界的精准感知与实时监测，它们产生的多源异构数据为联邦学习提供了丰富的训练素材，同时更好地模拟了未来6G网络中的工业物联网场景。

(3) 联邦学习框架：在6G通信网络的支持下，联邦学习框架实现了跨设备、跨领域的分布式模型训练。各IoT设备在本地进行数据预处理与模型训练，仅上传模型参数至服务器侧进行聚合，从而在保证数据隐私的同时，提升了模型的泛化能力与训练效率。

如图1系统架构所示，在基于6G工业物联网

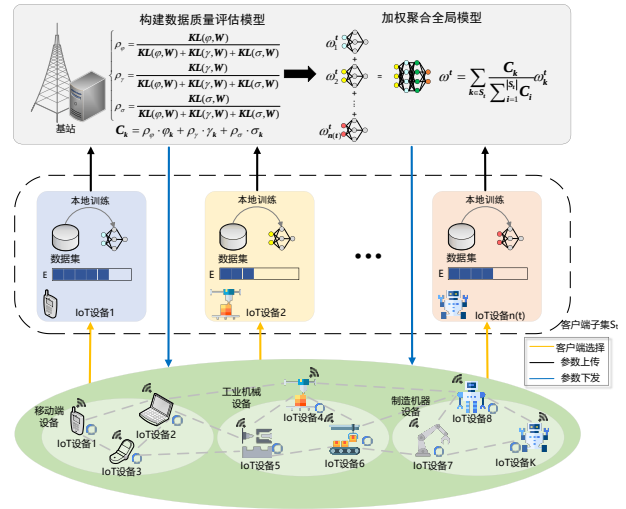


图1 系统模型架构图

的联邦学习场景中，我们考虑单个基站覆盖的多个IoT设备节点（客户端）协作进行联邦训练，假设总共有 K 个客户端，其中 k 是客户端的索引，每个客户端 k 存储一个本地数据集 D_k ，其大小由 $|D_k|$ 表示，其中 $|\cdot|$ 表示集合的大小，总数据量大小为 $|D| = \sum_{k=1}^K |D_k|$ ，本地数据集 D_k 中存在输入特征集 X_k 和对应的标签集 Y_k ，对于属于 D_k 的训练数据样本 i ，它由局部神经网络的输入向量 x_i 和所需的输出标量 y_i 两部分构成。训练的目的是通过损失函数 $f_i(\omega)$ 找到表征输出 y_i 的模型参数 ω 。对于客户端 k 本地数据集上的损失函数定义为：

$$\min_{\omega} F_k(\omega) = \frac{1}{|D_k|} \sum_{i \in D_k} f_i(\omega) \quad (1)$$

在训练开始时，我们通过全局参数 ω^0 初始化局部模型。 ω_k^t 表示联邦学习第 t 轮训练中的第 k 个客户端的模型参数。在第 t 轮训练中，目标是最小化训练损失，如(2)所示：

$$\min_{\omega_k^t} F_k^t(\omega_k^t) = \frac{1}{|D_k|} \sum_{i \in D_k} f_i(\omega_k^t) \quad (2)$$

我们通过确定 ω_k^t 来得到损失函数的最小值。对于联邦学习中每个轮次的类似过程，为了简单起见，省略了训练轮次的索引 t ，如(3)所示：

$$\min_{\omega_k} F_k(\omega_k) = \frac{1}{|D_k|} \sum_{i \in D_k} f_i(\omega_k) \quad (3)$$

联邦学习中的全局损失函数可以表示为 $F(\omega)$ ， $F_k(\omega_k)$ 表示第 k 个客户端的损失函数，通过最小化全局损失函数 $F(\omega)$ 来优化可训练参数 ω 。因此，式(3)中的损失函数可以概括为下式(4)：

$$\min_{\omega} F(\omega) = \sum_{k=1}^K \frac{|D_k|}{|D|} F_k(\omega_k) \quad (4)$$

由式(4)可知, 优化全局损失函数 $F(\omega)$ 等价于最小化局部损失函数 $F_k(\omega_k)$ 的加权平均值。

2.3 联邦学习的总通信成本

本节分析了联邦学习过程中的总通信成本, 并制定了通信成本的上限, 该成本代表了通信效率的关键指标。

联邦学习过程通常包括两个通信阶段, 下行链路通信阶段和上行链路通信阶段。 K 为客户端的总数, $n(t) (n(t) \leq K)$ 为参与第 t 轮训练的客户端数量, T 代表联邦学习过程中的总训练轮次。考虑同质环境, 即本地模型参数的数量是相同的, 表示为 ω^* 。 M 表示联邦学习中的总通信成本, 其具体计算公式为方程式(5):

$$\begin{aligned} M &= \sum_{t=1}^T \{ n(t) + K \} \cdot \omega^* \\ &= \sum_{t=1}^T \{ n(t) \cdot \omega^* + K \cdot \omega^* \} \end{aligned} \quad (5)$$

式中的两项分别对应所有训练轮次中每个参与客户端的下行链路通信 (分发全局模型参数) 成本和上行链路通信 (参数聚合) 成本。当所有 K 个客户端都参与训练阶段和聚合阶段时, 我们能够得到通信开销的上限, 如公式(6)所示:

$$\tilde{M} = 2T(K \cdot \omega^*) \quad (6)$$

根据上述公式(6)中的上限, 我们从以下几点优化通信开销。首先, 考虑到训练过程较长, 我们缩短了本地训练过程以加快收敛速度并减少通信轮次。其次, 考虑到下行链路和上行链路通信信道之间的不对称性, 我们在联邦学习过程中减少了参与设备的数量, 合理选择了 $n(t)$ 。

3 基于联邦学习的高效训练优化算法描述

本章将介绍我们提出的一种基于联邦学习的高效训练优化算法, 采用定制化本地训练策略, 配合部分客户端选择方法, 结合一种更合理的权值计算方法, 降低了通信成本, 提升了模型精度, 实现了通信和模型训练性能的联合优化。

3.1 定制化的本地训练策略

在联邦学习过程中, 梯度多样性量化了分布式计算中损失函数的各个梯度之间的差异。每个客户端 k 的本地模型 ω_k^t 在聚合后都会进行更新, 然而由

于本地数据集的不同分布, 不同客户端的梯度向不同的方向更新, 因此本地模型和全局模型之间存在权值发散现象。如果在第 t 轮训练中执行聚合, 则有 $\omega_k^{t+1} \neq \omega_k^t$, 我们令 $\omega_k^{t+1} = \omega^t$, 其中 ω^t 是 ω_k^t 的加权平均值。

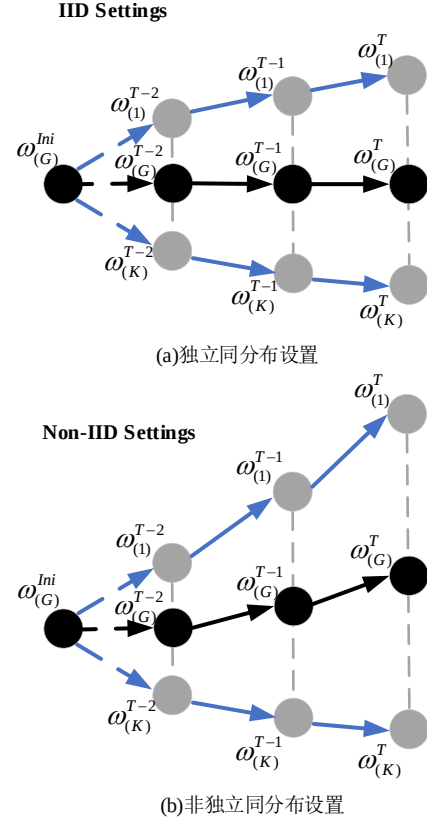


图2 联邦学习中的权值发散现象

如图2所示, 和传统的中心式学习类似, 不同的客户端从相同的初始模型参数 $\omega_{(G)}^{ini}$ 开始, 蓝色曲线表示参与联邦学习的客户端, 随着通信轮次的增加, 客户端使用本地数据在不同方向上更新模型参数。理想的情况是, 联邦学习过程可以更接近传统集中式训练的性能, 但经过 T 轮训练后, 权值发散现象显著。我们假设标签空间中有 C 个标签。权值发散的的根本原因是客户端 k 上的数据分布和整体数据分布之间的差异, 表示为 $\sum_{i=1}^C \|p^{(k)}(y=i) - p(y=i)\|$, 其中 $\|\cdot\|$ 表示 L2 范数。

基于上述讨论, 不同的客户端具有不同的资源条件, 例如网络连接、计算硬件和数据集。如果所有本地模型在其本地数据集上执行相同的迭代周期 E , 则局部损失函数的梯度将出现显著差异。因

此，我们需要为每个客户端制定一个合适的本地训练策略，提供不同的迭代周期数。

本文引入了一个约束项来有效地缩小局部函数之间的梯度多样性。客户端 k 不只是最小化局部损失函数 $F_k(\cdot)$ ，而是用其局部损失函数近似最小化以下目标方程(7)：

$$\min_{\omega} g_k(\omega_k; \omega^t) = F_k(\omega_k) + \frac{\phi}{2} \|\omega_k - \omega^t\|^2 \quad (7)$$

定义1 (θ_k^t 定制本地迭代次数)：对于目标方程(7)，我们提出了一个变量 θ 来表示不同的本地迭代次数， θ_k^t 表示在第 t 次联邦训练中在客户端 k 上迭代 θ_k^t 次。若有 $\|\nabla g_k(\hat{\omega}_k; \omega^t)\| \leq \|\nabla g_k(\omega_k^t; \omega^t)\|$ ，则 $\hat{\omega}_k$ 是经过 θ_k^t ($\theta_k^t \leq E$)次局部训练后满足目标方程(7)的模型参数，其中 $\nabla g(\omega_k; \omega^0) = \nabla F_k(\omega_k) + \phi(\omega_k - \omega^0)$ ， ϕ 为(7)中约束项的惩罚因子，这种定制化的本地训练策略用较少的本地迭代次数实现了更高的准确率。

由于私有数据的潜在异构性，一轮训练中过多的本地迭代次数可能会导致全局模型收敛速度缓慢，我们通过定制化的本地训练策略允许客户端根据需要执行最优的本地迭代次数，这有效地减轻了权值发散的负面影响。

3.2 部分客户端选择方法

在6G网络场景中，客户端与参数服务器的通信在两个方面会影响模型训练效率。首先，考虑到上行通信信道的限制，设备发送的训练输出在参数服务器上排队等待处理，从而显著降低模型的训练速度。其次，一些客户端只有少量的数据和较低的计算效率，这些劣质客户端的模型参数将在聚合阶段成为噪声。

为了解决上述问题，我们提出了一种基于计算效率的客户端选择方法。该方法通过评估客户端的计算能力和数据量，选择一部分计算效率高的客户端参与模型训练和参数上传，并采用 ε -greedy策略平衡客户端选择过程的探索与利用。在每一轮训练中，服务器都会选择一个包含 $n(t)$ 个客户端的子集，记为 S_t 。

定义2 (e_k 本地计算效率)：利用当前轮次的本地训练时间 ζ_k 和本地数据集大小 $|D_k|$ ，计算客户端 k 的本地计算效率 e_k 为：

$$e_k = \frac{|D_k|}{\zeta_k} \quad (8)$$

假设本地数据量 $|D_k|$ 恒定，客户端 k 在第 t 个训练轮次中的本地计算效率表示为 $e_k = \frac{|D_k|}{\zeta_k^t}$ ， $t < T$ 。

由于在初始化阶段没有对所有客户端进行本地训练，采用客户端 k 的本地数据量 $|D_k|$ 来初始化计算效率 e_k 为：

$$e_k = \begin{cases} |D_k|, t = 0 \\ |D_k|/\zeta_k^t, t \neq 0 \end{cases}, k \in \{1, \dots, K\} \quad (9)$$

在一般情况下，由于客户端的初始化计算效率远高于其真实的计算效率，即 $|D_k| \gg e_k = |D_k|/\zeta_k^t$ 。因此，在迭代开始时，未能完成训练的客户端更有可能被选中，从而快速获得真正的本地计算效率。

基于本地客户端上传的信息，我们用 p_k 表示客户端被选择的概率，参数服务器计算每个客户端的选中概率 p_k 如下所示：

$$p_k = \frac{e_k}{\sum_{k=1}^K e_k} \quad (10)$$

在每一轮训练中，所选客户端的子集是不同的，这使得全局模型考虑尽可能多的客户端信息，因此，式(4)和式(7)变为式(11)：

$$\min_{\omega} F(\omega) = \sum_{k=1}^K p_k g_k(\omega_k; \omega^t), \sum_{k=1}^K p_k = 1 \quad (11)$$

考虑到在联邦学习系统中，客户端的计算能力差异可能导致算力资源丰富的客户端被频繁选中，而资源较少的客户端则可能在多轮迭代后被边缘化，这种现象被称为“饥饿者”问题，我们采用一种 ε -greedy策略，通过平衡探索与利用来优化客户端的选择过程。

具体而言，对于客户端子集 $n(t)$ 中的每个客户端，我们以探索率 ε 进行随机选择，促进选择的多样性并减少对特定客户端的依赖。同时，我们以 $1 - \varepsilon$ 的概率根据客户端的计算效率进行选择，这有助于提升整体训练的效率。

我们通过指数衰减函数来实现动态 ε 值的调整，在训练初期鼓励探索，随着训练的深入增加高效客户端的参与，表示为：

$$\varepsilon^t = \varepsilon^0 \cdot e^{-\lambda(t-1)} \quad (12)$$

其中 ε^0 表示探索率 ε 的初始化值， ε^t 表示当前轮次的探索率， λ 是用于控制 ε 值减小的衰减系数。

部分客户端选择方法有效减少了每轮训练中参与的客户端数量，充分考虑了不同客户端的计算效

率和训练过程中可能出现的“饥饿者”问题。这种方法通过限制每轮上传参数更新的客户端数量，显著降低了因通信而产生的延迟和带宽消耗，大大节省了联邦学习每轮的通信成本。

3.3 权值计算方法

在联邦学习环境中，客户端之间的数据分布差异是非独立同分布特性的直接体现，这对全局模型的准确度和收敛性构成了显著挑战^[31-32]。因此，开发一种能够有效应对数据异质性，合理聚合本地模型的策略，对于提高联邦学习方法的准确度和收敛性有重要价值。

本节我们提出了一种基于数据质量的加权聚合方法。该方法通过评估数据量、梯度相似度和数据多样性等关键指标，构建了一个数据质量评估模型，并利用KL散度量化不同数据特征之间的差异，为本地模型分配了合理的聚合权重。

定义3 (ϕ_k 数据量) 我们将客户端 k 的数据量定义为客户端数据与总体数据之比的归一化值，如方程式(13)所示：

$$\phi_k = \text{Normalize}\left(\frac{|D_k|}{|D|}\right) \quad (13)$$

定义4 (γ_k 梯度相似度) 在联邦学习的场景中，客户端 k 的数据对全局模型训练的贡献也可以通过梯度相似度进行描述。梯度相似度通过比较本地模型梯度和全局模型梯度来衡量客户端 k 的数据质量^[33]。联邦学习的拟合过程所采用的梯度在一定程度上反映了优化的方向，因此，如果客户端 k 的数据产生的训练梯度与假定的最优梯度方向越接近，那么认为其数据质量越高。所以我们考虑利用梯度相似度来缓解客户端漂移现象，通过聚合多轮次的余弦相似度并归一化来构建客户端 k 的梯度相似度：

$$\begin{aligned} \gamma_k &= \text{Normalize}\left(\sum_{t=0}^{\text{iter}} \cos(u_k^t, u^t) / \text{iter}\right) \\ &= \text{Normalize}\left(\sum_{t=0}^{\text{iter}} \frac{\langle u_k^t, u^t \rangle}{\|u_k^t\| \|u^t\|} / \text{iter}\right) \end{aligned} \quad (14)$$

其中， u_k^t 表示第 t 轮训练的客户端本地模型梯度向量， u^t 表示第 t 轮训练的全局梯度向量， iter 表示累计训练轮次。

定义5 (σ_k 数据多样性) 当联邦学习中提供的数据充分可靠时，我们发现数据分布的多样性能够在一定程度上反映数据的质量。每个数据样本都包含多个特征维度，客户端 k 的格拉姆矩阵 G_k 可以计

算数据样本的特征表示向量 $x_i \in X_k$ 在欧式空间内构建的平行多面体体积，从之前的研究得知，当所构建的平行多面体的体积越大时，数据的特征分布越发散，对应更高的数据质量^[34]，因此我们使用格拉姆行列式衡量客户端本地数据的多样性，并对数据多样性进行归一化，如下式(15)所示：

$$\begin{aligned} \sigma_k &= \text{Normalize}\left(\sqrt{|G_k|}\right) \\ &= \text{Normalize}\left(\sqrt{|X_k^T X_k|}\right) \end{aligned} \quad (15)$$

考虑KL散度能够衡量两种分布之间的差异程度，我们将通过KL散度构建一个包含数据量、梯度相似度和数据多样性的数据质量评估模型，可以评估上述指标的影响^[35]。

设 $\phi = \{\phi_1, \phi_2, \dots, \phi_K\}$ ， $\gamma = \{\gamma_1, \gamma_2, \dots, \gamma_K\}$ 和 $\sigma = \{\sigma_1, \sigma_2, \dots, \sigma_K\}$ 分别为数据量、梯度相似度和数据多样性的权重参数集，概率分布 W 表示当三个指标无法区分时的一组权重，我们基于KL散度计算权重 ρ_ϕ 、 ρ_γ 和 ρ_σ 来衡量这三个指标的重要程度：

$$\rho_\phi = \frac{KL(\phi, W)}{KL(\phi, W) + KL(\gamma, W) + KL(\sigma, W)} \quad (16)$$

$$\rho_\gamma = \frac{KL(\gamma, W)}{KL(\phi, W) + KL(\gamma, W) + KL(\sigma, W)} \quad (17)$$

$$\rho_\sigma = \frac{KL(\sigma, W)}{KL(\phi, W) + KL(\gamma, W) + KL(\sigma, W)} \quad (18)$$

定义6 (C_k 数据质量评估模型) 客户端 k 的数据质量评估模型 C_k 如方程式(19)所示：

$$C_k = \rho_\phi \cdot \phi_k + \rho_\gamma \cdot \gamma_k + \rho_\sigma \cdot \sigma_k \quad (19)$$

服务器通过采集选中客户端的本地模型来构建客户端数据的质量评估模型 C_k ，并将其归一化作为全局模型加权聚合的权重。第 t 轮的全局模型可以通过以下方程式(20)获得：

$$\omega^t = \sum_{k \in S_t} \frac{C_k}{\sum_{i=1}^{|S_t|} C_i} \omega_k^t \quad (20)$$

本文提出的权值计算方法通过综合考量数据量的权重、模型梯度的相似度和数据分布的多样性这三个核心因素，提升了数据质量较高的客户端的权重，优化了全局模型的聚合过程。

我们将本文算法的整个过程总结为算法1：

客户端选中概率更新算法用于更新联邦训练中不同客户端被选中参与训练的概率，算法2使用客户端的本地数据量来初始化它们的局部计算效率，本地训练结束后参与训练的客户端的局部计算效率会得到更新并用于重构客户端选中概率。具体细节

算法1 基于联邦学习的高效训练优化算法(CEFED)

输入: K 个客户端的索引 k , 总训练轮次 T , 约束项的惩罚因子 ϕ , 探索率初始值 ϵ^0 , 衰减因子 λ , 客户端上的本地数据集 $D_k, k = 1, \dots, K$, 第 t 轮训练中所选择的客户端数量 $n(t)$, 并构成一个子集 S_t , 本地迭代次数 θ_k^t , 模型的初始化参数 ω^0

输出: 更新后的全局模型参数 ω'^{+1}

- 1) 使用 ω^0 来初始化所有本地模型参数
- 2) **for** $t \leftarrow 0$ **to** $T - 1$ **do**
- 3) 下行通信阶段:
服务器根据式(12)更新探索率 ϵ^t
服务器基于 ϵ - greedy 策略决策随机选择客户端或以概率 p_k 选择客户端构建客户端子集 S_t .
服务器将模型参数 ω^t 发送给子集 S_t 中的客户端。
- 4) 本地训练阶段:
- 5) **for** 每个客户端 $k \in S_t$ **in parallel do**
- 6) 每个客户端用方程(11)计算 θ_k^t 迭代周期, 通过约束项最小化全局损失函数。
- 7) 每个客户端根据式(14)计算当前轮次本地模型的梯度相似度
- 8) **end for**
- 9) 上行通信阶段:
考虑计算效率的评估, 每个客户端将当前轮次的训练时间 ζ_k^t 上传至服务器端。
考虑灵活的聚合策略, 集合 S_t 中的每个客户端将模型参数 $\omega_k'^{+1}$ 、数据量 ϕ_k 、梯度相似度 γ_k 、数据多样性 σ_k 上传至服务器端。
- 10) 模型聚合阶段:
服务器通过**算法2**更新客户端选择概率 p_k
服务器通过式(19)计算数据质量评估模型 C_k
服务器通过式(20)加权聚合本地模型更新全局模型 ω'^{+1}
- 11) **end for**

在算法2中详细描述:

算法2 客户端选中概率更新算法

输入: 当前训练轮次 t , 本地客户端数据集大小 $|D_k|, k = 1, \dots, K$, 第 t 轮训练中参与训练的客户端子集 S_t 及其本地训练时间 ζ_k^t

输出: 每个客户端的选中概率 p_k

- 1) **if** $t = 0$ **then**
- 2) **for** $k \leftarrow 1$ **to** K **do**
- 3) $e_k = |D_k|$
- 4) **end for**
- 5) **else**
- 6) **for** k **in** S_t **do**
- 7) $e_k = |D_k| / \zeta_k^t$
- 8) **end for**
- 9) 通过式(10)更新客户端选中概率 p_k
- 10) **end if**
- 11) **return** 客户端选中概率 p_k

4 仿真结果与分析

4.1 数据集与仿真设置

本文基于 PyTorch 实现了 CEFED 算法, 并通过一系列模拟实验来评估所提出的算法和策略的有效性。

实验选用了两个广泛使用的公开数据集进行评估: FashionMNIST 和 CIFAR-10。Fashion-MNIST 包含 60,000 张训练图和 10,000 张测试图, 其具有 10 个不同的类别, 这些类别覆盖了各种服装, 比如 T 恤、裤子和鞋子等, 每张都是 28×28 像素的灰度图。CIFAR-10 数据集包含 10 个类别的 60000 张 32×32 的彩色图像, 这些类别分别是飞机、汽车、鸟类、猫、鹿、狗、青蛙、马、船和卡车, 其中训练集包含 50,000 张图像, 测试集包含 10,000 张图像, 是更具有挑战性的基准数据集。

本文采用 CNN 模型^[36]来训练数据集。该模型包含两个卷积层, 分别拥有 32 和 64 个通道。在每个卷积层后, 通过 2×2 的池化层来提取图像特征。此外, 模型还包含一个 512 单元的全连接层和一个线性层, 后者用于输入图像最终的分类判断。

本实验设置了 1 个中心服务器和 20 个本地客户端, 客户端通过随机梯度下降方法^[37]进行本地训练, 采用交叉熵损失函数以更好地适应分类问题, 实验设置本地迭代周期的上限固定为 $E = 20$, 总训练轮次为 100, 每一轮选取的本地客户端数量为 4, 学习率设置为 0.005, 为了评估本文所提出算法的准确度以及对通信效率的提升, 将本文所提出的方法与 FedAvg^[5]、FedProx^[14]、SCAFFOLD^[16]、Per-Avg^[17] 算法进行对比, 以突出该方法的优势。

(1) 实际 Non-IID

在联邦学习场景中, 数据的非独立同分布特性是一个常见的挑战, 考虑采用迪利克雷分布 $\text{Dir}(\beta)$ 来模拟和分析现实中 Non-IID 数据分布的情况。通过调整迪利克雷分布的参数 β , 我们可以生成不同异质性程度的数据, 参数 β 越小, 对应的数据异质性程度越高。

图3中展示了 $\beta = 0.1$ 时的数据分布情况。每个客户端的点的颜色和大小代表了其数据集中各类别的相对比例和分布。可以观察到, 不同客户端对不同类别的偏好程度不一, 例如客户端 6 主要集中于类别 0 的数据, 而客户端 18 中包含大量类别 1 的数

据。这种分布的不均衡性是实际 Non-IID 特性的直观体现。

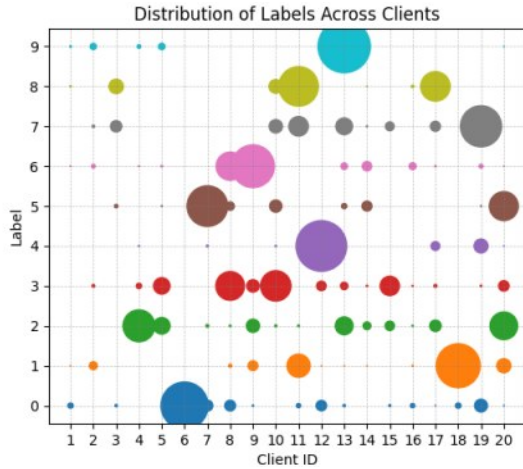


图3 实际 Non-IID 数据分布图

(2) 病态 Non-IID

考虑固定分类数量可能会限制模型精确识别不同客户端数据分布的能力，我们采用 $\text{Pat}(\psi)$ 来模拟病态 Non-IID 分布，其中 ψ 代表病态异构数据的类别占比。 ψ 等于 20% 意味着每个客户端将从整个数据集的所有类别中随机选取 20% 的类别进行训练。病态 Non-IID 分布能够限制模型在这些客户端上学习到的类别多样性和泛化能力。

图 4 展示了 $\text{Pat}(20\%)$ 时我们模拟的客户端数据的分布状态，例如客户端 1 中只包含类别 3 和类别 5 的数据，同时客户端 12 和客户端 16 的数据都只包含类别 1 和类别 2，通过这种方式能够模拟数据分布的极端情况。

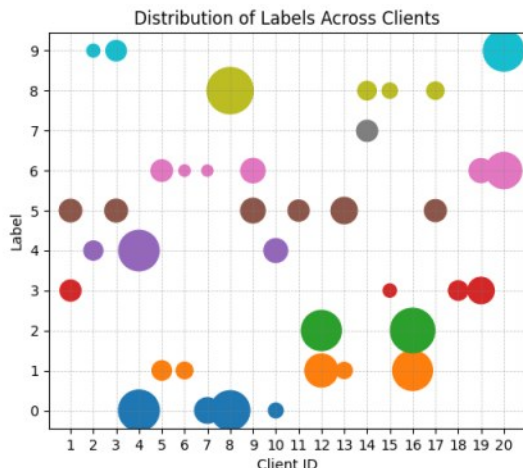


图4 病态 Non-IID 数据分布图

4.2 超参数的影响

为了更好地理解定制化的本地训练策略是如何在 6G 网络中工作的，我们探讨了统计异质性对收敛速度的影响。

(1) 迪利克雷分布的浓度参数 β

首先，我们演示了统计异质性如何影响收敛性，通过利用不同迪利克雷分布浓度参数 β 生成数据集样本进行训练来比较算法在不同异质性情况下的性能，迪利克雷参数 β 的值越小，说明生成样本标签的统计异构性越高。

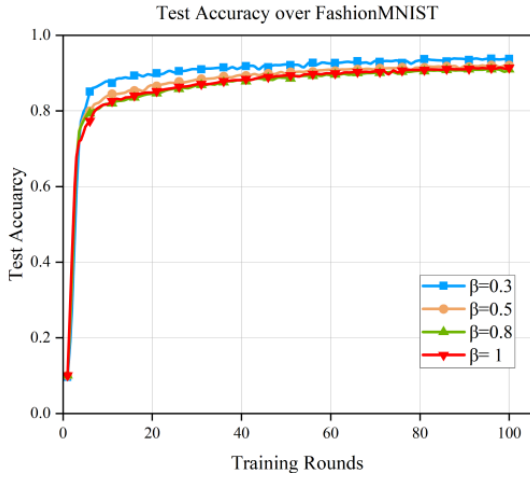
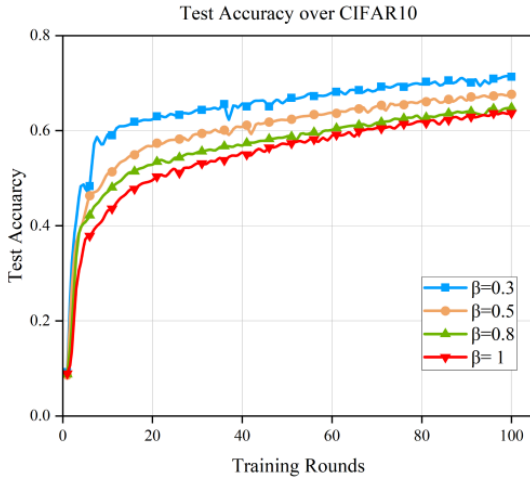
通过图 5 仿真结果，我们可以观察到随着迪利克雷分布浓度参数 β 的降低，即数据异质性增加，CEFED 算法的准确度表现出显著的提升，并且收敛速度加快。这一现象表明，CEFED 算法对于处理非独立同分布数据具有显著的优势，有效解决了高度异构的数据分布产生的权值发散现象。

(2) 定制化本地训练策略的惩罚因子 ϕ

客户端上的本地计算量和定制化本地训练策略的惩罚因子 ϕ 是影响 CEFED 性能的关键参数。一般来说，选择更多的本地计算可以减少总训练轮次 T ，从而降低通信成本，但较大的本地迭代次数 E 可能会导致本地模型偏离初始起点太远，从而导致潜在的权值发散现象。因此，在第 3.1 节中提出了一个减少本地模型与全局模型之间距离的约束项。我们通过为每个客户端设置相同数量的本地迭代总次数来消除系统异构性的影响，在本节中将本地迭代次数的上限固定为 $E = 20$ 。对于不同数据集， ϕ 分别设置为 0、0.3、0.8 和 1，以表示不同的惩罚程度。

如式(7)所示，约束项测量的是每一轮训练后本地模型和全局模型之间的距离，其会惩罚过于偏离全局模型的本地模型，系数 ϕ 表示整个损失函数 $g_k(\omega_k; \omega')$ 中的惩罚权重，增加统计异质性会导致更差的收敛性，但约束项缓解了这个问题。图 6 中实验结果表明，当惩罚系数权重越大时，即 ϕ 值越大时，在不同情况下具有较好的性能。因此，根据上述结果，在后续实验中将 ϕ 设为 1。实验表明，CEFED 算法中引入的定制化局部训练策略适用于具有不同统计异质性的场景。因此，本文方法通过定制化的本地训练策略减少了总训练轮数 T ，加速了模型收敛。

(3) 不同数据指标的重要程度

(a) FashionMNIST 数据集上不同 β 值的测试准确率曲线(b) CIFAR10 数据集上不同 β 值的测试准确率曲线图5 不同迪利克雷分布参数 β 的影响

实验分别在非独立同分布和独立同分布数据环境下, 评估客户端数据量、梯度相似度和数据多样性作为训练过程中个体模型聚合权重的重要性, 通过比较单索引模型训练准确度来评估每个指标的重要程度。

非独立同分布数据环境下, 从表2中可以看出数据多样性是影响模型训练准确度的最关键因素。在所有测试数据集上, 与梯度相似度指标相比, 数据多样性指标的准确率提高了约两个百分点。同样, 与数据量指标相比, 它增加了约三到四个百分点。

独立同分布数据环境下, 从表2中可以得出结论, 梯度相似度的重要程度最高, 数据量被认为是仅次于梯度相似度的第二大关键因素。当单独作为聚合权重使用时, 它具有比数据多样性更明显的优

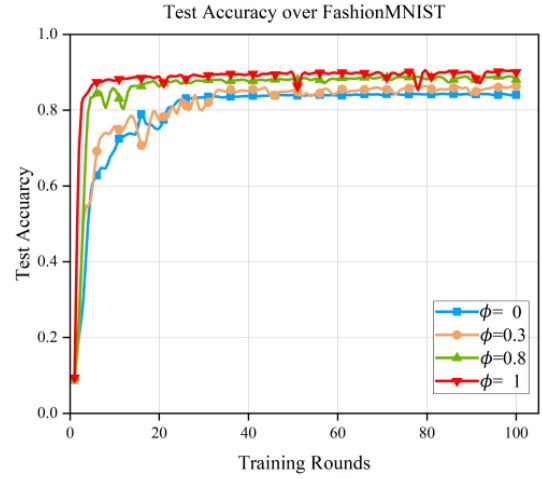
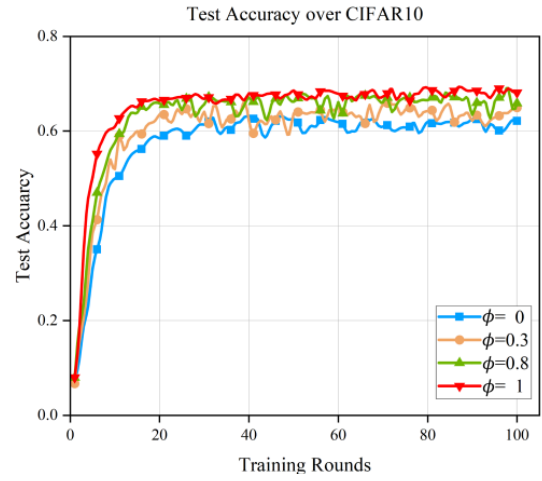
(a) FashionMNIST 数据集上不同 ϕ 值的测试准确率曲线(b) CIFAR10 数据集上不同 ϕ 值的测试准确率曲线图6 不同约束项惩罚因子 ϕ 的影响

表2 不同数据分布设置下, 单索引 CEFED 算法的准确率(%)

数据指标	非独立同分布(Non-IID)		独立同分布(IID)	
	FashionMNIST	CIFAR10 $\times 10^{-3}$	FashionMNIST	CIFAR10 $\times 10^{-3}$
数据多样性 σ	89.7	69.3	90.6	62.1
梯度相似度 γ	87.9	67.2	95.2	71.5
数据量 ϕ	85.4	66.4	93.3	70.1

势。此时数据多样性的性能在两个数据集上都是最差的, 这是因为在每个客户端的数据都是独立同分布的情况下, 不同客户端的数据多样性基本一致, 难以从中获取有效信息。

4.3 与相关联邦学习算法的比较

在本节中, 将所提出的 CEFED 算法与几种相关的联邦学习算法进行比较, 包括 FedAvg、Fed-

Prox、SCAFFOLD 和 PerAvg, 以突出该算法的优势。

● **FedAvg**: 首个提出的联邦学习算法, 每轮训练中随机选择客户端参与训练, 将数据量的归一化值作为权重进行模型聚合。

● **FedProx**: 通过在其优化目标中引入正则项来规范训练过程, 防止局部训练过度拟合, 从而减轻统计异质性。

● **SCAFFOLD**: 通过为客户端和服务端设置额外控制变量来减少客户端之间的数据异质性, 从而提高模型的准确性和鲁棒性。

● **PerAvg**: 引入了元学习的思想, 通过对初始共享模型进行微调, 使得新用户可以快速地将其应用到自己的数据集上。

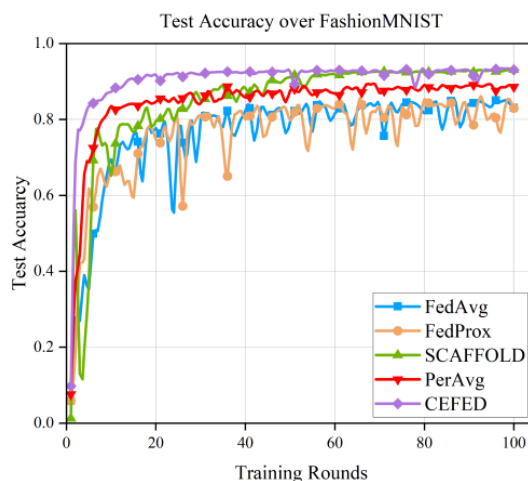
(1) 实际 Non-IID 设置下的性能对比

我们将本文提出 CEFED 算法和不同联邦学习基线算法在实际非独立同分布设置下训练 100 轮次后进行测试准确度对比。从图 7 中可以观察到在实际非独立同分布设置下, 相比于传统的联邦平均算法(FedAvg), CEFED 算法在 CIFAR-10 数据集上的精度提升了约十二个百分点, 在 Fashion-MNIST 数据集上也取得了一定的优势, 这一结果证实了本文提出算法能够更有效地处理异构数据。

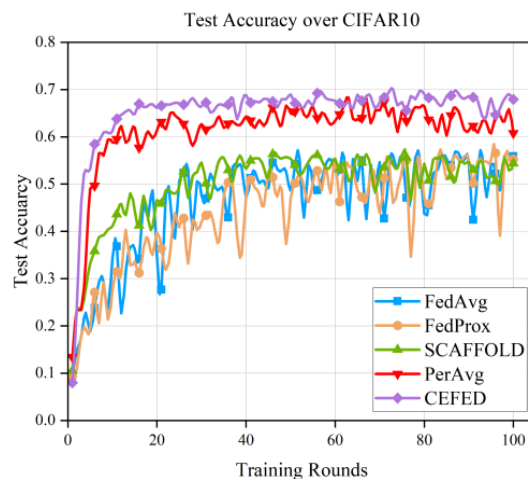
同时, 在与 FedProx、SCAFFOLD、PerAvg 等经典个性化联邦学习算法竞争中, CEFED 算法也取得了显著提升。例如, 在 Fashion-MNIST 数据集中, CEFED 成功将模型精度提升至 90% 左右, 显著提升了全局模型的性能。而对于更复杂的 CIFAR-10 数据集, CEFED 相比个性化方法的最大精度提升超过了十三个百分点, 同时收敛速度也有较大的提升, 这是因为 CEFED 算法中基于数据质量的加权聚合方法能够深入分析客户端数据以及模型参数之间的内在联系, 从而更好地克服异构数据产生的客户端漂移现象, 同时部分客户端选择方法会选择计算效率更高的客户端参与训练, 加速模型收敛。

(2) 病态 Non-IID 设置下的性能对比

虽然利用迪利克雷分布生成的非独立同分布场景能够让模型采集到更广泛的特征和内在联系, 但缺乏对分类数固定时的极端场景的性能比较, 于是我们考虑在 $\psi = 20\%$ 时对不同联邦学习算法在不同数据集上训练 100 轮次的表现进行对比分析。



(a) FashionMNIST 数据集上不同算法的测试准确率曲线



(b) CIFAR10 数据集上不同算法的测试准确率曲线

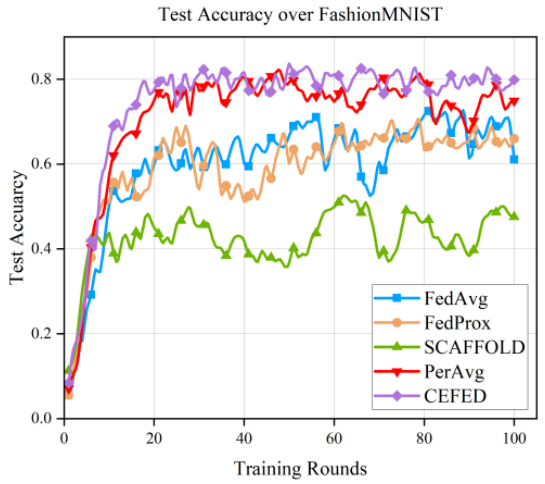
图 7 实际 Non-IID 设置下不同算法的性能比较

根据图 8 中所示的对比结果, 我们可以发现 CEFED 算法相较于其他基准算法能够实现更快的收敛速度和更高的准确度。在 FashionMNIST 数据集的应用场景下, CEFED 算法相较于联邦平均算法实现了超过十个百分点的精度显著提升, 同时对于表现出色的 PerAvg 算法, CEFED 算法也能小幅提升测试准确度。

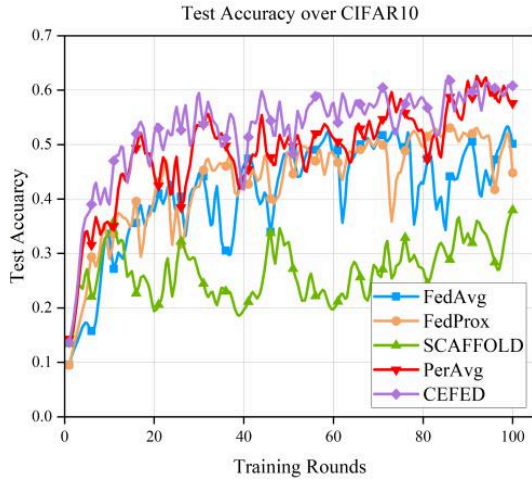
不过, 在对比不同个性化联邦学习方法的实验中, CEFED 算法在 CIFAR-10 数据集上的表现与 PerAvg 算法相比, 其精度提升并不明显, 提升幅度未及五个百分点, 同时波动幅度较大, 这可能与实验中所处的病态 Non-IID 环境有关, 其中各个客户端的数据类别极其有限, 阻碍了模型学习到数据的典型特征与联系。

(3) CEFED 算法的通信开销对比

图 9 展示了不同算法在两种数据集上训练 100



(a) FashionMNIST数据集上不同算法的测试准确率曲线



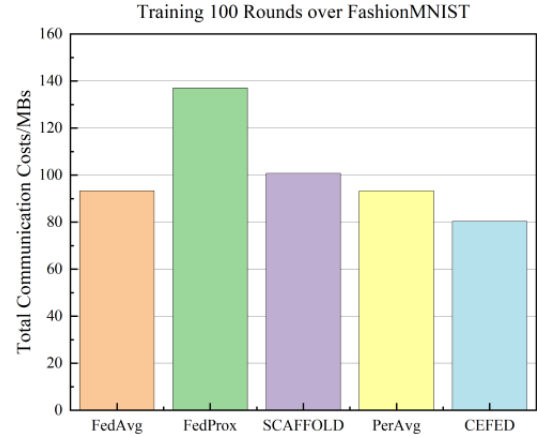
(b) CIFAR10数据集上不同算法的测试准确率曲线

图8 病态 Non-IID 设置下不同算法的性能比较

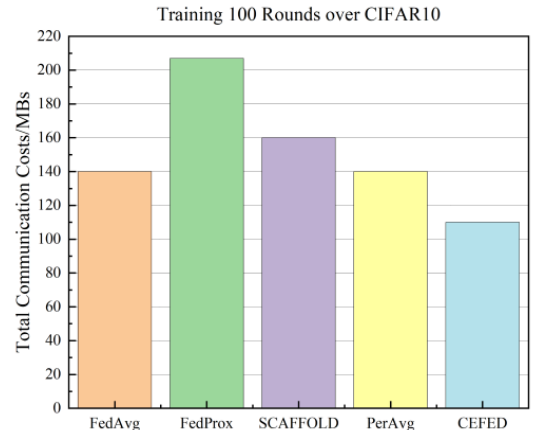
轮次的通信开销对比, 本文提出的 CEFED 算法在减少通信开销方面有显著的优势。该算法通过定制化的本地训练策略, 在保持模型精度的同时显著减少了所需的通信成本; 通过选择计算效率高的客户端, 有效降低了参与训练的客户端数量, 从而减少了通信开销; 同时基于数据质量的加权聚合策略为不同本地模型分配了合理的权重, 不仅提升了模型精度, 也减少了需要通信来修正模型的次数。

相比之下, 传统的联邦平均算法在每一轮随机选择客户端参与训练, 需要更多的通信轮次来达到相同的精度。FedProx 虽然通过引入正则化项来考虑客户端更新的多样性, 但如果客户端选择和聚合策略不够完善, 仍可能导致较高的通信成本。SCAFFOLD 算法通过为客户端和服务端引入额外控制变量来减少数据异质性的影响, 但由于客户端参与度不加以控制, 通信开销依然较高。PerAvg 算

法引入了元学习的思想, 但通信开销仍会因为客户端数量大或数据分布极不均匀增加。综上所述, CEFED 算法在不同数据集上的实验结果证明了其在减少通信开销方面的卓越性能。



(a) FashionMNIST数据集上不同算法的通信开销对比



(b) CIFAR10数据集上不同算法的通信开销对比

图9 不同算法训练100轮次的通信开销对比

5 结束语

本文针对6G网络环境下的联邦学习挑战, 提出了一种高效的训练优化方法——CEFED 算法。该方法通过实施定制化的局部训练策略, 有效减少了通信轮次, 并加快了模型的收敛速度。同时, 引入了基于计算效率的部分客户端选择方法, 通过 ϵ -greedy 策略平衡探索与利用, 显著降低了每轮的通信成本。此外, 我们还设计了一种基于数据质量的加权聚合方法, 进一步提升了全局模型的准确度和泛化能力。通过一系列的实验验证, CEFED 算法在处理不同异质性水平的数据集时, 都展现出了优异的训练性能和通信效率, 为未来6G网络中

的联邦学习提供了一个高效、可靠的解决方案。

参考文献:

- [1] 耿光磊, 高博, 熊轲, 等. 联邦学习赋能 6G 网络综述[J]. 物联网学报, 2023, 7(2): 50-66.
GENG G L, GAO B, XIONG K, et al. A survey of federated learning for 6G networks [J]. Chinese Journal on Internet of Things, 2023, 7(2): 50-66.
- [2] 李文璟, 喻鹏, 张平. 6G 智能内生网络架构及关键技术分析[J]. 中兴通讯技术, 2023, 29(05): 2-8.
LI W J, YU P, ZHANG P. Analysis of the architecture and key technologies of 6G intelligent endogenous networks[J]. ZTE Technology Journal, 2023, 29(05): 2-8.
- [3] ZHOU F Q, LI W J, YANG Y, et al. Intelligence-endogenous networks: Innovative network paradigm for 6G[J]. IEEE Wireless Communications, 2022, 29(1): 40-47.
- [4] 栾宁, 熊轲, 张煜, 等. 6G: 典型应用, 关键技术与面临挑战[J]. 物联网学报, 2022, 6(1): 29-43.
LUAN N, XIONG K, ZHANG Y, et al. 6G: typical applications, key technologies and challenges[J]. Chinese Journal on Internet of Things, 2022, 6(1): 29-43.
- [5] MCMAHAN H B, MOORE E, RAMAGE D, et al. Communication-efficient learning of deep networks from decentralized data[C]//Artificial intelligence and statistics. PMLR, 2017: 1273-1282.
- [6] GHADIMI S, LAN G H. Stochastic first-and zeroth-order methods for nonconvex stochastic programming[J]. SIAM journal on optimization, 2013, 23(4): 2341-2368.
- [7] PENG Y Q, HAO Z Y, YUN X C. Lock-free parallelization for variance-reduced stochastic gradient descent on streaming data[J]. IEEE Transactions on Parallel and Distributed Systems, 2020, 31(9): 2220-2231.
- [8] ZHANG J, GUO S, QU Z H, et al. Adaptive federated learning on non-iid data with resource constraint[J]. IEEE Transactions on Computers, 2021, 71(7): 1655-1667.
- [9] 王勇, 李国良, 李开宇. 联邦学习贡献评估综述. 软件学报, 2023, 34(3): 1168-1192.
WANG Y, LI G L, LI K Y. Survey on contribution evaluation for federated learning[J]. Journal of Software, 2023, 34(3): 1168-1192.
- [10] 王波, 代晓蕊, 王伟, 等. 面向联邦学习的对抗样本投毒攻击[J]. 中国科学: 信息科学, 2023, 53(3): 470-484.
WANG B, DAI X R, WANG W, et al. Adversarial examples for poisoning attacks against federated learning[J]. Science in China (Information Sciences), 2023, 53(3): 470-484.
- [11] SOLTANI B, HAGHIGHI V, MAHMOOD A, et al. A survey on participant selection for federated learning in mobile networks[C]//Proceedings of the 17th ACM Workshop on Mobility in the Evolving Internet Architecture. 2022: 19-24.
- [12] GUPTA A, LUO T, NGO M V, et al. Long-short history of gradients is all you need: Detecting malicious and unreliable clients in federated learning[C]//European Symposium on Research in Computer Security. Cham: Springer Nature Switzerland, 2022: 445-465.
- [13] WU J J, DONG F, LEUNG H, et al. Topology-aware federated learning in edge computing: A comprehensive survey[J]. ACM Computing Surveys, 2024, 56(10): 1-41.
- [14] LI T, SAHU A K, ZAHEER M, et al. Federated optimization in heterogeneous networks[J]. Proceedings of Machine learning and systems, 2020, 2: 429-450.
- [15] LI Q B, HE B S, SONG D. Model-contrastive federated learning [C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2021: 10713-10722.
- [16] KARIMIREDDY S P, KALE S, MOHRI M, et al. Scaffold: Stochastic controlled averaging for federated learning[C]//International conference on machine learning. PMLR, 2020: 5132-5143.
- [17] FALLAH A, MOKHTARI A, OZDAGLAR A. Personalized federated learning with theoretical guarantees: A model-agnostic meta-learning approach[J]. Advances in neural information processing systems, 2020, 33: 3557-3568.
- [18] CHO Y J, WANG J Y, JOSHI G. Towards understanding biased client selection in federated learning[C]//International Conference on Artificial Intelligence and Statistics. PMLR, 2022: 10351-10375.
- [19] WANG H, KAPLAN Z, NIU D, et al. Optimizing federated learning on non-iid data with reinforcement learning[C]//IEEE INFOCOM 2020-IEEE conference on computer communications. IEEE, 2020: 1698-1707.
- [20] NISHIO T, YONETANI R. Client selection for federated learning with heterogeneous resources in mobile edge[C]//ICC 2019-2019 IEEE international conference on communications (ICC). IEEE, 2019: 1-7.
- [21] CHAI Z, ALI A, ZAWAD S, et al. Tiff: A tier-based federated learning system[C]//Proceedings of the 29th international symposium on high-performance parallel and distributed computing. 2020: 125-136.
- [22] LAI F, ZHU X F, MADHYASTHA H V, et al. Oort: Efficient federated learning via guided participant selection[C]//Proceedings of the 15th USENIX Symposium on Operating Systems Design and Implementation, 2021: 19-35.
- [23] CHAI Z, CHEN Y J, ANWAR A, et al. FedAT: A high-performance and communication-efficient federated learning system with asynchronous tiers[C]//Proceedings of the international conference for high performance computing, networking, storage and analysis. 2021: 1-16.
- [24] WU H D, WANG P. Fast-convergent federated learning with adaptive weighting[J]. IEEE Transactions on Cognitive Communications and Networking, 2021, 7(4): 1078-1088.
- [25] AAMER B, CHERGUI H, BENJILLALI M, et al. Entropy-driven stochastic policy for fast federated learning in beyond 5G edge-

- RAN[C]//2021 IEEE Global Communications Conference (GLOBECOM). IEEE, 2021: 1-6.
- [26] WANG D, ZHANG N F, TAO M X. Clustered federated learning with weighted model aggregation for imbalanced data[J]. China Communications, 2022, 19(8): 41-56.
- [27] ZHAO B R, CUI Q M, LIANG S Y, et al. Green concerns in federated learning over 6G[J]. China Communicaitons, 2022, 19(3): 50-69.
- [28] YU A, YANG Q K, DOU L H, et al. Federated imitation learning: a cross-domain knowledge sharing framework for traffic scheduling in 6G ubiquitous IoT[J]. IEEE Network, 2021, 35(5): 136-142.
- [29] ZHU Z Q, Wan S, Fan P Y, et al. Federated multiagent actor - critic learning for age sensitive mobile-edge computing[J]. IEEE Internet of Things Journal, 2021, 9(2): 1053-1067.
- [30] LU Z L, PAN H, DAI Y Y, et al. Federated learning with non-iid data: A survey[J]. IEEE Internet of Things Journal, 2024, 11(11), 19188-19209.
- [31] ZENG Y, MU Y K, YUAN J F, et al. Adaptive federated learning with non-IID data[J]. The Computer Journal, 2023, 66(11): 2758-2772.
- [32] YAN Z, YI Y Z, ZHANG J L, et al. Federated learning model training method based on data features perception aggregation[C]// 2021 IEEE 94th Vehicular Technology Conference (VTC2021-Fall). IEEE, 2021: 1-7.
- [33] XU X Y, LYU L J, MA X J, et al. Gradient driven rewards to guarantee fairness in collaborative machine learning[J]. Advances in Neural Information Processing Systems, 2021, 34: 16104-16117.
- [34] XU X Y, WU Z X, FOO C S, et al. Validation free and replication robust volume-based data valuation[J]. Advances in Neural Information Processing Systems, 2021, 34: 10837-10848.
- [35] MA X D, ZHU J, LIN Z H, et al. A state-of-the-art survey on solving non-iid data in federated learning[J]. Future Generation Computer Systems, 2022, 135: 244-258.
- [36] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. Imagenet classification with deep convolutional neural networks[J]. Advances in neural information processing systems, 2012, 25.
- [37] FAN X, WANG Y, HUO Y, et al. Bev-sgd: Best effort voting sgd against byzantine attacks for analog-aggregation-based federated learning over the air[J]. IEEE Internet of Things Journal, 2022, 9 (19): 18946-18959.