

联邦长尾学习研究综述

周弈志^{1),2)} 王军晓³⁾ 谢鑫⁴⁾ 王鹏飞^{1),2)} 贾西贝⁵⁾
齐恒^{1),2)} 秦宇辰^{1),2)}

¹⁾(大连理工大学计算机科学与技术学院 辽宁 大连 126024)

²⁾(大连理工大学社会计算与认知智能教育部重点实验室 辽宁 大连 126024)

³⁾(广州大学网络空间安全学院 广州 510006)

⁴⁾(天津大学智能与计算学部 天津 300072)

⁵⁾(深圳市华傲数据科技有限公司 广东 深圳 518000)

摘 要 联邦学习是一种基于分布式训练的机器学习技术,有效地解决了因联合建模而引发的用户间数据隐私泄露问题,因此在多个领域得到了广泛应用。然而,在实际的联邦学习应用中,统计异质性与长尾分布的共存成为一大挑战,严重影响了模型性能。因此,如何在保护数据隐私的前提下解决长尾问题,已成为当前的研究热点。本文综述了联邦长尾学习的研究成果,首先简要阐述了联邦学习的架构,并介绍了统计异质性、长尾学习及联邦长尾学习的核心概念与定义。接着,依据优化方法的差异,将联邦长尾学习的算法分为两大类:模型组件改进和基于算法的优化,并深入分析了每种算法的实现细节及其优缺点。同时,为了更好地为不同任务提供参考,本文整合了一些具有代表性的开源数据集、长尾划分策略、评价指标与对比实验。最后,针对未来的应用场景和研究方向,对联邦长尾学习进行了展望。期望通过本文的深入研究,能为这类问题提供更全面的解决方案,进一步推动联邦长尾学习技术在各个领域的广泛应用和发展。

关键词 联邦学习;长尾学习;联邦长尾学习;统计异质性;隐私保护;边缘智能

中图法分类号 TP18 **DOI号** 10.11897/SP.J.1016.2025.00779

A Survey on Federated Long-Tailed Learning

ZHOU Yi-Zhi^{1),2)} WANG Jun-Xiao³⁾ XIE Xin⁴⁾ WANG Peng-Fei^{1),2)} JIA Xi-Bei⁵⁾
QI Heng^{1),2)} QIN Yu-Chen^{1),2)}

¹⁾(School of Computer Science and Technology, Dalian University of Technology, Dalian, Liaoning 116024)

²⁾(Key Laboratory of Social Computing and Cognitive Intelligence (Dalian University of Technology),
Ministry of Education, Dalian, Liaoning 116024)

³⁾(School of Cyberspace Security, Guangzhou University, Guangzhou 510006)

⁴⁾(College of Intelligence and Computing, Tianjin University, Tianjin 300072)

⁵⁾(Shenzhen Audaque Data Technology Co., Ltd., Shenzhen, Guangdong 518057)

Abstract Federated learning, a machine learning method built on distributed training, effectively navigates the problem of data privacy leakage among users that joint modeling introduces. This effectiveness has led to the widespread use of federated learning-based methods in many fields.

收稿日期:2024-07-07;在线发布日期:2024-12-16。本课题得到国家自然科学基金面上项目(62072069)、中央高校基本科研业务费项目(DUT24JC001)、国家自然科学基金青年项目(62202080,62402336)、中国博士后科学基金面上项目(2023M733354)、辽宁省“揭榜挂帅”科技计划项目(2023JH1/10400083)、大连市优秀青年科技人才项目(2023RY023)资助。周弈志,博士研究生,中国计算机学会(CCF)会员,主要研究领域为边缘计算、联邦学习。Email:zhouyizhi@mail.dlut.edu.cn。王军晓,博士,副教授,中国计算机学会(CCF)会员,主要研究领域为基座模型、联邦学习。谢鑫,博士,副教授,中国计算机学会(CCF)会员,主要研究领域为无线感知、射频识别、联邦学习。王鹏飞,博士,副教授,中国计算机学会(CCF)会员,主要研究领域为分布式人工智能、大数据智能处理、联邦学习。贾西贝,博士,主要研究领域为分布式人工智能、数据挖掘。齐恒(通信作者),博士,教授,中国计算机学会(CCF)高级会员,主要研究领域为新一代互联网体系、数据中心网络、联邦学习。Email:hengqi@dlut.edu.cn。秦宇辰,博士研究生,主要研究领域为大数据挖掘、联邦学习。

Yet, in practical applications of federated learning, the combined existence of statistical heterogeneity and long-tailed distribution presents a considerable hurdle, dramatically affecting the model's performance. Consequently, a principal objective of contemporary research is the mastery of the long-tailed challenge in a manner that maintains data privacy. Most of the existing researches on long-tailed learning and statistical heterogeneity are not viable in a federated learning setting. Furthermore, there is a paucity of detailed and comprehensive overviews of the cutting-edge methods in this area. This paper presents a comprehensive review of the research accomplishments in federated long-tailed learning. To begin with, the architecture of federated learning is briefly explained, followed by an introduction to core concepts and definitions concerning statistical heterogeneity, long-tailed learning, and federated long-tailed learning. The main challenges of federated long-tailed learning are mentioned as well, including differences in data distribution, imbalance in categories, and inconsistent updates between clients during training. Basing on different optimization methods, the algorithms for federated long-tailed learning are classified into two major categories: Model Component Improvement and Algorithm-based Optimization. Model Component Improvement methods include virtual feature re-training, classifier calibration, and personalized federal learning. Algorithm-based Optimization methods include cost-sensitive learning, knowledge distillation, client selection, and balance enhancement. Each algorithm's implementation details, as well as its advantages and disadvantages, are thoroughly analyzed. Meanwhile, striving to provide useful insights for diverse tasks, this review compiles some representative open-source datasets, long-tailed division strategies, evaluation metrics, and comparative experiments. The evaluation metrics are classified into two major categories: performance evaluation metrics, which include Accuracy, Precision, Recall Rate, F1-Score, and computation or communication overhead, as well as long-tailed evaluation metrics, which include Multiclass Imbalance Degree, and Weighted Cosine Similarity. Additionally, potential application scenarios are considered, including: (1) the design of federated intelligence models for the detection of diseases in the medical field; (2) the application of federated long-tailed learning to intelligent marketing, as well as risk control in the financial field; (3) the development of intelligent recommender systems based on federated long-tailed learning in the web industry; (4) the use of federated long-tailed learning to enhance the vehicles' ability to handle complex situations in the field of intelligent driving. In the end, we provide insights into the future of federated long-tailed learning and directions for further research: (1) analyze the issue of security and privacy in the training process of federated long-tailed learning; (2) establish more general mechanisms to cope with the demands of multimodal data tasks and complex scenarios; (3) explore the combination of long-tailed learning with personalized federated methods based on knowledge distillation and model segmentation; (4) design efficient light-weighting schemes to reduce the communication and computation overhead of federated long-tailed algorithms; (5) combining federated learning with multi-task learning, and designing federated multi-task learning methods that can balance inter-task relationships and data privacy requirements. Through this work, we strive to present comprehensive solutions for these types of issues, and to promote the application and growth of federated long-tailed learning technologies in various fields.

Keywords federated learning; long-tailed learning; federated long-tailed learning; statistical heterogeneity; privacy preserving; edge intelligence

1 引言

当前,新一轮科技革命和产业变革持续深化,新质生产力展现出鲜明的时代特征。人工智能作为新质生产力的重要驱动力^[1],加快发展新一代人工智能对于抓住数字经济时代的机遇、加速形成新质生产力具有重要意义。然而,人工智能的持续发展面临着数据和法律两大困境。

首先,数据困境指的是人工智能与机器学习算法对数据的高度依赖。然而,在现实中,许多行业的数据量有限且质量参差不齐,往往以碎片化的形式分布,无法满足人工智能技术的需求。同时,各个数据源之间存在难以逾越的壁垒。由于行业竞争、隐私安全、行政手续复杂等因素,数据常常以孤岛的形式存在,而数据收集者往往并非数据使用者,这使得将分散的数据整合用于机器学习的成本极高^[2]。其次,法律挑战同样不可忽视。随着社会对数据隐私安全关注不断增加,我国相继制定并强化了对数据安全和隐私保护的相关法规。例如,全国信息安全标准委员会于2017年12月和2020年3月发布了两版《信息安全技术个人信息安全规范》,对个人信息的收集、储存和使用进行了明确规定。此外,自2017年实施的《中华人民共和国网络安全法》和《中华人民共和国民法总则》也明确指出,网络运营者不得泄露、篡改或毁坏其收集的个人信息,且在与第三方进行数据交易时需确保在合同中明确约定拟交易数据的范围及数据保护义务。2024年,工业和信息化部更是起草了《工业领域数据安全能力提升实施方案》,为工业数据安全提供法律保障,并进一步规范了数据的合法合规使用。

为应对上述挑战,谷歌研究人员于2016年首次提出了联邦学习(Federated Learning, FL)的概念^[3],现已逐渐发展成为一种解决数据孤岛问题,同时满足隐私保护和数据安全需求的新范式^[4]。联邦学习的核心特性包括数据本地化处理、参与者身份地位平等、数据隐私保护。此外,联邦学习可以协同进行模型训练与预测,其模型效果与将所有数据集中处理的效果相当或仅有微小差异,从而实现企业间的数据融合建模,有效解决数据孤岛问题。公开资料显示,联邦学习已在金融、物联网和医疗等领域得到了成功应用^[5]。

尽管联邦学习取得了显著成就,但其成功主要依赖于参与者共享相同网络结构和相似的数据分

布。然而,在真实场景中,两种特殊的数据分布现象可能对联邦学习的鲁棒性构成威胁。第一种现象是统计异质性(Statistical Heterogeneity)^[6],即不同参与者收集的数据可能呈现出非独立同分布(Non-Independent and Identically Distributed, Non-IID)特性^[7],这可能导致参与者的更新优化方向出现差异,从而使原始方法在这种偏斜的协作环境下失效。第二种现象是全局类别的长尾分布^[8](Long-Tailed Distribution),即所有参与设备的本地数据在全局范围内展现出长尾或类别不平衡,其中头部类别的训练样本占比远超尾部类别。在这种数据分布条件下,直接采用联邦平均会降低生成的全局模型的泛化能力,聚合结果偏向头部类别。更为复杂的是,当本地训练数据同时呈现统计异质性和长尾分布时,联邦学习的复杂性相应增加,因为每个客户端可能都拥有独特的尾部类别数据。此外,标准的联邦学习客户端选择策略会导致拥有少数样本的客户端在某些迭代中无法加入聚合,这种排除减少了模型训练的多样性^[9]。

目前,针对统计异质性和集中式长尾的解决方案均不适用于联邦长尾分布环境。前者受非独立同分布数据分区的影响,导致局部数据分布与全局数据分布的长尾模式不一致,解决的仅是局部数据不平衡问题,无法有效缓解联邦学习中的全局数据不平衡。后者则由于数据隐私的限制,无法通过局部数据推测全局不平衡,这进一步限制了全局类重加权策略的应用^[10]。因此,联邦长尾学习(Federated Long-Tailed Learning, Fed-LT)提供了一种新的优化方案^[11],旨在保护客户数据分布不被泄露的同时,减轻客户更新优化方向的偏差。鉴于联邦长尾学习研究的飞速发展以及项目落地过程中长尾分布的实际存在^[12],加之我国在这一领域的系统研究仍显不足,本文对涉及联邦长尾学习的解决方案与应用场景进行了细致调研,旨在展现联邦长尾模型在这些策略优化下所表现出的更好鲁棒性。

与现有其他已发表的联邦长尾学习综述^[13]相比,本文有以下创新之处:(1)对联邦长尾任务进行了更为详尽的定义和描述,深入探讨了该领域的核心概念。不仅考虑了本地客户与全局服务器之间的关系,还将联邦长尾问题细分为三大类型,并介绍了长尾划分的具体实现方法,这对于在公开数据集上模拟长尾分布进行训练至关重要;(2)本文系统梳理了联邦长尾学习领域的最新成果,进行了综合性的总结与分类,有选择性地排除了质量较低或与联邦长尾学习相

关性较弱的文章;(3)在公开数据集 CIFAR-10 和 CIFAR-100 上测试了不同联邦长尾学习方法的性能与速度,深入分析了各类解决方案的优缺点;(4)本文是第一篇关于联邦长尾学习的中文综述性文章,相较于

于现有的英文综述,为国内研究人员提供了更清晰的研究视角,使其能够更系统、更快速地了解联邦长尾学习领域及其未来发展趋势。本文与其他联邦长尾学习综述的详细对比展示如表 1。

表 1 本文与其它联邦长尾学习综述对比

综述文章	联邦长尾任务定义	综述文献前沿性	综述文献代表性	联邦长尾算法分类	数据划分策略/评价指标	性能对比测试	总结与展望
本文	详细的定义与描述	整理了最新的 前沿文献	总结了最有 代表性文献	对联联邦长尾算法 清晰分类	详细的划分策略与 评价指标	包含清晰的 性能测试	完整的总结 与展望
Zhang 等人 ^[11]	未对联联邦长尾 明确定义	整理了较新的 前沿文献	总结了较有 代表性文献	对联联邦长尾算法 较清晰分类	仅包含详细的 评价指标	不包含性能 测试	较完整的总结 与展望
Li 等人 ^[12]	清晰的定义与 描述	整理了较新的 前沿文献	总结了不具有 代表性文献	对联联邦长尾算法 清晰分类	简单的划分策略与 评价指标	包含清晰的 性能测试	较完整的总结 与展望
Chen 等人 ^[13]	简单的定义与 描述	整理了少量 前沿文献	总结了有代表性的 少量文献	未对联联邦长尾算法 清晰分类	未包含划分策略与 评价指标	包含简单的 性能测试	简单的总结 与展望

本文深入探讨了长尾分布对联联邦学习造成的独特危害,并从多个角度进行了全面分析。具体章节安排如图 1 所示:第 2 节首先对联联邦学习的架构进行了简要阐述,并介绍了统计异质性、长尾学习和联邦长尾学习的核心概念和定义;第 3 节根据优化方法的差异,将联邦长尾学习的算法划分为两大类:模型组件改进和基于算法的优化,并对每种算法的实现细节及优缺点进行了深入分析;第 4 节整合了一些具有代表性的开源数据集、长尾划分策略、评价指标与对比实验;第 5 节对联联邦长尾学习的未来应用场景和研究方向进行展望;最后,第 6 节对全文进行了总结。

2 联邦长尾学习概述

在本节中,首先详细介绍联邦学习的通用训练流程,然后介绍联邦学习中统计异质性的相关理论。接着系统地阐述联邦长尾学习的特性,特别强调其与一般联邦学习的主要区别。最后,深入探讨在联邦长尾下所面临的挑战。

2.1 联邦学习

联邦学习作为一种新兴的分布式机器学习方法,强调不同设备之间的协同训练。在确保数据不离开本地的前提下,通过设备间的参数传递、梯度聚合和模型更新的有序机制,实现了联合建模。这一方法在维护数据隐私安全的基础上,解决了数据孤岛问题,为多方实体的合作求解提供了一种新的机器学习范式。

为了准确阐述联邦学习的核心思想,本文对联联邦学习架构进行定义(如图 2 所示)。假设联邦学习

系统包含 K 个客户端,由一个中心服务器进行协调。第 k 个客户端拥有私有数据集 $D_k = \{(\mathbf{x}_i^k, y_i^k)\}_{i=1}^{N_k}$,其中, N_k 表示客户端 k 的数据量, \mathbf{x}_i^k 和 y_i^k 表示客户端 k 中第 i 个样本的特征和对应标签。此外,客户端 k 通常具有可学习的本地网络模型或初始化模型 w_k ,并且 $f(\mathbf{x}^k, w_k)$ 表示局部模型 w_k 对私有样本 \mathbf{x}^k 的预测输出。

联邦学习的训练流程具体如下:

(1)模型初始化:服务器初始化全局模型参数 w_{global}^0 ,选择 m 个客户端($m \leq K$)参与联邦聚合任务,并向所有客户发送全局模型 w_{global}^0 。

(2)本地训练:在第 t 轮训练中,客户端 k 使用自己的本地数据集 D_k 训练局部模型参数 w_k^t 。这一过程可以通过随机梯度下降^[14](Stochastic Gradient Descent,SGD)或者其他优化算法完成。以随机梯度下降为例,其参数更新如公式(1)所示。

w_k^{t+1} ← w_k^t − η ∇_w L_k(f(x^k, w_k^t), y^k) (1)

其中, η 是学习率, $\nabla_w L_k(f(\mathbf{x}^k, w_k^t), y^k)$ 是损失函数 $L_k(\cdot)$ 对当前模型参数 w_k^t 的梯度。

(3)上传模型更新:客户端完成本地训练后,将更新后的模型权重上传到服务器。

(4)模型聚合:服务器收集所有客户端上传的模型权重后,将这些模型参数进行聚合,形成新的全局模型参数 w_{global}^{t+1} 。该步骤通过公式(2)计算所有客户端本地模型参数的加权平均来实现。

w_{global}^{t+1} ← ∑_{k=1}^K (N_k / N) w_k^{t+1} (2)

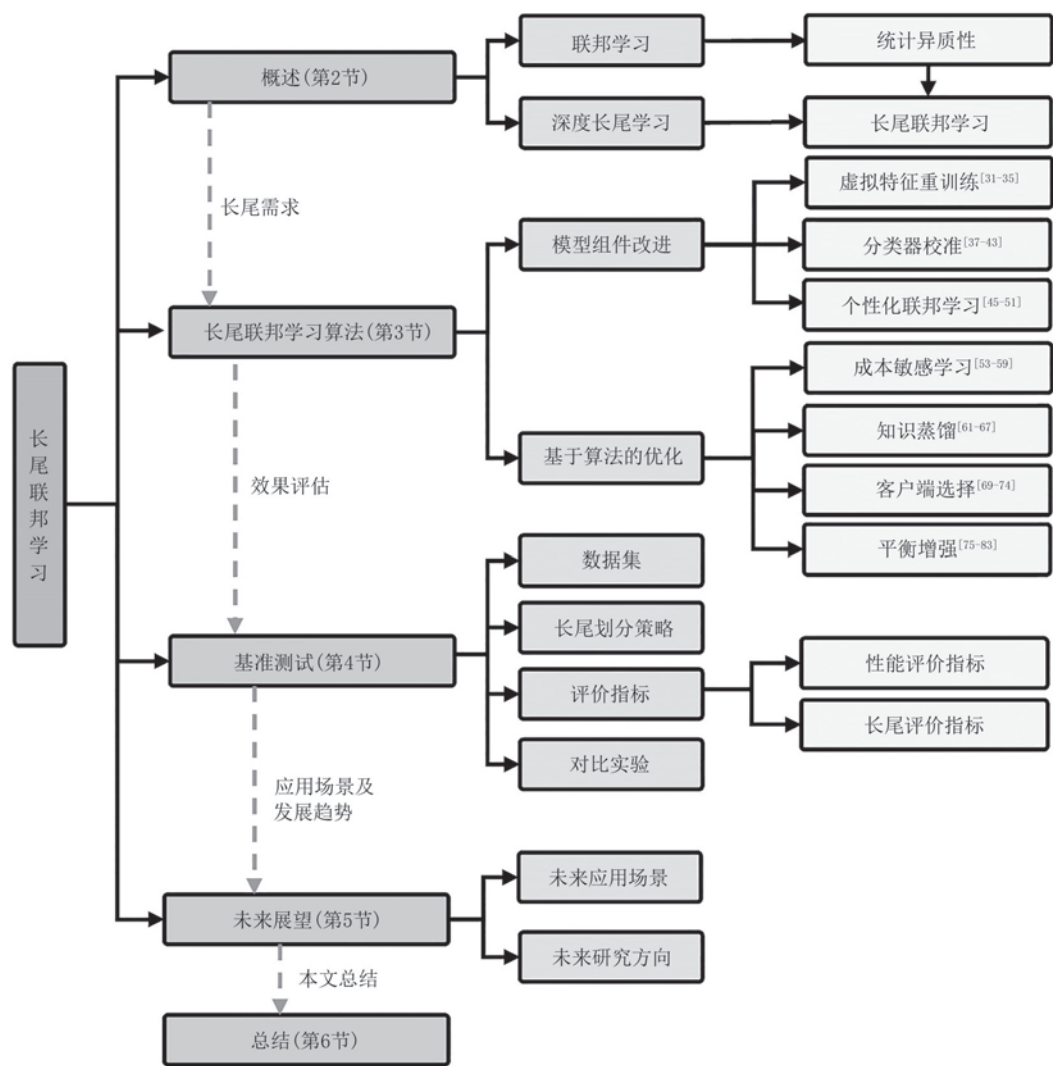


图1 文章结构安排

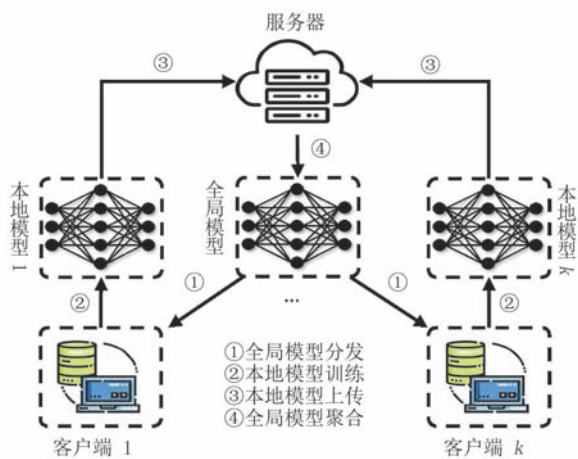


图2 联邦学习架构

其中, N 为所有客户端总的数量。
(5) 迭代更新: 服务器更新全局模型 w_{global}^{t+1} 后, 将

更新后的模型再次分发给各个客户端。客户端在接收到更新后的全局模型后, 使用新的模型在本地数据上进行下一轮训练。该过程不断迭代, 直至全局模型收敛或达到设定的终止条件。然而, 上述方法并不适用于数据非独立同分布的情况。

一方面, 每个本地客户端的数据可能具有不同的分布, 这可能导致本地更新在不同方向上推动全局模型, 最终造成更新的不一致性。另一方面, 每个客户端的数据量可能不同, 当这些数据不是独立同分布时, 模型的更新会被数据量较大的设备主导, 而数据量较小的设备的信息可能被忽视。

2.2 联邦学习中统计异质性

统计异质性是指水平联邦学习^[15]的各个客户端所持有的数据在统计分布上的差异, 也就是非独立同分布现象。现有的研究通过考虑数据分布的差异

对统计异质性进行了分类^[16]。简洁地说,统计异质性可被划分为以下五种主要情况:(1)标签分布偏斜;(2)特征分布偏斜;(3)相同标签但不同特征;(4)相同特征但不同标签;(5)数量偏斜。

第三种情况主要与垂直联邦学习^[17]相关(即各方共享相同的样本编号,但特征不同)。本文则聚焦于水平联邦学习,其中每个客户端共享相同的特征空间,但拥有不同的样本。除此之外,第四种情况在大多数联邦学习研究中并不适用,因为这些研究通常假设参与各方拥有共同的领域知识。因此,本文主要考虑标签分布偏斜、特征分布

偏斜和数量偏斜作为潜在的非独立同分布形式。接下来,本文将介绍针对每种情况的模拟策略。在此之前,定义两个不同的客户端 i 和 j 。客户端 i 的局部数据分布表示为 $P_i(x, y)$,其中 x 和 y 分别表示数据样本的特征和标签。

2.2.1 标签分布偏斜

标签分布偏斜^[18]指的是参与客户端之间的标签分布存在差异,即 $P_i(y) \neq P_j(y)$ 。为描述各种标签分布偏斜的情形,本文引入了两种不同的设置:基于数量的标签不平衡和基于分布的标签不平衡。一个视觉示例如图3(a)所示。

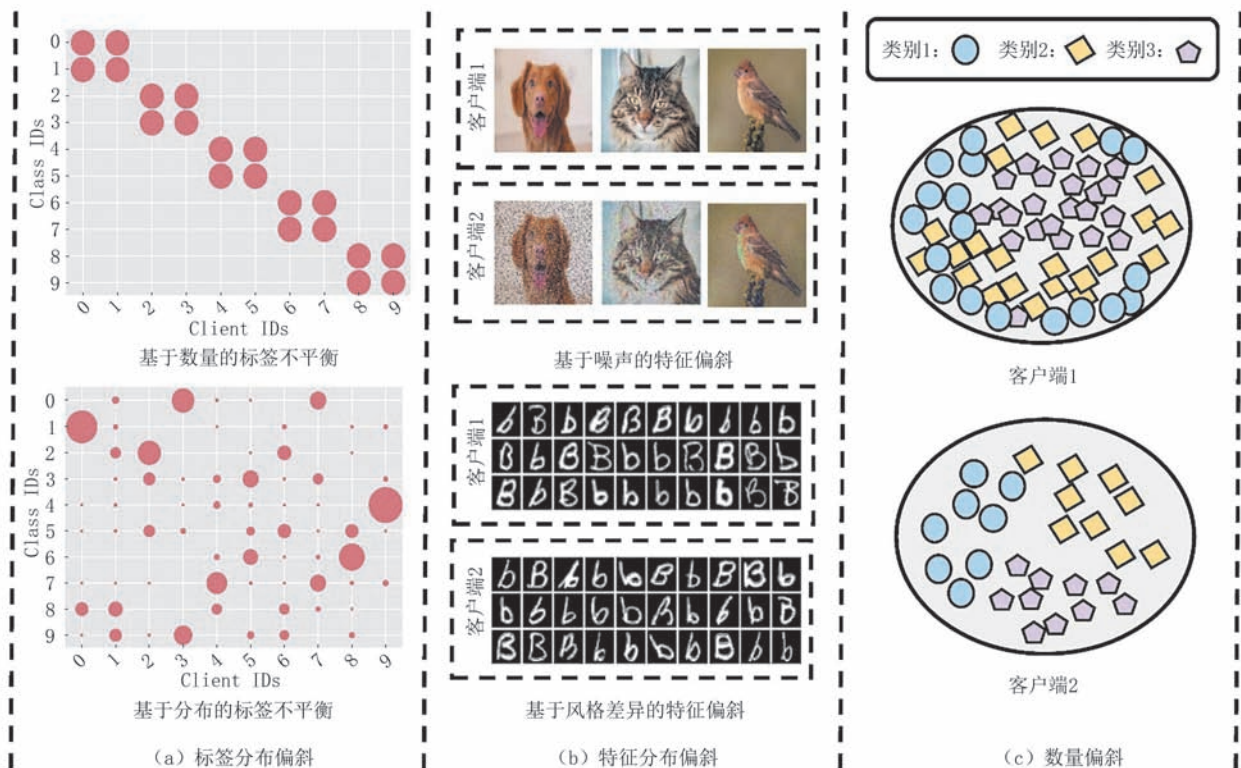


图3 统计异质性中三种不同的倾斜模式的示例

(1) 基于数量的标签不平衡:在面临数量的标签不平衡时,每个客户端拥有固定数量的标签样本。这种问题最初在联邦平均算法^[19] (Federated Averaging, FedAvg)的实验中被提及。实验中,相同标签的数据样本被切割为多个子集,而每个客户端只被分配到两个不同标签的子集。后续研究中,经常会见到类似的设定。后来, Li等人^[20]提出了一种通用的划分策略,用于设定每个客户端拥有的标签数量。假设每个客户端只拥有 k 个不同标签的数据样本。首先,为每个客户端随机分配 k 个不同的标签编号,然后,将每个标签的样本随机且均等地分配

给拥有该标签的客户端。通过这种方式,每个客户端拥有的标签数量是固定的,且不同客户端间的样本没有重叠。为了表述方便,使用 $C=k$ 来表示这种划分策略。

(2) 基于分布的标签不平衡:模拟标签不平衡的另一种方法是根据狄利克雷分布^[21]为每个客户端分配每个标签样本的比例。狄利克雷分布常用于贝叶斯统计作为先验分布,是模拟现实数据分布的理想工具。具体而言,该策略从狄利克雷分布中抽取 $p_k \sim \text{Dir}_N(\alpha)$,然后按比例 $p_{k,j}$ 将类别 k 的样本分配给客户端 j 。其中, $\text{Dir}(\cdot)$ 表示狄利克雷分布, α 为超参

数。这种划分策略已在多项近期研究中得到应用。其优势在于,可以通过调整参数 α 来灵活地改变不平衡程度。当 α 值较小时,划分结果会更加不平衡。

2.2.2 特征分布偏斜

分布是一致的,即 $P_i(y|x)=P_j(y|x)$,但特征分布可能不同,即 $P_i(x)\neq P_j(x)$ 。在这里,本文介绍了两种不同的设置来模拟特征分布偏差:基于噪声的特征偏斜与基于风格差异的特征偏斜。一个视觉示例如图3(b)所示。

(1) 基于噪声的特征偏斜:噪声特征偏斜是指由于各客户端数据采集过程和设备条件的不同,导致数据的质量存在显著的差异。这意味着,即使各个客户端可能在处理同样的任务,由于设备性能差异或采集条件的不同,数据质量可能有显著的不同。以声音识别为例,尽管所有用户可能说出同样的句子,但不同设备的环境噪声等各种因素可能导致采集到的数据质量存在差异,从而导致基于噪声的特征偏斜。

(2) 基于风格差异的特征偏斜:风格特征偏斜是指由于客户端地理位置、用户行为习惯和设备类型的不同,导致数据特征的边际分布有所差异。这意味着,尽管各个客户端可能处理相同类型的任务(如图像识别或文本分类),它们的数据特征仍可能存在显著不同。以手写识别任务为例,即便所有用户都在书写相同的词语,不同用户的笔划宽度、倾斜度等特征仍可能存在较大差异。

2.2.3 数量偏斜

数量偏斜指的是不同客户端拥有的数据量存在显著差异^[22-23]。例如,某些客户端可能拥有大量数据,而其他客户端的数据量则相对较少。如图3(c)所示,这种现象可能导致训练不稳定、数据冗余、有偏聚合等问题,进而对模型的总体性能及公平性产生负面影响。

2.3 长尾学习

长尾学习是深度学习在处理类别高度不平衡数据时面临的一个重大挑战^[24]。在长尾数据集中,尾部类的样本数量远远少于头部类,这往往导致模型在训练过程中倾向于过拟合头部类,生成有偏的分类决策。长尾学习的主要目标是在类别不平衡的条件下,合理确定决策边界,并为尾部类提供与头部类同等的关注,从而提高整体性能。

为了解决长尾学习存在的问题^[25],研究人员提出了多种方法,包括重采样、成本敏感学习、解耦训练和集成学习等。其中,重采样方法^[26]通过调整训

练数据的分布来平衡各类别的样本数量。常见的重采样技术包括随机过采样和随机欠采样;成本敏感学习^[27]则通过为不同类别分配不同的分类成本,使得模型在训练过程中更关注少数类样本的正确分类。常见的成本敏感方法包括加权损失函数和代价敏感决策树等;解耦训练是一种将特征学习与分类器学习分离的策略^[28],通过在两个阶段进行优化来提升模型的长尾识别能力。第一阶段模型专注于提取通用的特征,而第二阶段则进行分类器的微调,特别关注长尾类别的准确性;集成学习通过组合多个弱分类器来构建一个更强的分类器,从而提升模型的鲁棒性^[29]。对于长尾学习,集成学习方法可以通过融合多个模型的预测结果,有效增强少数类样本的检测能力。常见的集成方法包括自助聚合、堆叠泛化与多专家模型等。

这些方法为提高算法在长尾分布下的鲁棒性和稳定性提供了宝贵途径。然而,长尾学习方法通常依赖于全局类别分布信息,以强调类别属性及其不平衡性。在联邦学习中,收集每个客户端的类分布信息以构建全局分布是不切实际的,因为这会侵犯客户端的隐私。此外,各客户端可能有不同的尾部类别,这使得某些长尾学习方法在联邦学习场景中难以适用。

2.4 联邦长尾学习

在当前大多数针对联邦学习中非独立同分布问题的解决方案中,实验通常假设全局数据分布是平衡的,而忽视了数据可能呈现长尾分布的情况。事实上,长尾分布的存在为解决非独立同分布带来了更大的挑战。首先,由于非独立同分布数据的特性,局部数据和全局数据的长尾分布往往不一致。即使本文解决了局部数据的不平衡问题,也无法从根本上消除全局数据固有的不平衡性。其次,在保证数据隐私的前提下,无法直接获得全局不平衡数据分布的详细信息,这使得实施全局类别重加权策略变得复杂。因此,在联邦学习框架下,深入探讨全局长尾分布对非独立同分布的影响显得尤为重要,并基于此提出联邦长尾学习的概念。

联邦长尾学习是一类特殊的联邦学习任务,其中本地客户端的数据表现出统计异质性,而全局数据分布则呈现长尾特性。以医疗应用为例,不同医疗机构可能希望基于其私有的患者记录共同构建一个诊断模型。在每个机构的本地数据中,数据分布可能差异显著,且疾病频率呈长尾分布,一些疾病非常常见,而另一些则相对罕见。在这种情况下,由于统计异质

性和长尾分布的共同作用,本地模型在罕见类别上的表现往往较差,进而影响全局模型的聚合效果^[30]。

为进一步阐述联邦长尾学习,本文分析了本地客户端和全局服务器数据分布之间的关系及其差异,将联邦长尾学习任务主要分为如图4所示的

三种类型:类型1是指本地客户端为标签不平衡,并且全局数据呈现长尾分布;类型2为本地客户端分布不平衡,并且全局数据长尾分布;类型3是指本地客户端呈现分布不平衡,各个客户端间的样本数量倾斜,并且全局数据长尾分布的情况。

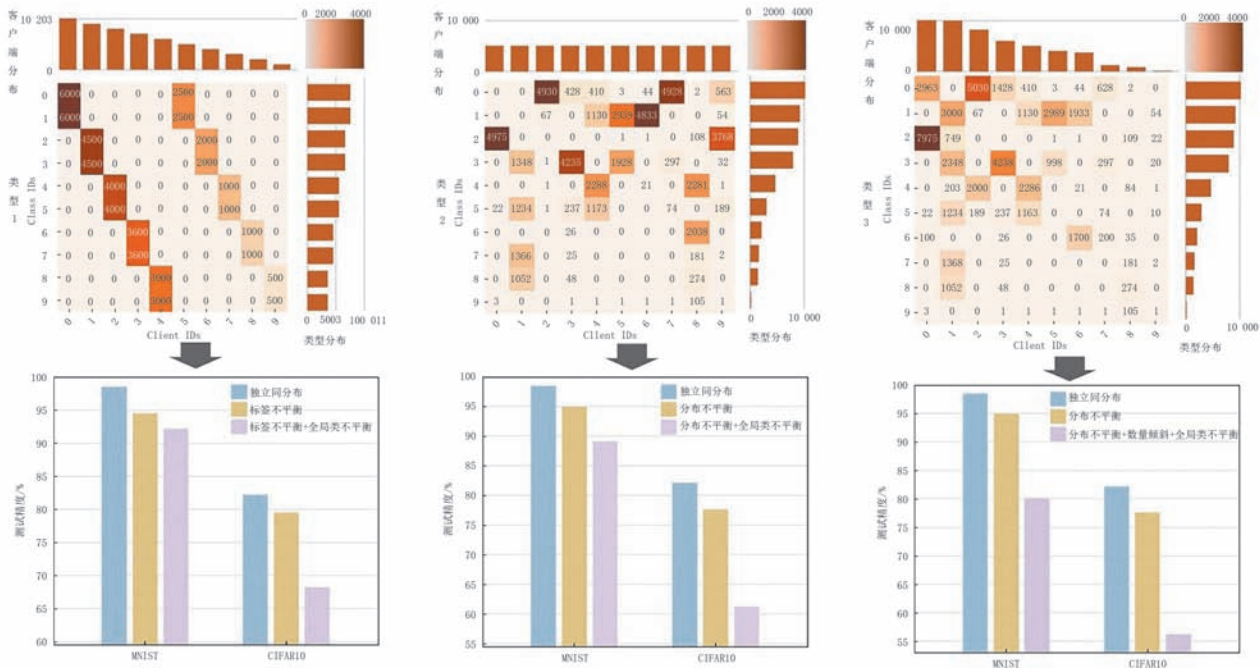


图4 三种类型的联邦长尾数据分布示例及影响

这三种情况代表了典型联邦学习系统中长尾分布可能遇到的所有场景。通过观察可发现,数据的全局不平衡性会进一步加剧非独立同分布场景下模型精度的下降。需要强调的是,联邦长尾学习并未考虑特征分布偏斜问题,这是因为全局长尾分布并不会显著加剧特征分布偏斜所带来的挑战。

2.5 联邦长尾学习挑战

在联邦长尾学习中,由于无法直接掌握用户数据的分布情况,如何有效聚合出理想的全局模型面临以下挑战:

(1)客户端与服务器之间的分布差异与不可知性:在联邦长尾学习中,由于每个客户端持有的数据分布不同,客户端之间的更新可能表现出极大的不一致性,尤其是在某些客户端数据集中存在少量类别甚至缺失类别的情况下。服务器端无法直接观察各客户端的实际数据分布,因此在全局模型的聚合过程中,服务器无法精确判断某些类别的贡献大小。由于数据分布的不可知性,服务器可能高估或低估某些客户端在全局模型中的重要性,导致最终

模型的泛化性能下降。

(2)类别不平衡和类别缺失:联邦长尾学习中的一个核心问题是类别不平衡,即某些客户端数据集中的某些类别占据绝大部分,而另一些类别则可能极其稀缺,甚至完全缺失。这种类别不平衡导致的后果是全局模型容易过拟合于那些在大多数客户端中占主导地位的类别,忽视了少数类别的学习效果。此外,某些类别的缺失会导致模型对这些类别的预测能力不足,进而影响全局模型的整体性能。

(3)训练过程中客户端之间的更新不一致:传统机器学习通常假设训练和测试数据集是独立同分布的,但在联邦长尾学习中,无法获得全局或各客户端的类分布信息。这会导致全局模型在聚合阶段的不稳定性。为了解决这个问题,需要充分考虑具有精确标签的分布数据,以决定是否调整训练与测试数据集中少数类别的概率,同时合理处理本地客户端和其他客户端之间多数类别与少数类别的概率差异。

3 联邦长尾学习算法

本节综述了当前联邦长尾学习领域的主要算法。首先,本文根据不同的优化方法,将联邦长尾学习算法划分为两大类:模块组件改进和基于算法的优化。模块组件改进类算法主要包括虚拟特征重训练、分类器校准和个性化联邦学习。另一方面,基于算法优化的策略则包括成本敏感学习、客户端选择策略、知识蒸馏和平衡增强。此外,本文对每种算法的具体实现细节进行了详细剖析,并深入探讨了它们各自的优势与局限性。

3.1 模型组件改进

面向模型组件的改进通过优化长尾分布下得到的特征提取器或分类器,从而改进模型预测性能。这些方法分为虚拟特征重训练、分类器校准、个性化联邦学习。

3.1.1 虚拟特征重训练

虚拟特征重训练算法的核心思想是在现有联邦学习框架的基础上,通过服务器端生成虚拟特征,并利用这些特征对分类器进行再训练。具体流程如图5所示。将神经网络 ϕ_w 的参数 $w=\{u, v\}$ 分为特征提取器 u 和分类器 v 两部分。首先,本地客户端基于已有模型权重,计算每个类别的类别信息(如梯度、均值、协方差等),并将模型权重与类别梯度一同上传至服务器。随后,服务器对上传的模型权重进行加权聚合,生成全局模型权重,并根据类别信息反向推导出虚拟特征。最后,服务器使用生成的虚拟特征对全局模型中的分类器进行重新训练。

鉴于类别信息在虚拟特征重训练中的重要性,

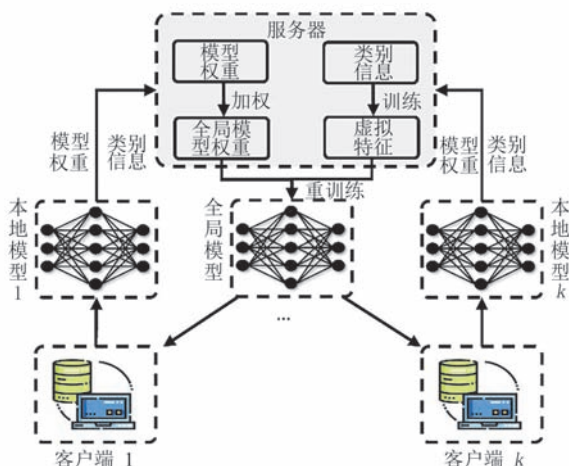


图5 虚拟特征重训练算法基本架构

CCVR^[31]假设在语义相关的分类任务中,深度学习网络学习到的特征可以近似为混合高斯分布。因此,数据集 D_k 中每个类别的特征被视为服从高斯分布。在此背景下,类别信息主要指均值和协方差,而虚拟特征则通过高斯分布生成。具体来说,服务器通过以下公式收集每个类别在客户端本地数据中的均值和协方差:

$$\mu_{c,k} = \frac{1}{N_{c,k}} \sum_{j=1}^{N_{c,k}} z_{c,k,j} \quad (3)$$

$$\Sigma_{c,k} = \frac{1}{N_{c,k} - 1} \sum_{j=1}^{N_{c,k}} (z_{c,k,j} - \mu_{c,k})(z_{c,k,j} - \mu_{c,k})^T \quad (4)$$

其中, $z_{c,k,j}$ 表示数据集 $D_{\{k,c\}}$ 中第 j 个样本的特征。

接下来,服务器计算全局均值和全局协方差矩阵,并通过高斯分布生成一组带有真实标签的虚拟特征 S :

$$\mu_c = \frac{1}{N_c} \sum_{k=1}^K \sum_{j=1}^{N_{c,k}} z_{c,k,j} = \sum_{k=1}^K \frac{N_{c,k}}{N_c} \mu_{c,k} \quad (5)$$

$$\Sigma_c = \sum_{k=1}^K \frac{N_{c,k} - 1}{N_c - 1} \Sigma_{c,k} + \sum_{k=1}^K \frac{N_{c,k}}{N_c - 1} \mu_{c,k} \mu_{c,k}^T - \frac{N_c}{N_c - 1} \mu_c \mu_c^T \quad (6)$$

$$S = \{\mathcal{N}(\mu_c, \Sigma_c)\}_1^c \quad (7)$$

最后,将生成的虚拟特征 S 传递至全局模型 w 中,并根据公式(8)对分类器 \hat{v} 进行重训练。

$$\hat{v} \leftarrow \tilde{v} - \eta \nabla_{\tilde{v}} \mathcal{L}(\tilde{v}, S) \quad (8)$$

Chen等人验证了CCVR生成的全局均值和协方差会偏向于头部类。为了解决这一问题,提出了加权协方差(Weighted Covariance, WC)与重采样特征(Re-Sample Features, RS)两种方法来生成高斯虚拟特征^[32]。具体而言,加权协方差根据全局类别分布为每个类别的协方差分配权重,头部类分配较小的权重,而尾部类分配较大的权重。重采样特征则通过构建不平衡分布,使尾部类的数量更多,头部类的数量更少。这种方式使模型在特征重叠时能够更好地区分尾部类与其他类别。

然而,通过高斯混合模型采样的虚拟特征与真实特征存在差距。此外,上传客户端的均值与协方差可能会泄露本地类别信息。为此,Shang等人^[33]提出联邦特征重训练算法CReFF,该方法不再上传均值和协方差,而是上传通过公式(9)计算的每个类别的真实特征梯度:

$$\mathbf{g}_c^k = \frac{1}{n_c^k} \sum_{i=1}^{n_c^k} \nabla_{\hat{\mathbf{v}}} l(h_{\hat{\mathbf{v}}}(\mathbf{z}_{c,i}^k), y_i) \quad (9)$$

其中, \mathbf{g}_c^k 是客户 k 中第 c 类的真实特征梯度。

在接收到真实特征梯度后, 服务器首先会计算全局平均特征梯度 $\mathbf{g}_c^{\text{agg}}$ 。然后, 通过优化梯度匹配损失, 学习联邦特征 \mathbf{S} , 从而生成靠近真实梯度的特征来更新分类器。以下是具体的优化公式:

$$\mathbf{g}_c^{\text{fed}} = \frac{1}{m} \sum_{i=1}^m \nabla_{\hat{\mathbf{v}}} l(h_{\hat{\mathbf{v}}}(\mathbf{s}_{c,i}), y_i) \quad (10)$$

$$D(\mathbf{g}_c^{\text{fed}}, \mathbf{g}_c^{\text{agg}}) = \frac{1}{C} \sum_{j=1}^C \left(1 - \frac{\mathbf{g}_c^{\text{fed}}[\mathbf{j}] \cdot \mathbf{g}_c^{\text{agg}}[\mathbf{j}]}{\|\mathbf{g}_c^{\text{fed}}[\mathbf{j}]\| \|\mathbf{g}_c^{\text{agg}}[\mathbf{j}]\|} \right) \quad (11)$$

然而, CReFF 算法用于分类器再训练的虚拟特征仍然来自长尾数据, 因此其改进仍受到每类特征间高度相似性的限制。为此, Zhou 等人^[34]提出联邦平衡奖励算法 FedCI。该算法由评估、混合数据增强和扰动三个部分组成, 其主旨在于生成平衡的本地数据集 D^{bal} 。然后利用 D^{bal} 训练真实特征梯度 \mathbf{g}_c^k 上传至服务器。下面是混合数据增强与扰动使用的公式:

$$D^{k,b} = \text{HybridAug}(D^k, p, q) \quad (12)$$

$$D^{k,bal} = \left\{ (\mathbf{x} + x^p, y) \right\}, x^p \sim \mathcal{N}(0, \sigma^p), (\mathbf{x}, y) \in D^{k,b} \quad (13)$$

其中, $\text{HybridAug}(\cdot)$ 是包含 20 种数据增强算法的集合, p 是使用基本数据增强算法的数量, q 是所有变换的幅度区间, x^p 是随机高斯噪声, σ^p 控制扰动强度。值得一提的是, FedCI 是第一个对虚拟特征重训练算法的收敛性和时间复杂度进行分析的工作。然而, 平衡监视模块的加入需要消耗大量算力。

除此之外, Shi 等人^[35]发现通过类别梯度生成的联邦特征缺乏语义监督, 可能导致重训练后的分类器在类别判别上存在不足。为此, 他们提出基于对比语言-图像预训练引导^[36]的联邦学习方法 CLIP2FL。具体而言, 该方法在服务器端利用预训练模型进行原型对比学习, 将联合特征 \mathbf{S}_c 及其类内原型 f_r^c 视为正样本, 而将联合特征 \mathbf{S}_c 及其类间联合特征视为负样本。原型对比学习的损失为

$$L_{\text{pcl}} = \sum_{i=1}^{C \times m} - \log \frac{\exp(\langle \mathbf{v}_{c,i}, f_r^c \rangle / \tau)}{\sum_{j=1}^{C \times m} 1_{[j \neq i]} \exp(\langle \mathbf{v}_{c,i}, \mathbf{v}_j \rangle / \tau)} \quad (14)$$

其中, $\langle \cdot, \cdot \rangle$ 表示余弦相似度, τ 是温度超参数。获得优化联邦特征的总损失如下:

$$L_{\text{total}} = D(\mathbf{g}_c^{\text{fed}}, \mathbf{g}_c^{\text{agg}}) + \eta \cdot L_{\text{pcl}} \quad (15)$$

虽然 CLIP2FL 能够提升模型的精度, 但其提升主要依赖于提前训练好的图像-文本预训练模型, 对于非分类任务并无显著帮助。

3.1.2 分类器校准

在分类任务中, 分类器通过学习输入数据的特征表示和对应的类别标签来预测数据的类别。然而, 长尾数据分布以及非独立同分布场景可能导致分类器难以有效区分某些类别, 尤其是样本数量较少的尾部类别。为应对这一问题, 分类器校准方法通过调整不同类别间的边距来提高模型的区分能力。具体流程如图 6 所示: 首先, 将模型解耦为特征提取器和分类器; 接着, 在本地客户端或全局服务器上构建小规模平衡数据特征, 并利用这些平衡数据特征对分类器进行校准; 最后, 使用校准后的分类器替换原有分类器, 从而实现联邦长尾学习。

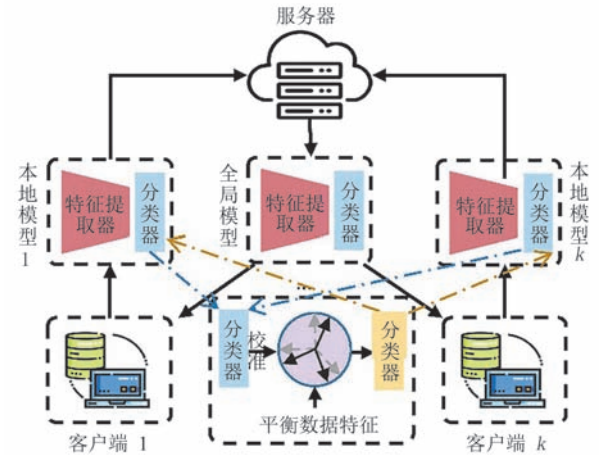


图6 分类器校准算法基本架构

Huang 等人^[37]提出了一种基于全局统一原型与分类器校准算法 RUCR, 用于处理长尾分布问题。该方法通过在客户端和服务端间交换全局统一原型来减少特征和分类器的偏差, 并通过生成平衡的虚拟特征集来修正分类器的表现。具体而言, 本地客户端上传本地原型与标签比例, 服务器则通过加权平均计算出全局原型 \mathcal{G} 与全局先验集 \mathcal{Q} 。之后, 利用全局原型和先验集来改进标准 softmax 损失函数, 减少头部类别与尾部类别之间的差距, 优化后的损失函数如下:

$$\mathcal{L} = \mathcal{L}_{\text{CE}} + \lambda \cdot \mathcal{L}_{\text{RUCR}} \quad (16)$$

$$\mathcal{L}_{\text{RUCR}} = -\log \frac{\phi(y) \times \exp(s(\mathbf{z}, \mathcal{G}^y) / \tau)}{\sum_{k \in I} \phi(k) \times \exp(s(\mathbf{z}, \mathcal{G}^k) / \tau)} \quad (17)$$

其中, $\phi(y)$ 表示类别 y 的先验概率, $s(\mathbf{z}, \mathcal{G}^y)$ 为特征 \mathbf{z} 和类别 y 的全局统一原型之间的相似度得分。该优

化损失函数旨在增强类内紧密性和类间距离,从而促使模型生成一个类间距离均匀的特征空间。通过这一过程,模型能够学习到更平衡的类别特征空间,同时保留尾类的独特特性。然而,尽管RUCR提升了模型的精度,但上传本地标签的比例也引发了隐私泄露问题,导致用户信息的公开。

Yang等人^[38]提出一种融合本地真实数据与全局梯度原型的分类器校准算法RedGrape。具体来说,客户端计算每个类别的真实特征梯度 \mathbf{g}_c^k 并上传至服务器,服务器对所有类别的梯度进行加权求和,得到 \mathbf{g}_c^{agg} 并返回给客户端。此时,所有客户端会检查样本数是否达到阈值 Q ,并对达到阈值的类别进行随机抽样,形成 $D_{c,k}^{bal}$ 。然后计算对应的梯度贡献 $\mathbf{g}_{c,k}$,对于未达到阈值的类别,则采用服务器返回的 \mathbf{g}_c^{agg} 作为该类别的梯度贡献。最终,所有本地客户端汇总形成包含所有类别的梯度贡献,经过加权求和得到平衡类别梯度 \mathbf{g}_k^{bal} 。最后,使用平衡梯度重新训练校准后的分类器,公式如下:

$$\mathbf{u}_k = \mathbf{u}_k - \eta \left[\mathbf{g}_k^{local} + \lambda \mathbf{g}_k^{bal} \frac{\|\mathbf{g}_k^{local}\|}{\|\mathbf{g}_k^{bal}\|} \right] \quad (18)$$

其中, \mathbf{u}_k 代表客户端 k 的分类器, η 为学习率, λ 为平衡参数。尽管该方法有效校准了类别间的平衡性,但生成的全局梯度原型仍缺乏尾部类别的实质性特征,未能彻底解决尾部类数据稀少的问题。

Zhang等人^[39]从统计角度分析长尾分布的倾斜现象,而且理论上证实了交叉熵损失函数可能引发局部模型对多数类产生过度拟合的问题。为了解决这一问题,他们设计了一种基于细粒度分类器的校准算法FedLC,通过成对标签间的边距来校准对数偏差,减少局部更新中的误差:

$$\mathcal{L}_{Cal}(y; f(\mathbf{x})) = -\log \frac{1}{\sum_{i \neq y} e^{-f_y(\mathbf{x}) + f_i(\mathbf{x}) + \Delta_{(y,i)}}} \quad (19)$$

其中, $\Delta_{(y,i)}$ 为类别 y 与类别 i 间的成对标签边距, $f(\mathbf{x})$ 表示模型对输入样本 \mathbf{x} 的预测输出。然而,FedLC的性能依赖于从本地多数类与少数类随机采样得到样本对,模型性能好坏具有很强的随机性。

Zheng等人^[40]提出了改进的三元组度量学习方法FedMeta,该方法为每个锚点样本选择一个正样本和若干个负样本,并通过优化使锚点样本与负样本之间的距离大于与正样本之间的距离。然而,该方法对三元组选择流程的敏感度极高。任意随机选

择的三元组往往无法有效地收敛,而挑选出最为困难的三元组可能会使模型陷入劣质的局部最小值,这样反过来会削弱三元组损失学习的效果。

Li等人^[41]发现现有分类器校准算法在修正由分类器偏差导致的不良特征表示方面存在不足。为此,提出了基于神经塌缩^[42]的联邦学习算法FedETF,该算法使用合成且固定的单纯形等角分类器进行训练,以实现所有客户端学习统一且最优的特征表示。具体而言,模型将权重 w 拆分为特征提取器 u ,投影层 p 以及可学习温度 β ,并在训练过程中仅学习与聚合特征提取器与投影层,本地所采用的损失具体如下:

$$l(w, V_{ETF}; D) = -\log \frac{n_{k,y}^\gamma \exp(\beta \cdot \mathbf{v}_y^T \boldsymbol{\mu})}{\sum_{c \in [C]} n_{k,c}^\gamma \exp(\beta \cdot \mathbf{v}_c^T \boldsymbol{\mu})} \quad (20)$$

$$\boldsymbol{\mu} = \hat{\boldsymbol{\mu}} / \|\hat{\boldsymbol{\mu}}\|_2, \quad \hat{\boldsymbol{\mu}} = g(p; h), \quad h = f(u; x) \quad (21)$$

其中, $n_{k,c}$ 是指客户端 k 的 c 类样本数, γ 是可学习温度, $\boldsymbol{\mu}$ 是归一化特征, \mathbf{v}_c 是 V_{ETF} 中第 c 类的分类器向量。然而,FedETF对特征维度的要求较高,类别数越少,所需特征维度也越低。在实际任务中,类别数的确定往往困难,因此选择合适的特征维度成为该方法面临的挑战。

Wicaksana等人^[43]提出的联邦分类器锚定算法FCA,旨在处理客户类别分布的不均衡性。具体而言,FCA为每个客户端添加个性化分类器作为引导联邦模型的锚点,并采用平衡交叉熵损失对分类器进行校准,校准的目的是通过根据每个客户端的本地类别分布调整联邦和个性化分类器的预测,从而消除每个客户端的偏差。具体损失如下:

$$\mathcal{L}_k^{fed} = \frac{1}{n_k} \sum_{x_i \in D_k} -y_i \log \sigma(p_i^{fed} + \log \pi_k) \quad (22)$$

$$\mathcal{L}_k^{local} = \frac{1}{n_k} \sum_{x_i \in D_k} -y_i \log \sigma(p_i^{local} + \log \pi_k) \quad (23)$$

其中, $\sigma(\cdot)$ 表示平衡交叉熵损失, p_i^{fed} 与 p_i^{local} 分别表示联邦分类器和个性化分类器的预测, $\pi_k = [\pi_k^0, \dots, \pi_k^{|class|}]$, $\pi_k^c = |D_k^c|/|D_k|$ 是客户端 k 的类别 c 的频率。校准后,通过在联邦分类器的预测结果与每个客户端的个性化分类器之间进行一致性正则化,以增强长尾特征提取的鲁棒性。然而,当两个分类器进行相互正则化时,它们的目标对齐可能存在不一致,这可能会带来不利影响。

3.1.3 个性化联邦学习

在传统联邦学习的设定中,单一的全局共享模

型被训练用于适应平均化的客户端。然而,当各客户端的数据分布呈现明显的长尾分布时,单一全局模型难以处理与整体数据分布显著不同的局部分布。因此,研究人员提出了个性化联邦学习^[44],为每个参与的客户端提供定制化的模型,其具体过程如图7所示。首先,本地客户端将训练后的模型划分为共性知识和个性知识。其中,个性知识保留在本地,而共性知识被上传到服务器进行聚合,最终得到全局模型。随后,客户端利用其本地的个性知识对聚合后的共性知识进行进一步优化。

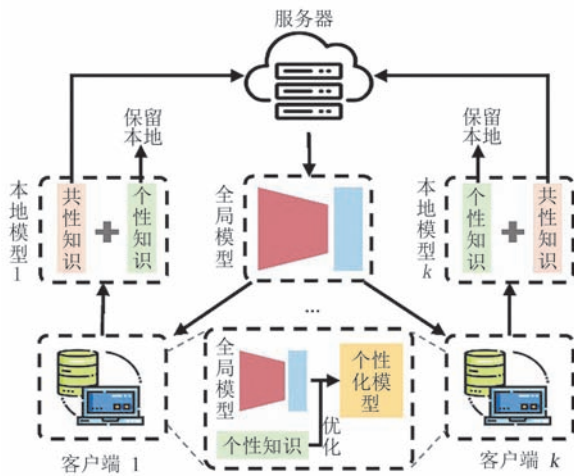


图7 个性化联邦学习算法基本架构

早期针对长尾分布的个性化联邦学习研究主要关注全局模型与个性化模型之间的优先级。例如,Chen 等人^[45]提出了联邦鲁棒解耦个性化算法 FedROD。并且引入非同类分布的类别平衡损失,使客户端在各任务间保持一致目标。此外,还将个性化分类器建模为轻量级自适应模块,通过最小化每个客户端的经验风险来学习该模块。总损失函数如下:

$$\mathcal{L}_{rod} = \mathcal{L}_m^{BR}(\{u, v\}) + \mathcal{L}_m(\{u, v, v_m\}) \quad (24)$$

其中, u 表示共享特征提取器的参数, v 表示全局分类器的参数, v_m 表示个性化分类器的参数。然而, FedROD 使用类别平衡损失来训练局部模型时,仅考虑本地的类别平衡分布,而未考虑全局类别分布。这种差异可能导致本地优化目标与全局目标之间不一致。

Xiao 等人^[46]提出联合本地和全局模型的联邦学习算法 FedLoGe,旨在解决长尾数据下的表示学习问题。该方法首先使用静态稀疏等角紧框^[47]分类器进行表示学习,随后服务器根据每类权重范数重新分配类别向量。最终,通过调整全局辅助分类器

的权重范数,将全局趋势与各客户端的偏好相结合,从而构建个性化模型。然而,在长尾分布下,特征空间中的某些特征可能退化,即均值较小且受到噪声污染,这种退化与神经崩溃特性不一致,可能导致模型性能下降。此外,尽管 FedLoGe 采用了自适应特征对齐模块,但在实践中,如何有效对齐全局与本地模型仍是一个挑战。

Chen 等人^[48]提出个性化的抗倒退弹性联邦学习框架 CPA。具体来说,该算法通过类别分布来定义全局联合目标的掩码 \hat{p}_c 。然后从每个客户端收集类别原型,并利用原型之间的相似性来计算原型距离。接着根据原型距离调整类别惩罚因子 γ_c 。最后,利用惩罚因子优化交叉熵损失,改进的损失如下:

$$L_{CPA}(\mathbf{z}) = -\sum_{c=1}^C \gamma_c y_c \log(\hat{p}_c) \quad (25)$$

然而,在长尾分布的场景中,使用余弦相似性计算客户端间的原型差异存在一定局限性。多数类与少数类之间的差异可能导致模型偏向多数类,而余弦相似性仅关注方向,未能有效反映数据分布的不平衡性。

Guo 等人^[49]提出了一种个性化联邦原型校正算法 FedPRP,旨在为类别分布不均的异构客户端创建个性化分类器。该算法通过优化客户端的经验风险,确保本地模型的个性化性能。此外, FedPRP 设计了类间差异和类内一致性损失,以校正有限样本中推导的经验原型,确保客户端本地训练的一致性,从而增强聚合后的全局模型在不均衡分布下的鲁棒性。然而,在模型上传阶段,除了上传已有模型权重外,还需要传递所有类别的经验原型,这可能导致用户数据分布泄露的风险。

Dai 等人^[50]提出基于类别原型的均匀性和语义信息的个性化联邦算法 FedNH,有效解决了类别不平衡问题。具体而言, FedNH 在本地训练过程中通过固定分类器训练得到特征提取器 u_k ,确保所有客户端具有一致的学习目标。接着,每个客户端计算并向服务器发送其所有类别的本地原型 $\{\mu_{k,c}^t\}_{c \in \mathcal{C}_k}$ 。随后,服务器运用下述公式将类别的语义信息平滑地融入到每个类别原型中:

$$\mathbf{v}_c^{t+1} \leftarrow \rho \mathbf{v}_c^t + (1 - \rho) \sum_{k \in \mathcal{S}^c} \alpha_k^{t+1} \mu_{k,c}^{t+1} \quad (26)$$

其中, \mathbf{v}_c 代表前一轮的全局分类器,聚合权重 $\{\alpha_k^{t+1}\}_{k \in \mathcal{S}^c}$ 是正标量,光滑参数 $\rho \in (0, 1)$ 。然而,在长尾分布条件下,单靠固定分类器无法获得理想的特征提取器,导致本地训练的原型不够可靠。此外, FedNH 仅适用于分类任务,其语义融合目前仅适用

于神经网络架构的头部层级,未来的研究可探索将其扩展至神经网络的中间层及其他视觉任务。

Zhang 等人^[51]提出的多级个性化联邦学习框架 MuPFL,通过在本地客户端引入偏置激活值的随机失活机制,并根据实时激活映射选择性保留关键特征,从而减少多数类的过拟合并加速训练。同时,该框架在模型聚合过程中动态聚类客户端模型,并在服务器端利用先验知识优化分类器。然而,假设服务器拥有平衡的先验知识在现实场景中并不合理。

表2对比了具有代表性的模型组件改进算法在性能、应用场景和优化目标上的差异。性能指标包括模型准确性、计算开销和通信开销,其中准确性指模型在测试集上的最终精度;计算开销是指在本地设备上执行训练所需的资源;通信开销则是指本地客户端与全局服务器之间模型交换产生的网络流量。此外,实心圆表示最佳准确度且计算与通信开销最大;半实心圆表示中等准确度和中等开销;空心圆则对应较低准确度和较小的开销。

表2 模型组件改进的联邦长尾学习算法对比

类型	算法名称	算法性能			应用场景	优化目标	
		准确性	计算开销	通信开销		本地客户	全局服务器
虚拟特征重训练	CCVR ^[31]	○	●	●	通用分类	✓	×
	DFL-FS ^[32]	●	●	●	通用分类	✓	×
	CRcFF ^[33]	●	●	●	通用分类	×	✓
	FedCI ^[34]	●	●	●	通用分类	×	✓
	CLIP2FL ^[35]	●	●	●	通用分类	✓	✓
分类器校准	RUCR ^[37]	●	○	○	通用分类	✓	✓
	RedGrape ^[38]	●	●	●	通用分类	✓	×
	FedLC ^[39]	●	●	○	通用分类	✓	×
	FedMeta ^[40]	●	●	●	金融预测	✓	×
	FedETF ^[41]	●	○	○	通用分类	✓	✓
	FCA ^[43]	●	●	●	医疗分类	✓	×
个性化联邦学习	FedROD ^[45]	●	○	○	通用分类	✓	×
	FedLoGe ^[46]	●	●	●	通用分类	✓	×
	CPA ^[48]	○	●	○	医疗分类	✓	×
	FedPRP ^[49]	●	●	●	通用分类	✓	✓
	FedNH ^[50]	○	○	●	通用分类	✓	×
	MuPFL ^[51]	●	●	●	通用分割	✓	×

通过对比发现,虚拟特征重训练算法的准确率较高,但需要额外的计算和通信开销来上传类别梯度和均值方差,同时还需服务器进行额外训练,整体训练时间较长。分类器校准算法同样具有较高准确率,且无需服务器额外训练,但部分算法在本地占用较大空间,需要模拟局部真实特征进行校准。个性化联邦学习算法的全局精度较低,但本地个性化模型精度较高,这主要与其优化目标集中于本地客户端有关。

3.2 基于算法的优化

除了模型组件改进之外,研究人员还探索了基于算法优化的方法。此类方法通过成本敏感学习、客户端选择、知识蒸馏、类别平衡增强等来缓解全局类别的不平衡。

3.2.1 成本敏感学习

成本敏感学习算法旨在优化机器学习模型,使

其对不同类别错误的成本敏感^[52]。通过为每个类别分配特定的错误成本,模型在训练时更关注高成本错误。算法调整各类别的权重,引导模型关注高成本类别,最终训练出最小化总体成本的模型。在联邦长尾环境中,成本敏感学习算法集中在设计算法以估计服务器的全局分布,并调高全局尾部类的权重,具体流程如图8所示。

Sarkar 等人^[53]借鉴焦点损失^[54]的思想,提出一种客户无需花费更多计算来纠正类别不平衡的损失函数 Fed-Focal Loss。该损失通过修改交叉熵来降低对分类正确的样本的惩罚,具体公式如下:

$$\mathcal{L}_{F-Focal} = -\alpha(1 - p_i)^{\gamma} \log(p_i) \tag{27}$$

其中,超参数 $\gamma \in [0, 5]$ 用来降低易分样本的贡献, α 用来缩放客户端的不同类别。

Wang 等人^[55]提出了一种基于联邦学习的自动调制分类方法 FedAMC,用于解决类别不平衡和

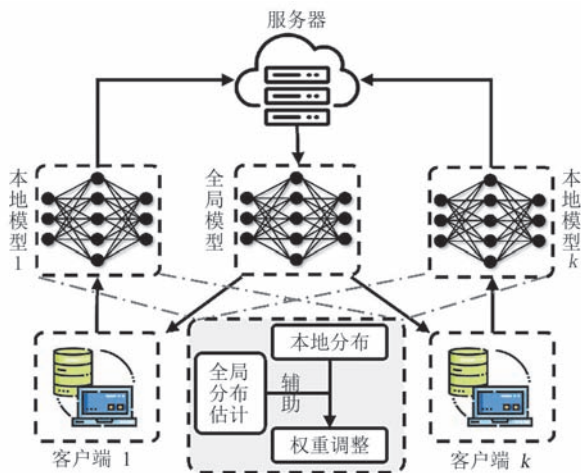


图8 成本敏感学习算法基本架构

变化噪声条件下的调制识别问题。具体而言，FedAMC引入平衡交叉熵作为损失函数，并增加光滑正则项以防止模型过度拟合训练数据，其具体损失如下：

$$\mathcal{L}_{AMC} = \alpha \cdot L_{CE} + \lambda L_{Smooth} \quad (28)$$

其中， α 是平衡因子， λ 是平衡两个损失的超参数。然而，Fed-Focal Loss与FedAMC都只考虑全部客户端参与聚合的场景，并没有考虑部分客户端参与时全局分布变化。

Wang等人^[56]发现真实场景下客户端会不断产生新数据，类别不平衡也会随之改变。因此，为了精确探测训练过程中的不平衡情况，他们建议利用服务器上的平衡辅助数据来推断全局类别分布。当某个特定的不平衡结构持续出现，监控系统将对此作出反应。基于这种思路，设计了新的损失函数Ratio Loss，以消除类别不平衡的影响，其具体损失如下：

$$\mathcal{L}_{RL} = (\alpha + \beta R) \cdot p \cdot L_{CE} \quad (29)$$

其中， R 表示每个类别对应的权重更新在所有更新中所占的比例， α 与 β 为超参数。当输入为少数类时，其对应的 R 相对较大，从而增加其对总损失的贡献；反之，则减少贡献。

Lu等人^[57]设计了一种估计全局类分布算法FedRE，通过在服务器引入辅助平衡数据集来估计各客户端的类分布，然后将估计的类分布应用在损失重加权来解决全局类别不平衡问题。然而，正如之前提到的，服务器上引用辅助数据并不可取。

相较于上述降低多数类所占权重，Shuai等人^[58]假设所有本地节点自平衡时，全局分布也将平衡，这避免了估计全局数据分布的难度。为了实现上述自

平衡，提出一种联邦局部更新算法Balance FL。算法将本地数据不平衡问题分解为类缺失和类间不平衡。针对类缺失采用知识继承进行增广，对于类间不平衡采用平衡采样、特征空间增强以及平滑正则化进行重平衡。并对自平衡后的数据设计了新的损失函数进行本地训练：

$$\mathcal{L}_{B-FL} = L_{KL} + L_{CE} + \lambda_1 L_{Smooth} \quad (30)$$

与对客户端类别赋予不同权重的方法不同，Qian等人^[59]提出联邦聚合元映射算法FedAMM。该方法通过在本地图训阶段为每个样本分配权重，并在服务器上聚合这些权重，从而减轻了数据不平衡和全局长尾的影响。值得说明，FedAMM需要使用一个小的平衡元数据集作为元知识的真实数据分布，以评估权重分配的有效性。然而，客户通常不允许拥有额外的辅助数据集。

3.2.2 客户端选择

客户端选择^[60]可以视为解决类别不平衡问题的一种策略。该策略通过选择与全局数据分布较为一致的客户端参与训练来实施。正如图9所示，将客户端选择过程视为一个优化任务，目标是在给定的约束和奖励条件下，找到与全局数据分布相似且类别分布平衡的客户端子集。随后，这些选定的客户端将参与后续的联邦聚合。

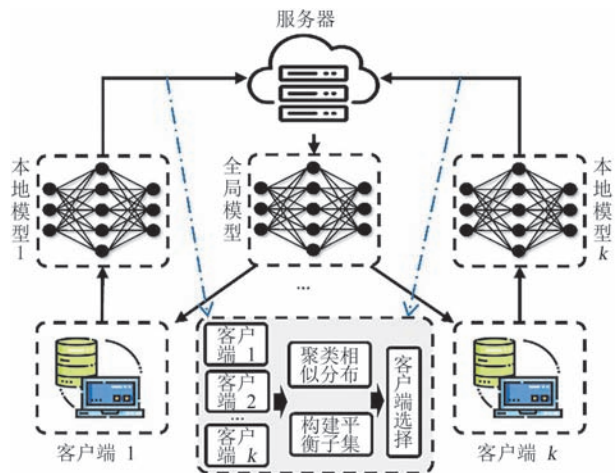


图9 客户端选择算法基本架构

早期的研究通常尝试寻找类别不平衡最小的客户端集进行聚合，其主要挑战在于如何通过迭代更新来捕捉类分布的统计特征。Yang等人^[61]针对上述问题提出FedCIR长尾学习框架。首先在服务器端构建了一个辅助平衡数据集，以估计不同类别间的梯度平方期望。接着，通过计算类别比例，得到表示全局数据分布的向量 U 。最后，采用强化学习中

的多臂老虎机理念^[62],将客户端选择过程视为在有限臂老虎机上进行投注的过程。在此过程中,每个客户端代表一个臂,所有客户端的集合则构成一个超臂。通过奖励与全局数据分布相似的客户端,算法能够选择出类别不平衡最小的客户端集合。具体的奖励机制如下:

$$r^k = 1/(D_{KL}(R^k||U)) \quad (31)$$

其中, R^k 表示本地客户端的数据分布, r^k 表示强化学习的奖励得分。然而,当客户端数量过多时,性能提升不明显。此外,服务器端持有辅助平衡数据集在实际应用中可能不可行。

相比直接寻求最小的标签类不平衡的客户端子集,Zhang等人^[63]假设本地设备中具有平衡类标签的客户端应更频繁地参与本地更新。基于此假设,设计了一种结合好奇心驱动^[64]的强化学习模型的局部更新方案FedSens。通过内在奖励 RI_i^t 与外部奖励 RE_i^t 来选择客户端。具体的,外在奖励与内在奖励的定义如下所示:

$$RE_i^t = \begin{cases} F1(M') - F1(M), t \bmod K == 0 \\ 0, t \bmod K \neq 0 \end{cases} \quad (32)$$

$$RI_i^t = \frac{|M_i^t - M_i^t|}{|M_i^t|} - \lambda \times e_i \quad (33)$$

其中, $F1(M')$ 为本地更新后全局模型的F1分数, $F1(M)$ 是全局模型在更新前的F1分数, $\frac{|M_i^t - M_i^t|}{|M_i^t|}$

表示更新后的本地模型与更新前的本地模型之间的差异, e_i 表示在执行本地训练时的能量成本。通过这种设计,算法能够在提升全局模型准确性的同时,最小化边缘设备的能耗。然而,强化本地类别平衡的策略并不完全合理,因为本地数据平衡并不代表包含所有类别的平衡。

为了有效选择客户端参与模型更新,Chen等人^[65]提出一种无偏客户端选择算法FedCSS。通过最大化全局损失来推断全局类别分布,并使用客户端的加权因子来估计全局类分布 D_g 。然后定义了一个基于客户端与全局类别分布余弦相似度的奖励机制,并加入正则化项以防止某些客户端被过度选择。奖励公式如下:

$$R_k = \frac{D_k \cdot D_g}{\|D_k\| \cdot \|D_g\|} + \gamma \log \left\| \frac{D_k \cdot D_g}{\|D_k\| \cdot \|D_g\|} \right\|^{-1} \quad (34)$$

其中, D_k 是本地类别分布,正则超参数 $\gamma \in (0, 1)$ 。然而,FedCSS奖励的是与全局分布相似的客户端,这并不能有效应对全局长尾分布带来的挑战。

相对于每轮通过强化学习选择参与聚合的客户端,Duan等人^[66]提出自平衡联邦长尾学习框架Astraea,其核心理念是在服务器与客户端之间设置中介层。中介层通过计算KL散度来评估与均匀分布的偏差,并使用以下公式将偏差相近的客户端进行聚类:

$$k \leftarrow \arg \min_i D_{KL}((P_m + P_i)||P_u), i \in \mathcal{S}_{client} \quad (35)$$

其中, $P_m + P_i$ 表示中介已有概率分布 P_m 与客户端 i 的概率分布 P_i , P_u 是均匀分布的概率分布。聚类完成后,每个中介负责协调其客户端进行训练,然后将训练权重发给服务器。服务器使用联邦平均聚合所有中介的权重。然而,对中介应用联邦平均可能会引入不公平聚合,因为不同中介负责的客户端数量可能会有巨大的差异,平均策略会导致那些只负责少量客户端的中介拥有了过大的权重。

不同于聚类采样,Zeng等人^[67]提出一种联邦多专家学习框架GBME。具体而言,服务器根据客户端的本地偏置信息与全局偏置信息之间的相似性,将所有客户端划分为不同的专家。接下来,在训练不同专家时,以集成方式选择不同的客户端。对于每个客户端,它们可能被选中多次以更新其所在组对应的专家以及其他组的专家。通过这种客户端选择策略,不同组的客户端可以相互协作,从而在保持各组知识的同时,从其他组中隐式地学习平衡知识。然而,GBME缺乏有效的负载均衡机制,这会导致某些客户过载而另一些客户闲置。并且随着专家数量的增加,算法可扩展性也将成为一大问题。

3.2.3 知识蒸馏

知识蒸馏通常涉及教师模型向学生模型传递知识的过程^[68]。在联邦长尾学习中,知识蒸馏通过共享模型的中间层特征、预测结果或其他形式的知识来实现,如图10所示。首先,客户端或服务器构建小型平衡的真实或虚拟数据集并进行训练。训练后的模型作为教师模型,用于指导学生模型进行分类器的校准训练。

Li等人^[69]受到虚拟特征重训练方法的启发,提出了一种全局虚拟特征蒸馏校准算法FedCRAC。核心思想是计算本地客户端所有类别的均值 μ 与方差 σ ,然后根据类相似性感知边际矩阵优化损失函数,其计算公式如下:

$$\mathcal{L}_{CRAC} = -\log \frac{e^{v_j^T x}}{\sum_{i \in C} e^{v_i^T x + \beta \cdot \frac{\langle \mu_i, \mu_j \rangle}{\|\mu_i\| \cdot \|\mu_j\|} I_{i \neq j}}} \quad (36)$$

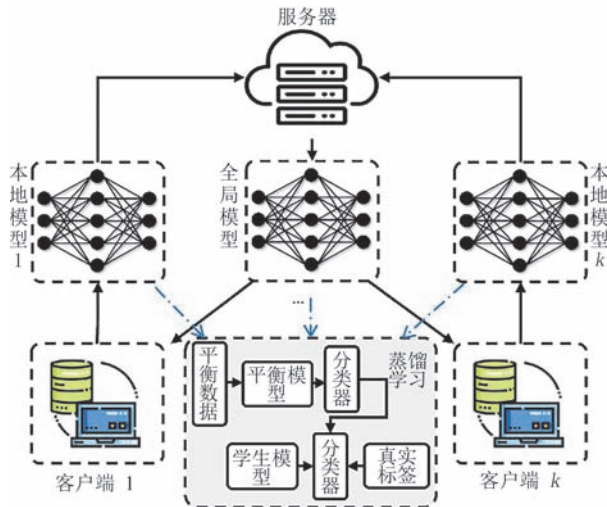


图10 知识蒸馏算法基本架构

其中, \mathbf{v} 和 \mathbf{x} 是分类器的参数和输入, i 和 j 为客户端的两个类别, $I_{i \neq j}$ 是输出类别 i 是否等于目标类别 j 的指标, β 是缩放边距值的超参数。在本地模型训练完成后, 客户端将权重、均值与方差上传至服务器。服务器利用这些均值与方差加权生成虚拟数据集 S , 并在自蒸馏模块^[70]中对分类器进行校准, 其校准损失函数如下:

$$\mathcal{H}(\mathbf{v}_{avg}, \tilde{\mathbf{v}}; S) = \mathcal{L}_{CE}(y_{cl}, y) + a \cdot (y_{cl}, \tilde{y}_{cl}) \quad (37)$$

其中, y_{cl} 与 \tilde{y}_{cl} 是平均全局分类器 \mathbf{v}_{avg} 和校准全局分类器 $\tilde{\mathbf{v}}$ 的输出。然而, 将客户端的均值与方差上传至服务器会泄露用户的数据分布, 且均值与方差生成的虚拟特征与真实特征存在较大差异。

不同于虚拟特征蒸馏校准, Shang 等人^[71]提出了一种服务器蒸馏集成校准算法 FedIC, 用于消除本地模型训练中的偏差。其核心思想是在服务器端存放少量真实的平衡辅助数据集 D_{aux} 作为先验知识, 通过固定特征提取器, 训练得到平衡的分类器作为教师模型, 对全局分类器进行蒸馏校准。然而, FedIC 仅关注分类器的校准, 忽视了中间层特征和关系的蒸馏, 这限制了学生模型的性能。

比服务器端的蒸馏校准算法, Elbatel 等人^[72]提出自监督蒸馏校准算法 Fed-MAS。具体而言, 客户端通过计算自监督预训练模型与全局模型之间的分布差异, 生成类感知的距离度量。然后加权平均该度量, 得到类感知的偏差估计, 并利用该估计进行蒸馏校准, 损失函数如下:

$$\mathcal{L}_{total} = \mathcal{L}_{sup} + \lambda_f \mathcal{L}_f \quad (38)$$

其中, \mathcal{L}_{sup} 表示原始监督损失, \mathcal{L}_f 为类感知距离。然而, Fed-MAS 的性能高度依赖于自监督预训练模型

的质量, 如果自监督模型无法捕捉长尾数据的本质特征, 蒸馏得到的学生模型可能表现不佳。

Wu 等人^[73]提出两阶段联邦长尾学习框架 FedNoRo。在第一阶段, 通过高斯混合模型和损失指标识别噪声客户端; 在第二阶段, 针对无噪声客户端使用标准交叉熵损失, 针对噪声客户端则通过软标签知识蒸馏训练本地模型。与硬标签相比, 软标签提供更多信息, 有助于模型更好地学习长尾数据中的有用特征, 其蒸馏损失函数为:

$$\mathcal{L}_{noise} = \mathcal{L}_{CE} + \lambda \mathcal{L}_{KL}(y_p, \text{softmax}(\frac{f_g(x)}{T})) \quad (39)$$

其中, T 是温度系数, $f_g(x)$ 是样本 x 的全局 logit 输出, y_p 表示本地模型的预测结果, \mathcal{L}_{KL} 为 KL 散度, λ 表示权衡系数。然而, 通过联邦平均聚合得到的软标签质量较低, 这可能影响客户端蒸馏的效果。

除了单独对服务器或客户端进行蒸馏校准外, Wu 等人^[74]提出一种联邦长尾双向蒸馏框架 FedICT。具体来说, 在客户端, 通过引入本地数据分布的先验知识, 加强模型对本地数据的拟合。在全局服务器蒸馏过程中, 通过余弦相似度为本地与全局知识分配权重, 使得传输的本地知识更好地匹配全局模型的类别平衡。然而, FedICT 中先验知识的引入往往带有一定的主观性, 可能导致不同的后验分布和推断结果。

3.2.4 平衡增强

平衡增强是应对长尾分布问题的主流方法之一。在联邦长尾学习中, 平衡增强策略通过本地数据增广和类别梯度平衡等技术, 针对尾部类别进行优化学习。其具体流程如图11所示。

Tang 等人^[75]提出联邦数据重采样算法 IWDS,

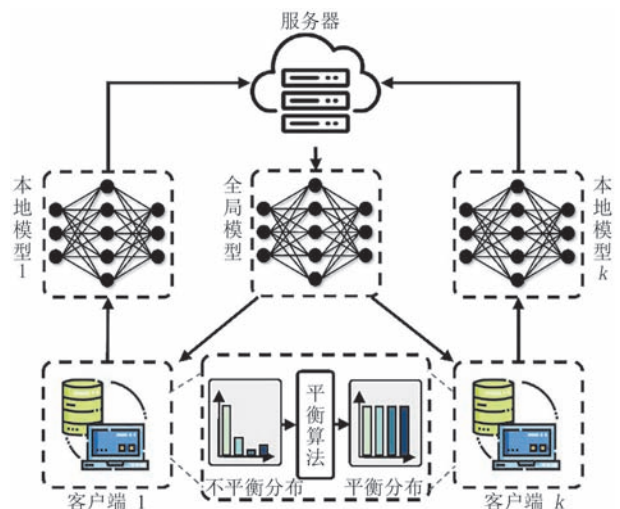


图11 平衡增强算法基本架构

通过在训练过程中逐步衰减尾部类采样权重,从而达到对于本地训练的平衡增强。具体来说,对于所有尾部样本,采样权重在训练初期较高,随着训练的进行,采样权重逐渐衰减。这种动态调整会使训练初期客户端的标签分布更加平衡。而在训练后期,通过减少重采样的影响,使得模型能够更好地学习到每个客户端特有的知识。整个训练过程中采样策略如下:

$$q(k,c)=\frac{N_{k,c}^\beta}{\sum_c N_{k,c}^\beta}$$

(40)

其中, $N_{k,c}$ 表示客户端 k 中类别 c 的样本数量, β 是一个衰减参数。虽然, IWDS 平衡了类别的样本数量,但不同类别的梯度可能仍然不平衡,结果同样会

对某些类别过于敏感。

不同于重采样平衡, Geet 等人^[76]提出联邦长尾过采样算法 FedSMOTE。具体来说,对于每一个少数类样本,随机选择一个邻居,然后在当前样本和选中的邻居之间随机插值,生成一个新的合成样本,直到少数类样本的数量达到期望的平衡状态。

Weinger 等人^[77]提出联邦组合数据增强算法 FedIoT。核心思想是采用随机过采样^[78]、合成少数类过采样^[79]、生成对抗网络^[80]和自适应合成^[81]对本地客户的数据进行增强平衡。然而,本地客户的样本短缺并不等同于全局的样本短缺。此外,当少数类样本中存在噪声或离群点时, FedSMOTE 和 FedIoT 算法可能会生成无效的合成样本。

表3 基于算法优化的联邦长尾学习算法对比

类型	算法名称	算法性能			应用场景	优化目标	
		准确性	计算开销	通信开销		本地客户	全局服务器
成本敏感学习	Fed-Focal Loss ^[53]	○	○	○	通用分类	✓	×
	FedAMC ^[55]	◐	◐	◐	物联网检测	✓	×
	Ratio Loss ^[56]	○	◐	◐	通用分类	✓	×
	FedRE ^[57]	◐	◐	○	通用分类	✓	✓
	Balance Loss ^[58]	○	○	◐	通用分类	✓	×
	FedAMM ^[59]	◐	●	●	通用分类	✓	✓
客户端选择	FedCIR ^[61]	◐	○	○	通用分类	×	✓
	FedSens ^[63]	◐	●	◐	边缘计算	×	✓
	FedCSS ^[65]	○	◐	◐	通用分类	×	✓
	Astraea ^[66]	○	◐	○	通用分类	✓	×
	GBME ^[67]	●	◐	◐	通用分类	✓	✓
知识蒸馏	FedCRAC ^[69]	●	◐	●	流量分类	×	✓
	FedIC ^[71]	◐	◐	●	通用分类	×	✓
	Fed-MAS ^[72]	◐	●	◐	医疗分类	✓	×
	FedNoRo ^[73]	●	◐	◐	通用分类	✓	✓
	FedICT ^[74]	●	◐	◐	边缘计算	✓	✓
平衡增强	IWDS ^[75]	○	●	○	通用分类	✓	×
	FedSMOTE ^[76]	○	●	○	金融预测	✓	×
	FedIoT ^[77]	○	●	○	物联网检测	✓	×
	ABC-GMM ^[82]	◐	◐	◐	通用分类	✓	✓
	FedGraB ^[83]	●	◐	◐	通用分类	✓	✓

Hahn 等人^[82]提出了一种隐私保护的联邦贝叶斯学习框架 ABC-GMM,用于解决临床数据不平衡分类的问题。具体来说,中心服务器通过与本地节点交换信息来估计生成模型参数的后验分布。本地客户不共享其本地数据,而是通过测量生成样本与本地数据之间的相似性来传递信息。这种方法在不显式共享或传输数据的情况下实现了分布式数据的分析,进而对尾部类别进行数据增强。然而基于近

似贝叶斯计算方法依赖于间接信息,可能导致与全局分类器之间的差距。

相较于通过数据生成器进行数据增广, Xiao 等人^[83]提出联邦长尾自调整梯度平衡框架 Fed-GraB。其基本思想包含两个关键部分,即直接先验分析和自梯度平衡。直接先验分析部分通过分析全局分类器的权重参数,生成一个先验向量 P_c ,并借助该向量对全局长尾先验进行学习。而自梯度平衡部分则

是以闭环方式,依据从全局长尾分布经过直接先验分析模块评估得来的反馈重新调整客户端的梯度。

表3比较了基于算法优化的代表方法在性能、应用场景和优化目标上的差异。其中,成本敏感学习方法尽管准确性较低,但计算和通信开销较小。为了寻找最佳成本,通常需要上传本地分布或平衡辅助数据集,这可能带来隐私泄露风险。客户端选择算法具有中等准确性,通信开销较低,但由于在本地训练模型之外还需额外训练选择策略,计算开销最大。蒸馏学习算法准确性较高,但为了计算和上

传类别梯度等信息,需要额外的计算和通信开销。平衡增强算法主要通过样本增广来优化,但由于全局少数类样本未知,盲目增广本地少数类样本可能进一步加剧全局不平衡。

3.3 联邦长尾学习算法对比总结

本节将联邦长尾学习算法根据改进策略分为两类:模型组件改进与基于算法的优化。针对不同联邦长尾学习算法进行对比,并分析其优缺点和适用场景。表4总结了各类联邦长尾学习算法的优劣势及适用场景,具体如下:

表 4 不同种类联邦长尾学习算法对比

长尾解决策略	类型	优点	缺点	适用场景
模型组件改进	虚拟特征重训练	能够有效提升全局识别准确率	服务器需要大量算力进行重训练,随着客户端增加,通信计算开销会大幅增加	适用于客户数量较少,服务器具有高计算能力,不考虑网络延迟的场景
	分类器校准	能够快速有效的提升全局识别精度,无需占用大量服务器算力	需要全局数据分布作为先验知识,隐私保护性低,需要额外的计算资源	适用于对模型准确率,泛化性较高的场景
	个性化联邦学习	能够适应不同客户端的数据分布,提高本地客户端上的性能	增加客户端的计算负担,全局模型的性能下降	适用于移动终端,医疗保健等用户行为和偏好高度个性化的场景
基于算法的优化	成本敏感学习	能够提升尾部类别精度,优化决策边界,无需额外计算资源	总体识别性能没有明显提升,需要全局数据分布作为先验知识,隐私保护性低	适用于低延迟,开销较小的场景
	客户端选择	能够减少必要的通信轮次,优化模型性能,减轻长尾分布影响	需要额外的通信和计算开销,需要复杂的优化策略,增加了系统设计的复杂度	适用于资源受限,实时性要求高,服务器具有高计算能力的环境
	知识蒸馏	能够有效提升全局识别准确率,算法鲁棒性较高	增加联邦学习的训练复杂性,需要额外的步骤来训练学生模型	适用于客户端,服务器具有高算力和存储空间场景
	平衡增强	能够平衡样本分布,提升本地模型精度	引发过拟合风险,并且需要消耗客户端额外的计算开销	适用于客户端具有高算力的场景

在模型组件改进策略中,本文探讨了虚拟特征重训练、分类器校准和个性化联邦学习三种方法。其中,虚拟特征重训练能够有效提高全局识别准确率,但其缺点是需要大量服务器计算资源进行重训练,且随着客户端数量增加,计算和通信开销大幅上升。此类算法适用于客户端较少、服务器计算能力强且无需考虑网络延迟的场景;分类器校准可快速提升全局识别精度,且不占用大量服务器算力,但依赖于全局数据分布作为先验知识,隐私保护较差,并需要额外的计算资源。适合对模型准确率和泛化性要求较高的场景;个性化联邦学习能适应不同客户端的数据分布,提高各客户端的模型表现,但会增加客户端的计算负担,可能导致全局模型性能下降。适用于用户行为和偏好高度个性化的场景,如移动

终端和医疗健康等领域。

在基于算法优化中,本文讨论了成本敏感学习、客户端选择、知识蒸馏和平衡增强四种方法。成本敏感学习能够提升尾部类别的精度,优化决策边界,且无需额外计算资源,但整体识别性能提升不显著,依赖全局数据分布,隐私保护性较弱。适合低延迟、开销小的场景;客户端选择可减少通信轮次,优化模型性能,缓解长尾分布影响,但需额外通信和计算开销,且设计复杂,适合资源受限、实时性高且服务器计算能力强的环境;知识蒸馏能有效提升全局识别准确率,且算法鲁棒性强,但增加了训练复杂性,需额外训练学生模型,适用于客户端和服务端均具备高计算能力和存储空间场景;平衡增强通过调整数据分布,缓解全局长尾带来的不平衡问题,显著提

高尾部类别的识别率,但可能引发过拟合风险,尤其在尾部类别数据量较少时。此外,平衡增强方法通常需要了解全局数据分布,依赖数据预处理,可能增加数据处理成本。

4 基准测试

本节首先介绍了不同领域中常用的长尾数据集,并概述了三种典型的长尾分布划分策略;随后,阐述联邦长尾学习算法常用的性能评价指标;最后对联联邦长尾学习算法的性能进行测试。

4.1 数据集

近年来,在执行联邦学习视觉识别任务时,主要

利用两种类型数据集进行长尾研究,并在表5进行了详细介绍。

第一种类型是通过合成操作修改的图像数据集的长尾版本,例如指数抽样 MNIST-LT^[84]、CIFAR10-LT^[85]、CIFAR100-LT^[86]、Fashion MNIST-LT^[87]数据集以及 Pareto 抽样的 ImageNet-LT^[88]、Places-LT 数据集^[89]。它们是从现有的平衡数据集中抽样形成的,长尾的程度可以通过任意不平衡率 (Imbalance Ratio, IR) 来控制,其中 $IR = \frac{\text{单类别最大值}}{\text{单类别最小值}}$ 。MNIST-LT 中训练样本 60 000-14 886 是指, MNIST 数据集有 60 000 个训练样本,采用不平衡率 $IR=100$ 的变体版本时的训练样本为 14 886。

表5 联邦长尾学习中开放数据集的总结

类型	数据集	训练样本数	类别	单类别最大值	单类别最小值
人工合成长尾数据集	MNIST-LT ^[84]	60 000-14 886	10	6000	60
	CIFAR10-LT ^[85]	50 000-12 406	10	5000	50
	CIFAR100-LT ^[86]	50 000-10 847	100	500	5
	FashionMNIST-LT ^[87]	60 000-14 886	10	6000	60
	ImageNet-LT ^[88]	115.8K	1000	1280	5
	Places-LT ^[89]	184.5K	365	4980	5
真实长尾数据集	iNaturalist17-LT ^[90]	57 9184	5089	2101	4
	iNaturalist18-LT ^[91]	43 7513	8142	2917	1
	Pascal VOC-MLT ^[92]	1142	20	775	4
	COCO-MLT ^[93]	1909	80	1128	6
	Lvis v0.5 ^[94]	57 000	1230	26 148	1
	Lvis v1 ^[95]	1 000 000	1203	125 481	10

第二种类型是使用具有高度非均衡标签分布的真实世界大规模数据集,这些数据集包括 iNaturalist 17/18 数据集^[90-91]、多标签分类的 Pascal VOC-MLT^[92]和 COCO-MLT^[93]数据集,以及物体检测的 Lvis v0.5/1 数据集^[94-95]。

4.2 长尾划分策略

在介绍数据集时提到联邦长尾学习中一类是将平衡数据集划分为长尾分布,进而进行模型训练。常用的长尾划分策略主要分为三种:指数划分^[96]、等差划分^[97]、Pareto 抽样^[98]。

(1) 指数划分是一种广泛应用的长尾数据划分策略。其核心特性就是借助指数函数对数据进行划分,从而形成符合指数分布的数据集。在这种划分策略下,较小的数据类别数量会比较多,而较大的数据类别数量则偏少。对于每个类别的图像数量 N_c , 公式如下:

$$N_c = N_{c, \max} \times \left(IR^{\frac{c_i}{C-1}} \right) \quad (41)$$

其中, $N_{c, \max}$ 是每个类别的最大图像数量, C 为类别总数, c_i 表示当前类别的索引号。

(2) 等差划分是另一种高效的长尾划分策略,其核心在于利用等差数列的原理来划分数据集,从而确保生成的数据集呈现出等差分布的特征。这种策略确保了每个数据类别的数量与前一个类别的数量之间存在一个恒定的差异。具体来说,每个类别的图像数量遵循以下分布公式:

$$N_c = N_{c, \max} - d \times c_i \quad (42)$$

其中,等差划分的步长 $d = \frac{(1-IR)N_{c, \max}}{C}$, 等差划分适合模拟类别间的差异明显的场景。

(3) Pareto 抽样是一种基于 Pareto 分布的数据划分策略。对于每一类数据,需要生成服从 Pareto

分布的随机数 X_c , 生成公式如下:

$$X_c = x_m(1 - U)^{-\frac{1}{\gamma}} \quad (43)$$

其中, γ 是分布的形状参数, 决定了分布的倾斜程度。 U 是一个服从均匀分布 $U(0, 1)$ 的随机数, x_m 是尺度参数, 代表分布可能的最小值。生成的随机数 X_c 可以看作是每一类数据在新的数据集中的比例。并可以通过以下公式计算每一类数据在新的数据集中的样本数量:

$$N_c = \lfloor N_{c, \max} \cdot X_c \rfloor \quad (44)$$

总的来说, 这三种划分策略都能够模拟现实世界中的数据分布情况, 但是每种策略适用的场景和模拟的现象是不同的。在进行模型训练时, 需要根据实际的需求和数据的特性来选择合适的划分策略。

4.3 评价指标

在联邦长尾学习中, 模型性能通常是通过不同的评价指标来进行判断的。这些评价指标涵盖了模型在处理大量数据时的准确度以及在处理少量数据时的表现, 以下是对这些评价指标的详细解释。

4.3.1 性能评价指标

准确度是指模型在所有样本中正确分类的比例, 通常用来衡量整体性能。针对于分类任务而言, 准确率的计算如下:

$$Acc = \frac{TP + TF}{TP + TN + FP + FN} \quad (45)$$

其中, TP 表示属于类别 C_i 且成功被分类为 C_i 的图像数量; TN 表示不属于类别 C_i 且没有被分配到 C_i 的图像数量; FP 表示被分类为 C_i 但不属于 C_i 的图像数量; FN 表示属于类别 C_i 但没有被分类为 C_i 的图像数量。

He^[99] 将准确率分解为多镜头准确率(类别图像超过 100 张)、中镜头准确率(类别图像在 20 至 100 张之间)和少镜头准确率(类别图像少于 20 张)。这种准确率的分解既可为特定需求提供更精确的评估, 同时也可揭示模型性能的瓶颈所在, 以便集中优化相应的任务类型。

在准确率的基础上, 联邦长尾学习还存在另外三个较少使用的度量分析方法^[100], 分别为精确度(Precision, P)、召回率(Recall Rate, R)和 F1-Score(F1)。三种方法用于衡量模型对正样本预测的能力, 以及评估模型在正样本识别与准确率两者间的综合预测能力, 相关的计算公式如下:

$$P = \frac{TP}{TP + FP} \quad (46)$$

$$R = \frac{TP}{TP + FN} \quad (47)$$

$$F1 = \frac{2 \times P \times R}{P + R} \quad (48)$$

在使用个性化联邦学习解决长尾问题时, 通常采用 PM(V) 和 PM(L) 作为性能评估指标^[101]。其中, PM(V) 表示测试集类别平衡下的个性化模型的平均准确率。PM(L) 则表示假设测试集和训练集具有相同分布条件下的个性化模型的平均测试准确率。

此外, 通信开销^[102]和计算开销^[103]也常用于衡量联邦学习在长尾数据上的表现。通信开销指的是设备或节点间为共享模型参数或梯度信息而消耗的网络资源和时间, 通常通过达到目标平均准确率所需的通信轮数与每轮传输的平均参数数量的乘积来量化。计算开销则指各节点进行模型训练所消耗的计算资源, 可通过达到目标准确率所需的通信轮数与每轮本地平均计算量的乘积来衡量。

4.3.2 长尾评价指标

多类不平衡程度(Multiclass Imbalance Degree, MID)是一种联邦长尾学习中用于度量全局类别不平衡程度的指标^[104]。它改进了传统的类别不平衡度量, 对数据集大小不敏感, 具体公式如下:

$$MID = \sum_{c=1}^C \frac{N_c}{N} \log_c \frac{C \cdot N_c}{N} \quad (49)$$

其中, N 代表所有样本, N_c 代表第 c 类样本, C 为类别数。多类不平衡程度越大, 全局数据分布越不平衡。

加权余弦相似度(Weighted Cosine Similarity, WCS)是一种用于衡量联邦学习中本地数据集和全局数据集类别不平衡差异的指标。它考虑了本地数据集的贡献, 并且能够更准确地评估本地和全局类别不平衡之间的关系, 具体公式如下:

$$WCS = \frac{1}{\|\mathbf{L}\|_1 \|\mathbf{L}\|_2} \sum_{i=1}^P \frac{\|\mathbf{l}_i\|_1}{\|\mathbf{l}_i\|_2} \mathbf{L} \cdot \mathbf{l}_i \quad (50)$$

其中, \mathbf{l}_i 为客户端 i 的标签分布向量, \mathbf{L} 表示所有客户端的标签分布的加权和。加权余弦相似度越大, 说明本地与全局类别不平衡的相似性越高。

4.4 对比实验

本节旨在评估联邦长尾学习算法的性能, 内容分为实验设置和主要结果两部分。

4.4.1 实验设置

实验使用按指数划分的 CIFAR10-LT 和 CIFAR100-LT 数据集,狄利克雷系数 α 设置为 0.5,训练模型采用 ResNet-8,客户端数量为 50,通信轮数设为 200。优化器选择随机梯度下降,初始学习率为 0.01,聚合时随机选取 60% 的客户端,每个客户端本地训练 5 轮。值得注意的是,部分对比算法包含特定的超参数设置,本文采用与原始论文一致的参数。此外,联邦长尾算法对客户端数据类别分布的敏感性极高,因此,为保证对比的公平性,本文对所有算法使用了相同的类别分布。

4.4.2 主要结果

为更直观地展示不同联邦学习算法的性能差异,本文在两种不平衡数据集上对代表性联邦长尾开源算法进行对比,结果见表 6。在类别较少的 CIFAR10-LT 数据集上,虚拟特征重训练和分类器校准方法表现优异。例如,CR_{EFF} 和 CLIP2FL 在 IR=100 时的准确率分别达到了 68.64% 和 70.22%;FedETC 和 FCA 的准确率分别为 69.64% 和 68.34%。这表明虚拟特征重训练通过强化特征表达,有效捕获了少数类别的信息,而分类器校准则通过调整分类边界,解决了类别不均衡问题,提升了长尾分布环境中的整体性能。

表 6 不同方法在 CIFAR10-LT 和 CIFAR100-LT 数据集上的性能对比

长尾解决策略	类型	方法	CIFAR10-LT			CIFAR100-LT		
			IR=10	IR=50	IR=100	IR=10	IR=50	IR=100
模型组件改进	虚拟特征重训练	CR _{EFF} ^[33]	78.55	71.64	68.64	46.11	35.94	32.18
		CLIP2FL ^[35]	79.16	72.21	70.22	47.62	38.64	35.87
	分类器校准	FedETC ^[41]	78.68	73.98	69.64	48.69	34.34	33.64
		FCA ^[42]	76.19	71.37	68.34	42.12	32.24	30.15
	个性化联邦学习	FedROD ^[45]	75.37	70.14	67.39	48.68	39.64	34.25
		FedNH ^[50]	75.48	70.36	68.37	49.36	38.24	36.51
基于算法优化	成本敏感学习	Ratio Loss ^[56]	78.97	68.99	65.12	48.52	33.25	28.67
		Balance Loss ^[58]	78.38	69.35	66.34	47.79	34.88	30.95
	客户端选择	FedCIR ^[61]	76.94	67.48	65.79	43.83	34.27	29.51
		Astraea ^[66]	75.38	68.37	62.67	45.52	33.57	31.58
	知识蒸馏	FedCRAC ^[69]	77.16	71.34	67.45	46.25	37.55	33.54
		FedIC ^[71]	76.37	70.66	66.64	44.52	36.54	32.10
	平衡增强	IWDS ^[75]	76.37	70.66	66.09	45.66	32.39	30.24
		FedGraB ^[83]	77.18	71.97	69.14	47.74	37.28	34.11

在类别较多的 CIFAR100-LT 数据集中,个性化联邦学习方法表现更为突出。例如,FedROD 和 FedNH 在 IR=100 时分别达到了 34.25% 和 36.51% 的准确率,显著优于其他算法。这得益于个性化联邦学习能够根据客户端的具体数据分布动态调整模型,从而缓解了类别增多带来的数据异质性问题,使其在跨客户端数据差异较大的场景下仍能保持较好的泛化性能。

此外,成本敏感学习方法 Ratio Loss 和 Balance Loss 在类别不均衡率 IR=10 时表现良好,准确率分别为 78.97% 和 78.38%,与其他领先方法相当。然而,随着不均衡程度的加剧,这些方法的性能明显下降。例如,在 IR=100 时,Ratio Loss 和 Balance Loss 的准确率分别下降到 65.12% 和 66.34%;在 CIFAR100-LT 数据集上,IR=100 时的准确率更是

降至 28.67% 和 30.95%。这种现象可能源于成本敏感学习试图通过调整不同类别的损失权重来平衡长尾分布,但在极端不均衡的情况下,损失权重的调整难以有效弥补少数类别的学习困难。

5 应用场景与未来展望

为了进一步推动联邦学习在真实环境中落地使用,解决联邦长尾问题已经成为一个重要的科学研究问题。本节结合现有工作,从方法优化和任务创新的角度确定了联邦长尾学习未来几个应用场景与未来研究方向,具体如图 12 所示。

5.1 联邦长尾学习应用场景

(1)在医疗领域,联邦学习的应用案例日益增多,其巨大的潜力逐渐被认可^[105]。例如,同济大学

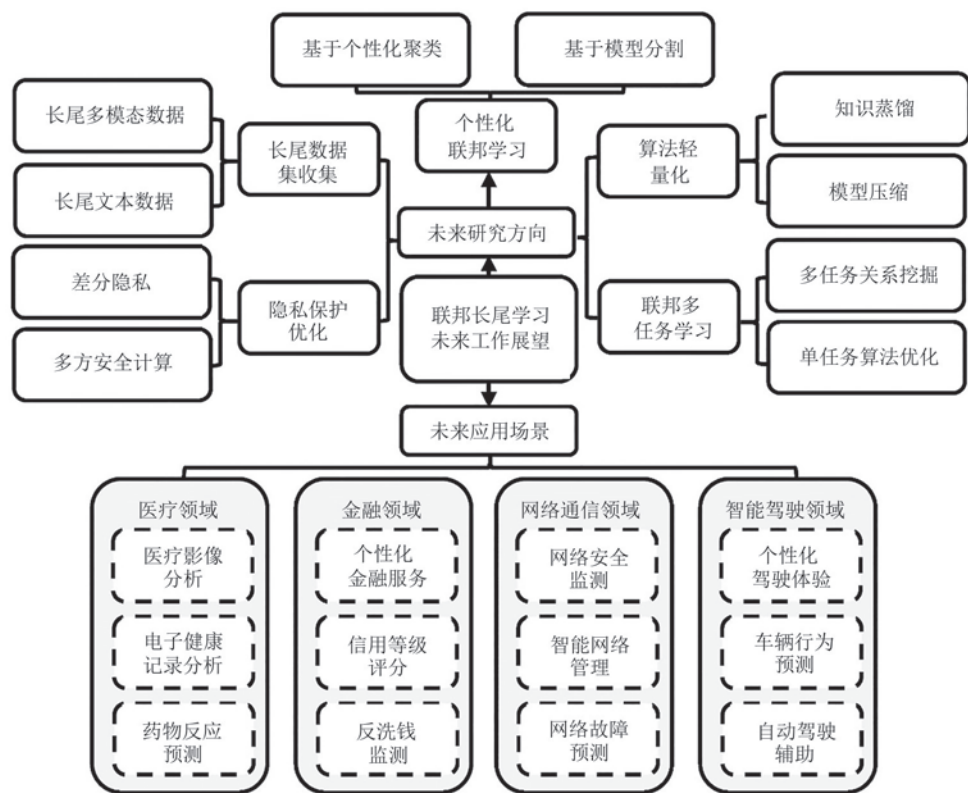


图 12 联邦长尾学习未来应用场景与研究方向展望

的刘琦课题组推出了协作药物发现平台 FL-QSAR,其基于联邦学习,有力推动了药物小分子领域的协同创新^[106]。同样,腾讯天衍实验室推出的医疗联邦学习 AI 工具成功提升脑卒中预测的精准度^[107]。此外,英特尔与宾夕法尼亚大学佩雷尔曼医学院联手开发了可以用于脑肿瘤检测的联邦智能模型^[108]。遗憾的是,现行诸多项目^[109]大多专注于处理常见疾病的监测难题,对于稀有或非典型疾病的诊断力仍显不足。这主要是因为医疗数据普遍呈现出长尾分布的特征,如何从有限的异常数据中提取有价值的信息,成为了提高罕见疾病诊断能力的关键研究点^[110]。对于未来的研究方向,应着眼于提升模型处理长尾数据的能力,例如运用稀疏样本的增强学习或生成模型,扩展稀有样本数据,从而提升模型的泛化性能。

(2)在金融领域,联邦学习开始被渐进式地运用到智能营销^[111]以及智能风控^[112]等环节,其间既发挥了数据的价值,也维护了用户隐私。以浦发银行构建的隐私计算平台^[113]和微众银行的信贷风控系统^[114]为例,这些系统均以联邦学习为基础,为企业信贷及个人贷款的风险管理改革提供了新颖的解决办法。然而,与医疗领域不同的是,金融领域中的

长尾分布问题主要展现在不同金融机构的数据差异性^[115],例如,银行拥有客户大量的交易数据,反观保险公司,则只拥有有限的健康记录。这种纵向长尾分布给数据对齐、隐私维护以及安全聚合策略设置带来诸多挑战^[116]。为解决上述问题,未来的研究方向应着眼于如何高效整合跨领域的数据资源,以构建安全且高效的多方数据计算模型,特别是在以联邦学习为背景的情境下,如何优化长尾训练方法显得尤为关键。

(3)在网络行业,得益于移动设备的广泛使用和互联网数据的爆炸式增长,联邦学习在智能推荐中的应用已取得显著进步^[117]。以字节跳动的 Fedlearner 平台为例,它已在电商、教育等众多领域得到实践,实现了广告个性化推送^[118]。另外,天津移动开发的联邦学习加区块链系统,为本地运营商的精准营销提供有效支持^[119]。虽然这些系统充分利用了联邦学习技术,但在实际运用中,它们往往忽视了用户行为数据时间长尾分布的现象,从而可能忽略了用户在不同时段内的偏好变化^[120]。因此,未来的研究重点应关注如何在时序场景中处理长尾分布问题,以便提升推荐系统在各个时间节点的精确度和适应性。

(4) 在智能驾驶领域,联邦学习的应用已逐步扩大。例如,百度的Apollo平台和特斯拉,它们都采用联邦学习技术对车辆与交通场景数据进行收集,以助力提升自动驾驶系统的感知和决策能力^[121]。然而,自动驾驶领域面临多模态数据的长尾分布问题^[122],如来自摄像头、雷达等不同传感器的数据,都呈现出非常显著的长尾特征。虽然常见的障碍物和行人对于识别相对容易,但在夜间低光环境下的小型障碍物识别,或是极端天气条件中行人的识别,现有模型的表现却不尽人意^[123]。因此,未来的研究应致力于处理多模态数据的长尾分布,以提升模型对复杂情境的感知能力^[124]。同时,在联邦学习环境中,如何实现多车辆数据的实时共享和高效训练,以确保自动驾驶系统的安全性和可靠性,也是当下亟待解决的挑战。

5.2 未来研究方向

目前,联邦学习中针对复杂长尾分布的研究尚处于初步阶段,仍有诸多问题亟待解决。以下五类问题值得进一步深入探讨:

(1) 长尾训练中的安全性研究。在解决联邦长尾问题时,许多算法在提升模型精度的过程中,要求客户端上传模型的类别梯度或少量真实数据,这可能在一定程度上泄露用户隐私^[125]。此外,攻击者可能通过分析模型的类别梯度推断用户的个人信息^[126],即便采用差分隐私技术,累积的噪声信息也可能被利用来推测用户隐私^[127]。因此,未来研究应重点关注模型性能与保障隐私间的平衡,探索新的方法或改进现有机制,更好地保护用户隐私。这一问题需要从数据收集、模型训练、结果解释等多个维度进行考量,可能还需结合机器学习、隐私保护、密码学、法律和伦理等多领域的知识。

(2) 多样化长尾分布的研究。目前的研究主要集中在现实世界中的长尾图像分类数据集,或是流行图像数据集的长尾版本上^[128]。然而,在非图像分类任务中,这些优化策略的效果通常不佳,主要原因是现有算法的改进措施多针对分类任务^[129]。因此,如何在多模态融合、时序预测、行为分析等真实长尾任务中,设计适用性广泛且具有鲁棒性的优化算法,成为一个重要挑战^[130]。未来的研究应不仅局限于图像分类领域,还应探讨如何设计更通用的损失函数和训练机制,以应对多模态数据融合及复杂任务场景的需求。

(3) 个性化联邦学习。目前,联邦长尾学习主要集中在提升全局模型性能,并保障隐私与全局分布

信息的获取。事实上,联邦长尾学习也可以根据用户的个性化需求,提供定制化的机器学习模型^[131]。不同用户的偏好和需求可能存在显著差异,为了满足这些差异,可以在训练过程中捕捉用户的个性化信息^[132]。在全局模型聚合后,通过本地数据微调模型,使其更符合用户的特定需求与兴趣^[133]。在此过程中,如何有效融合全局知识与本地个性化需求是关键挑战之一。此外,探索基于知识蒸馏和模型划分的个性化联邦方法与长尾学习的结合,也是未来值得关注的方向。

(4) 优化算法的轻量化。现有的联邦长尾算法往往需要客户端或服务器重新训练模型的类别梯度、分类器或客户端选择策略,导致额外的计算与存储开销^[134]。因此,这类算法不适用于工业传感器、智能手机、智能家居等计算资源有限的移动终端^[135]。如何在保证模型精度的前提下,构建轻量化的联邦长尾学习算法,成为亟待解决的核心问题^[136]。衡量轻量化程度需综合考虑信息留存量、模型复杂性、聚合模型的预期性能等多重因素。轻量化程度过高可能导致尾部类信息丢失,而算法过于复杂则会增加计算负担并影响全局模型性能。因此,未来研究可以结合模型、数据、通信等技术领域,设计高效轻量化方案,以降低通信与计算开销。

(5) 联邦多任务学习。在单任务学习中,长尾分布可能导致模型对少数类别的学习不足,影响整体性能^[137]。然而在多任务学习^[138]中,通过共享参数或隐藏层,不同任务间的关联性可以被挖掘,揭示低频任务与高频任务之间的关系,弥补长尾学习中的不足^[139]。因此,未来研究可聚焦于优化联邦多任务学习算法,结合联邦学习与多任务学习,设计出能够兼顾任务间关系与数据隐私需求的联邦多任务学习方法。

6 总 结

在数据使用过程中,由于隐私保护机制尚未完善,隐私泄露风险依然存在。联邦学习作为一种有效的隐私保护技术,能够在确保数据隐私的前提下实现数据共享。然而,全局数据的长尾分布与本地客户端的统计异质性共同作用,显著影响了联邦模型的稳定性与鲁棒性。基于这一背景,本文引入了联邦长尾学习,这是一个现实应用中普遍存在但研究较少的领域。本文首先概述了联邦长尾学习的概念,并提出了三种不同局部与全局长尾数据分布的

设置。此外,本文将解决方案分为模块组件改进和基于算法优化两大类,深入探讨了各种算法的优缺点。接着,本文介绍了常用的长尾数据集及其划分策略,概述了联邦长尾学习的评估指标,涵盖了模型性能和长尾效应的评价,并进行了对比分析。最后,本文讨论了联邦长尾学习在实际应用中的场景,并对未来的研究方向进行了展望。

致 谢 感谢《计算机学报》编辑和审稿专家,他们付出了辛勤的工作。

参 考 文 献

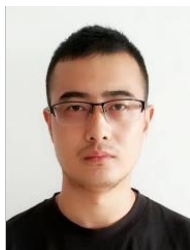
- [1] Wu F, Lu C, Zhu M, et al. Towards a new generation of artificial intelligence in China. *Nature Machine Intelligence*, 2020, 2(6): 312-316
- [2] Li Zhi-Peng, Guo Yong, Chen Yao-Fo, et al. Data generation based category balanced federated learning. *Chinese Journal of Computers*, 2023, 46(3): 609-625 (in Chinese)
(李志鹏, 国雍, 陈耀佛, 等. 基于数据生成的类别均衡联邦学习. *计算机学报*, 2023, 46(3): 609-625)
- [3] Zhang C, Xie Y, Bai H, et al. A survey on federated learning. *Knowledge-Based Systems*, 2021, 216: 106775
- [4] Xiao Xiong, Tang Zhuo, Xiao Bin, Li Ken-Li. A survey on privacy and security issues in federated learning. *Chinese Journal of Computers*, 2023, 46(5): 1019-1044 (in Chinese)
(肖雄, 唐卓, 肖斌, 李肯立. 联邦学习的隐私保护与安全防御研究综述. *计算机学报*, 2023, 46(5): 1019-1044)
- [5] Li L, Fan Y, Tse M, et al. A review of applications in federated learning. *Computers & Industrial Engineering*, 2020, 149: 106854
- [6] Wang J, Liu Q, Liang H, et al. A novel framework for the analysis and design of heterogeneous federated learning. *IEEE Transactions on Signal Processing*, 2021, 69: 5234-5249
- [7] Ma X, Zhu J, Lin Z, et al. A state-of-the-art survey on solving non-iid data in federated learning. *Future Generation Computer Systems*, 2022, 135: 244-258
- [8] Wang Z, Li H, Li J, et al. Federated learning on non-IID and long-tailed data via dual-decoupling. *Frontiers of Information Technology & Electronic Engineering*, 2024, 25(5): 728-741
- [9] Zhu G, Liu X, Niu J, et al. Learning by imitating the classics: Mitigating class imbalance in federated learning via simulated centralized learning. *Expert Systems with Applications*, 2024, 255: 124755
- [10] Zhang L, Wu Y, Chen L, et al. Scoring aided federated learning on long-tailed data for wireless iot based healthcare system. *IEEE Journal of Biomedical and Health Informatics*, 2023, 28(6): 3341-3348
- [11] Zhang J, Li C, Qi J, et al. A survey on class imbalance in federated learning. *arXiv preprint arXiv:2303.11673*, 2023
- [12] Li K, Li Y, Zhang J, et al. Federated deep long-tailed learning: A survey. *Neurocomputing*, 2024, 595: 127906
- [13] Chen Z, Liu S, Wang H, et al. Towards federated long-tailed learning. *arXiv preprint arXiv:2206.14988*, 2022
- [14] Amari S. Backpropagation and stochastic gradient descent method. *Neurocomputing*, 1993, 5(4-5): 185-196
- [15] Ye M, Fang X, Du B, et al. Heterogeneous federated learning: State-of-the-art and research challenges. *ACM Computing Surveys*, 2023, 56(3): 1-44
- [16] Vahidian S, Morafah M, Shah M, et al. Rethinking data heterogeneity in federated learning: Introducing a new notion and standard benchmarks. *IEEE Transactions on Artificial Intelligence*, 2023, 5(3): 1386-1397
- [17] Liu Y, Kang Y, Zou T, et al. Vertical federated learning: Concepts, advances, and challenges. *IEEE Transactions on Knowledge and Data Engineering*, 2024, 36(7): 3615-3634
- [18] Sheng T, Shen C, Liu Y, et al. Modeling global distribution for federated learning with label distribution skew. *Pattern Recognition*, 2023, 143: 109724
- [19] 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
- [20] Li Q, Diao Y, Chen Q, et al. Federated learning on non-iid data silos: An experimental study// *Proceedings of the 38th IEEE International Conference on Data Engineering*. Kuala Lumpur, Malaysia, 2022: 965-978
- [21] Yurochkin M, Agarwal M, Ghosh S, et al. Bayesian nonparametric federated learning of neural networks// *Proceedings of the 36th International Conference on Machine Learning*. Long Beach, USA, 2019, 97: 7252-7261
- [22] Zhou T, Zhang J, Tsang D H K. FedFA: Federated learning with feature anchors to align features and classifiers for heterogeneous data. *IEEE Transactions on Mobile Computing*, 2024, 23(6): 6731-6742
- [23] Fraboni Y, Vidal R, Kamení L, et al. Clustered sampling: Low-variance and improved representativity for clients selection in federated learning// *Proceedings of the 38th International Conference on Machine Learning*. Vienna, Austria, 2021: 3407-3416
- [24] Zhang Y, Kang B, Hooi B, et al. Deep long-tailed learning: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023, 45(9): 10795-10816
- [25] Yang L, Jiang H, Song Q, et al. A survey on long-tailed visual recognition. *International Journal of Computer Vision*, 2022, 130(7): 1837-1872
- [26] Chang N, Yu Z, Wang Y X, et al. Image-level or object-level? A tale of two resampling strategies for long-tailed detection// *Proceedings of the 38th International Conference on Machine Learning*. Online, 2021: 1463-1472
- [27] Li M, Cheung Y M, Hu Z. Key point sensitive loss for long-tailed visual recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022, 45(4): 4812-4825
- [28] Deng K, Cheng G, Yang R, et al. Alleviating asr long-tailed

- problem by decoupling the learning of representation and classification. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 2021, 30: 340-354
- [29] Zhang M L, Zhang X Y, Wang C, et al. Towards prior gap and representation gap for long-tailed recognition. *Pattern Recognition*, 2023, 133: 109012
- [30] Liu Z, Chen Y, Zhao Y, et al. Contribution-aware federated learning for smart healthcare//*Proceedings of the 36th AAAI Conference on Artificial Intelligence*. Online, 2022, 36 (11): 12396-12404
- [31] Luo M, Chen F, Hu D, et al. No fear of heterogeneity: Classifier calibration for federated learning with non-iid data//*Proceedings of the 35th Conference on Neural Information Processing Systems*. Online, 2021, 8: 5972-5984
- [32] Chen Z, Wu Z, Ji Y. Decoupled federated learning on long-tailed and non-IID data with feature statistics. *arXiv preprint arXiv:2403.08364*, 2024
- [33] Shang X, Lu Y, Huang G, et al. Federated learning on heterogeneous and long-tailed data via classifier re-training with federated features//*Proceedings of the 31st International Joint Conference on Artificial Intelligence*. Montreal, Canada, 2022: 2218-2224
- [34] Zhou Y, Wang J, Kong X, et al. Exploring amplified heterogeneity arising from heavy-tailed distributions in federated learning. *IEEE Transactions on Mobile Computing*, 2024, 23(12): 11519-11534
- [35] Shi J, Zheng S, Yin X, et al. CLIP-Guided Federated learning on heterogeneity and long-tailed data//*Proceedings of the 38th AAAI Conference on Artificial Intelligence*. Vancouver, Canada, 2024: 14955-14963
- [36] Radford A, Kim J W, Hallacy C, et al. Learning transferable visual models from natural language supervision//*Proceedings of the 38th International Conference on Machine Learning*. Online, 2021: 8748-8763
- [37] Huang W, Liu Y, Ye M, et al. Federated learning with long-tailed data via representation unification and classifier rectification. *IEEE Transactions on Information Forensics and Security*, 2024, 19: 5738-5750
- [38] Yang W, Chen D, Zhou H, et al. Integrating local real data with global gradient prototypes for classifier re-balancing in federated long-tailed learning. *arXiv preprint arXiv: 2301.10394*, 2023
- [39] Zhang J, Li Z, Li B, et al. Federated learning with label distribution skew via logits calibration//*Proceedings of the 39th International Conference on Machine Learning*. Baltimore, USA, 2022: 26311-26329
- [40] Zheng W, Yan L, Gou C, et al. Federated meta-learning for fraudulent credit card detection//*Proceedings of the 29th International Conference on International Joint Conferences on Artificial Intelligence*. Montreal, Canada, 2021: 4654-4660
- [41] Li Z, Shang X, He R, et al. No fear of classifier biases: Neural collapse inspired federated learning with synthetic and fixed classifier//*Proceedings of the IEEE/CVF International Conference on Computer Vision*. Paris, France, 2023: 5319-5329
- [42] Papayan V, Han X Y, Donoho D L. Prevalence of neural collapse during the terminal phase of deep learning training. *Proceedings of the National Academy of Sciences*, 2020, 117(40): 24652-24663
- [43] Wicaksana J, Yan Z, Cheng K T. FCA: Taming long-tailed federated medical image classification by classifier anchoring. *arXiv preprint arXiv:2305.00738*, 2023
- [44] Lu Y, Qian P, Huang G, et al. Personalized federated learning on long-tailed data via adversarial feature augmentation//*Proceedings of the 48th IEEE International Conference on Acoustics, Speech and Signal Processing*. Rhodes Island, Greece, 2023: 1-5
- [45] Chen H Y, Chao W L. On bridging generic and personalized federated learning for image classification//*Proceedings of the 10th International Conference on Learning Representations*. Online, 2022: 25-29
- [46] Xiao Z, Chen Z, Liu L, et al. FedLoGe: Joint local and generic federated learning under long-tailed data. *arXiv preprint arXiv: 2401.08977*, 2024
- [47] Xie L, Yang Y, Cai D, et al. Neural collapse inspired attraction-repulsion-balanced loss for imbalanced learning. *Neurocomputing*, 2023, 527: 60-70
- [48] Chen Z, Yang C, Zhu M, et al. Personalized retrogress-resilient federated learning toward imbalanced medical data. *IEEE Transactions on Medical Imaging*, 2022, 41(12): 3663-3674
- [49] Guo S, Wang H, Lin S, et al. Addressing skewed heterogeneity via federated prototype rectification with personalization. *IEEE Transactions on Neural Networks and Learning Systems*, 2024
- [50] Dai Y, Chen Z, Li J, et al. Tackling data heterogeneity in federated learning with class prototypes//*Proceedings of the 37th AAAI Conference on Artificial Intelligence*. Washington, USA, 2023: 7314-7322
- [51] Zhang R, Chen Y, Wu C, et al. Multi-level personalized federated learning on heterogeneous and long-tailed data. *IEEE Transactions on Mobile Computing*, 2024, 23 (12): 12396-12409
- [52] Zhou Z H, Liu X Y. On multi-class cost-sensitive learning. *Computational Intelligence*, 2010, 26(3): 232-257
- [53] Sarkar D, Narang A, Rai S. Fed-focal loss for imbalanced data classification in federated learning. *arXiv preprint arXiv: 2011.06283*, 2020
- [54] Lin T Y, Goyal P, Girshick R, et al. Focal loss for dense object detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020, 42(2): 318-327
- [55] Wang Y, Gui G, Gacanian H, et al. Federated learning for automatic modulation classification under class imbalance and varying noise condition. *IEEE Transactions on Cognitive Communications and Networking*, 2021, 8(1): 86-96
- [56] Wang L, Xu S, Wang X, et al. Addressing class imbalance in federated learning//*Proceedings of the 35th AAAI Conference on Artificial Intelligence*. Online, 2021, 35(11): 10165-10173

- [57] Cheng X, Shi F, Liu Y, et al. A class-imbalanced heterogeneous federated learning model for detecting icing on wind turbine blades. *IEEE Transactions on Industrial Informatics*, 2022, 18(12): 8487-8497
- [58] Shuai X, Shen Y, Jiang S, et al. Balancefl: Addressing class imbalance in long-tail federated learning//*Proceedings of the 21st ACM/IEEE International Conference on Information Processing in Sensor Networks*. Online, Italy 2022: 271-284
- [59] Qian P, Lu Y, Wang H. Long-Tailed federated learning via aggregated meta mapping//*Proceedings of the 30th IEEE International Conference on Image Processing*. Kuala Lumpur, Malaysia, 2023: 2010-2014
- [60] Huang Z, Shen X, Xing J, et al. Revisiting knowledge distillation: An inheritance and exploration framework//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Nashville, USA, 2021: 3579-3588
- [61] Yang M, Wang X, Zhu H, et al. Federated learning with class imbalance reduction//*Proceedings of the 2021 29th European Signal Processing Conference (EUSIPCO)*. Dublin, Ireland, 2021: 2174-2178
- [62] Vermorel J, Mohri M. Multi-armed bandit algorithms and empirical evaluation//*Proceedings of the European Conference on Machine Learning*. Berlin, Germany, 2005: 437-448
- [63] Zhang D Y, Kou Z, Wang D. Fedsens: A federated learning approach for smart health sensing with class imbalance in resource constrained edge computing//*Proceedings of the IEEE INFOCOM 2021-IEEE Conference on Computer Communications*. Vancouver, Canada, 2021: 1-10
- [64] Pathak D, Agrawal P, Efros A A, et al. Curiosity-driven exploration by self-supervised prediction//*Proceedings of the 34th International Conference on Machine Learning*. Sydney, Australia, 2017: 2778-2787
- [65] Chen B, Zheng X, Zhu Y, et al. A novel client sampling scheme for unbalanced data distribution under federated learning//*Proceedings of the International Conference on Smart Computing and Communication*. New York, USA, 2021: 438-449
- [66] Duan M, Liu D, Chen X, et al. Self-balancing federated learning with global imbalanced data in mobile systems. *IEEE Transactions on Parallel and Distributed Systems*, 2020, 32(1): 59-71
- [67] Zeng Y, Liu L, Liu L, et al. Global balanced experts for federated long-tailed learning//*Proceedings of the IEEE/CVF International Conference on Computer Vision*. Paris, France, 2023: 4815-4825
- [68] Li L, Gou J, Yu B, et al. Federated distillation: A Survey. *arXiv preprint arXiv:2404.08564*, 2024
- [69] Lv F, Shang X, Zhou Y, et al. Personalized federated learning on heterogeneous and long-tailed data via expert collaborative learning. *arXiv preprint arXiv:2408.02019*, 2024
- [70] Liu W, Zhou P, Zhao Z, et al. Fastbert: A self-distilling bert with adaptive inference time//*Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. Online, USA, 2020: 6035-6044
- [71] Shang X, Lu Y, Cheung Y, et al. Fedic: Federated learning on non-iid and long-tailed data via calibrated distillation//*Proceedings of the 2022 IEEE International Conference on Multimedia and Expo (ICME)*. Taipei, China, 2022: 1-6
- [72] Elbatel M, Wang H, Mart R, et al. Federated model aggregation via self-supervised priors for highly imbalanced medical image classification//*Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention*. Vancouver, Canada, 2023: 334-346
- [73] Wu N, Yu L, Jiang X, et al. FedNoRo: Towards noise-robust federated learning by addressing class imbalance and label noise heterogeneity //*Proceedings of the 32nd International Joint Conference on Artificial Intelligence*. Paris, France, 2023: 2218-2224
- [74] Wu Z, Sun S, Wang Y, et al. Fedict: Federated multi-task distillation for multi-access edge computing. *IEEE Transactions on Parallel and Distributed Systems*, 2023, 35(6): 1107-1121
- [75] Tang Z, Hu Z, Shi S, et al. Data resampling for federated learning with non-iid labels//*Proceedings of the 32nd International Joint Conference on Artificial Intelligence*. Paris, France, 2023: 2218-2224
- [76] Shingi G. A federated learning based approach for loan defaults prediction//*Proceedings of the 20th IEEE International Conference on Data Mining Workshops*. Sorrento, Italy, 2020: 362-368
- [77] Weinger B, Kim J, Sim A, et al. Enhancing IoT anomaly detection performance for federated learning. *Digital Communications and Networks*, 2022, 8(3): 314-323
- [78] Wei G, Mu W, Song Y, et al. An improved and random synthetic minority oversampling technique for imbalanced data. *Knowledge-Based Systems*, 2022, 248: 108839
- [79] Zhu T, Lin Y, Liu Y. Synthetic minority oversampling technique for multiclass imbalance problems. *Pattern Recognition*, 2017, 72: 327-340
- [80] Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial networks. *Communications of the ACM*, 2020, 63(11): 139-144
- [81] Jiang L, Dai B, Wu W, et al. Deceive d: Adaptive pseudo augmentation for gan training with limited data. *Advances in Neural Information Processing Systems*, 2021, 34: 21655-21667
- [82] Hahn S J, Lee J. Privacy-preserving federated bayesian learning of a generative model for imbalanced classification of clinical data. *arXiv preprint arXiv:1910.08489*, 2019
- [83] Xiao Z, Chen Z, Liu S, et al. Fed-grab: Federated long-tailed learning with self-adjusting gradient balancer. *Advances in Neural Information Processing Systems*, 2024, 36
- [84] Chen J, Su B. Transfer knowledge from head to tail: Uncertainty calibration under long-tailed distribution//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Vancouver, Canada, 2023: 19978-19987
- [85] Du F, Yang P, Jia Q, et al. Global and local mixture consistency cumulative learning for long-tailed visual recognitions//*Proceedings of the IEEE/CVF Conference on*

- Computer Vision and Pattern Recognition. Vancouver, Canada, 2023: 15814-15823
- [86] Cui J, Liu S, Tian Z, et al. Reslt: Residual learning for long-tailed recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022, 45(3): 3695-3706
- [87] Shi J X, Wei T, Xiang Y, et al. How re-sampling helps for long-tail learning? *Advances in Neural Information Processing Systems*, 2023, 36
- [88] Zhong Z, Cui J, Liu S, et al. Improving calibration for long-tailed recognition//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Nashville, USA, 2021: 16489-16498
- [89] Wang Y, Yu Z, Wang J, et al. Exploring vision-language models for imbalanced learning. *International Journal of Computer Vision*, 2024, 132(1): 224-237
- [90] Van Horn G, Mac Aodha O, Song Y, et al. The inaturalist species classification and detection dataset//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Salt Lake City, USA, 2018: 8769-8778
- [91] Koo K S, Oh J M, Park S J, et al. Accessing the accuracy of citizen science data based on iNaturalist data. *Diversity*, 2022, 14(5): 316
- [92] Tong K, Wu Y. Rethinking PASCAL-VOC and MS-COCO dataset for small object detection. *Journal of Visual Communication and Image Representation*, 2023, 93: 103830
- [93] Liang C, Yang Z, Zhu L, et al. Co-learning meets stitch-up for noisy multi-label visual recognition. *IEEE Transactions on Image Processing*, 2023, 32: 2508-2519
- [94] Gupta A, Dollar P, Girshick R. Lvis: A dataset for large vocabulary instance segmentation//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Long Beach, USA, 2019: 5356-5364
- [95] Tan J, Zhang G, Deng H, et al. 1st place solution of lvis challenge 2020: A good box is not a guarantee of a good mask. *arXiv preprint arXiv:2009.01559*, 2020
- [96] Qin J. A Survey of long-tail item recommendation methods. *Wireless Communications and Mobile Computing*, 2021, 2021(1): 7536316
- [97] Tuck A F. From molecules to meteorology via turbulent scale invariance. *Quarterly Journal of the Royal Meteorological Society*, 2010, 136(650): 1125-1144
- [98] Wang J, Zhou H, Zhou M, et al. A general model for long-tailed network traffic approximation. *The Journal of Supercomputing*, 2006, 38: 155-172
- [99] He Y Y, Wu J, Wei X S. Distilling virtual examples for long-tailed recognition//*Proceedings of the IEEE/CVF International Conference on Computer Vision*. Montreal, Canada, 2021: 235-244
- [100] Liu W, Mo J, Zhong F. Class imbalanced medical image classification based on semi-supervised federated learning. *Applied Sciences*, 2023, 13(4): 2109
- [101] Tan A Z, Yu H, Cui L, et al. Towards personalized federated learning. *IEEE Transactions on Neural Networks and Learning Systems*, 2022, 34(12): 9587-9603
- [102] Yi L, Yu H, Ren C, et al. FedMoE: Data-level personalization with mixture of experts for model-heterogeneous personalized federated learning. *arXiv preprint arXiv:2402.01350*, 2024
- [103] Ma X, Zhang J, Guo S, et al. Layer-wised model aggregation for personalized federated learning//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. New Orleans, USA, 2022: 10092-10101
- [104] Xiao C, Wang S. An experimental study of class imbalance in federated learning//*Proceedings of the 2021 IEEE Symposium Series on Computational Intelligence (SSCI)*. Orlando, USA, 2021: 1-7
- [105] Antunes R S, André da Costa C, Küderle A, et al. Federated learning for healthcare: Systematic review and architecture proposal. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 2022, 13(4): 1-23
- [106] Nguyen D C, Pham Q V, Pathirana P N, et al. Federated learning for smart healthcare: A survey. *ACM Computing Surveys (Csur)*, 2022, 55(3): 1-37
- [107] Xu J, Glicksberg B S, Su C, et al. Federated learning for healthcare informatics. *Journal of Healthcare Informatics Research*, 2021, 5: 1-19
- [108] Rahman A, Hossain M S, Muhammad G, et al. Federated learning-based AI approaches in smart healthcare: Concepts, taxonomies, challenges and open issues. *Cluster Computing*, 2023, 26(4): 2271-2311
- [109] Elhussein A, Savarraj J, Connolly E S, et al. A generalizable physiological model for detection of delayed cerebral ischemia using federated learning//*Proceedings of the 2023 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. Istanbul, Turkey, 2023: 1886-1889
- [110] Rieke N, Hancox J, Li W, et al. The future of digital health with federated learning. *NPJ Digital Medicine*, 2020, 3(1): 1-7
- [111] Wen J, Zhang Z, Lan Y, et al. A survey on federated learning: Challenges and applications. *International Journal of Machine Learning and Cybernetics*, 2023, 14(2): 513-535
- [112] Chatterjee P, Das D, Rawat D B. Federated learning empowered recommendation model for financial consumer services. *IEEE Transactions on Consumer Electronics*, 2023, 70(1): 2508-2516
- [113] Zheng Z, Zhou Y, Sun Y, et al. Applications of federated learning in smart cities: Recent advances, taxonomy, and open challenges. *Connection Science*, 2022, 34(1): 1-28
- [114] Dash B, Sharma P, Ali A. Federated learning for privacy-preserving: A review of PII data analysis in Fintech. *International Journal of Software Engineering Applications (IJSEA)*, 2022, 13(4)
- [115] Yu H, Liu Z, Liu Y, et al. A sustainable incentive scheme for federated learning. *IEEE Intelligent Systems*, 2020, 35(4): 58-69
- [116] Alazab M, RM S P, Parimala M, et al. Federated learning for cybersecurity: Concepts, challenges, and future directions. *IEEE Transactions on Industrial Informatics*, 2021, 18(5): 3501-3509
- [117] Liu Y, James J Q, Kang J, et al. Privacy-preserving traffic flow

- prediction: A federated learning approach. *IEEE Internet of Things Journal*, 2020, 7(8): 7751-7763
- [118] Zhang C, Dang S, Shihada B, et al. Dual attention-based federated learning for wireless traffic prediction//*Proceedings of the IEEE INFOCOM 2021-IEEE Conference on Computer Communications*. Vancouver, Canada, 2021: 1-10
- [119] Liu Y, Zhang L, Liu Y, et al. Model maturity-based model service composition in cloud environments. *Simulation Modelling Practice and Theory*, 2021, 113: 102389
- [120] Yang Z, Chen M, Wong K K, et al. Federated learning for 6G: Applications, challenges, and opportunities. *Engineering*, 2022, 8: 33-41
- [121] Nguyen D C, Ding M, Pathirana P N, et al. Federated learning for internet of things: A comprehensive survey. *IEEE Communications Surveys & Tutorials*, 2021, 23(3): 1622-1658
- [122] Du Z, Wu C, Yoshinaga T, et al. Federated learning for vehicular internet of things: Recent advances and open issues. *IEEE Open Journal of the Computer Society*, 2020, 1: 45-61
- [123] Zhang T, Gao L, He C, et al. Federated learning for the internet of things: Applications, challenges, and opportunities. *IEEE Internet of Things Magazine*, 2022, 5(1): 24-29
- [124] Imteaj A, Thakker U, Wang S, et al. A survey on federated learning for resource-constrained IoT devices. *IEEE Internet of Things Journal*, 2021, 9(1): 1-24
- [125] Li Q, Wen Z, Wu Z, et al. A survey on federated learning systems: Vision, hype and reality for data privacy and protection. *IEEE Transactions on Knowledge and Data Engineering*, 2021, 35(4): 3347-3366
- [126] Wei K, Li J, Ding M, et al. Federated learning with differential privacy: Algorithms and performance analysis. *IEEE Transactions on Information Forensics and Security*, 2020, 15: 3454-3469
- [127] Ma C, Li J, Ding M, et al. On safeguarding privacy and security in the framework of federated learning. *IEEE Network*, 2020, 34(4): 242-248
- [128] Fu Y, Xiang L, Zahid Y, et al. Long-tailed visual recognition with deep models: A methodological survey and evaluation. *Neurocomputing*, 2022, 509: 290-309
- [129] Gadre S Y, Ilharco G, Fang A, et al. Datacomp: In search of the next generation of multimodal datasets. *Advances in Neural Information Processing Systems*, 2024, 36
- [130] Birhane A, Prabhu V U, Kahembwe E. Multimodal datasets: Misogyny, pornography, and malignant stereotypes. *arXiv preprint arXiv:2110.01963*, 2021
- [131] Collins L, Hassani H, Mokhtari A, et al. Exploiting shared representations for personalized federated learning//*Proceedings of the 38th International Conference on Machine Learning*. Online, 2021: 2089-2099
- [132] Oh J, Kim S, Yun S Y. Fedbabu: Towards enhanced representation for federated image classification. *arXiv preprint arXiv:2106.06042*, 2021
- [133] Zhang J, Hua Y, Cao J, et al. Eliminating domain bias for federated learning in representation space//*Proceedings of the 37th Conference on Neural Information Processing Systems*. New Orleans, USA, 2023(36):14204--14227
- [134] So J, He C, Yang C S, et al. Lightsecagg: Aa lightweight and versatile design for secure aggregation in federated learning. *Proceedings of Machine Learning and Systems*, 2022, 4: 694-720
- [135] Wei Z, Pei Q, Zhang N, et al. Lightweight federated learning for large-scale IoT devices with privacy guarantee. *IEEE Internet of Things Journal*, 2021, 10(4): 3179-3191
- [136] Zhang Z, Wu L, Ma C, et al. LSFL: A lightweight and secure federated learning scheme for edge computing. *IEEE Transactions on Information Forensics and Security*, 2022, 18: 365-379
- [137] Marfoq O, Neglia G, Bellet A, et al. Federated multi-task learning under a mixture of distributions//*Proceedings of the 35th Conference on Neural Information Processing Systems*. Online, 2021: 15434-15447
- [138] Smith V, Chiang C K, Sanjabi M, et al. Federated multi-task learning//*Proceedings of the Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017*. Long Beach, USA, 2017: 4424--4434
- [139] Mills J, Hu J, Min G. Multi-task federated learning for personalised deep neural networks in edge computing. *IEEE Transactions on Parallel and Distributed Systems*, 2021, 33(3): 630-641



ZHOU Yi-Zhi, Ph. D. candidate. His main research interests include edge computing and federated learning.

WANG Jun-Xiao, Ph. D., associate professor. His main research interests include

foundation model and federated learning.

XIE Xin, Ph. D., associate professor. His

main research interests include wireless sensing, RFID, and federated learning.

WANG Peng-Fei, Ph. D., associate professor. His main research interests include distributed artificial intelligence, federated learning and big data intelligent processing.

JIA Xi-Bei, Ph. D., senior engineer. His main research interests include distributed artificial intelligence and data mining.

QI Heng, Ph. D. , professor. His main research interests include next generation internet system, data center networks, and federated learning.

QIN Yu-Chen, Ph. D. His main research interests include big data mining and federated learning.

Background

The problem examined in this paper falls within the domain of machine learning, specifically under the umbrella of federated learning and long-tail distribution. Federated learning, a decentralized approach to machine learning, and long-tail learning, the study of imbalanced data distribution, are two vital topics in the current international research landscape.

The issue of long-tail learning in a federated setting is an emerging problem, with partial solutions currently available. The primary challenge lies in dealing with the skewed distribution of data in a decentralized manner. While conventional approaches such as data classifier calibration and cost-sensitive learning have been adapted for this purpose, they often do not perform optimally in a federated setting due to privacy concerns and the non-IID nature of the data.

This paper presents an in-depth review of the current state of federated long-tail learning, aiming to provide a comprehensive understanding of the current solutions and their limitations.

Furthermore, we propose novel perspectives and potential pathways to tackle this problem more effectively, thus pushing the boundaries of the current understanding and solutions.

This work was supported by the National Natural Science Foundation of China under Grant No. 62072069. The project's significance lies in its potential to revolutionize machine learning applications in various fields, from healthcare to finance, where data privacy is paramount and global data distribution is often imbalanced.

Our research group has a robust background in federated learning and long-tail learning. We have previously published several works in these areas, including “Exploring Amplified Heterogeneity Arising from Heavy-Tailed Distributions in Federated Learning” , ” Rethinking Personalized Client Collaboration in Federated Learning” etc. These prior works have significantly contributed to the understanding of the challenges and potential solutions in these areas, laying a solid foundation for the current study.