

结合旋转自监督和 CLIP 指导的长尾数据联邦学习

刘海军¹, 付晓东^{1,2}

(1. 昆明理工大学信息工程与自动化学院, 云南 昆明 650500;
2. 昆明理工大学云南省计算机技术应用重点实验室, 云南 昆明 650500)

摘要: 现实世界中的数据通常遵循长尾分布, 假设全局数据分布平衡的联邦学习方法难以对长尾数据中的尾类数据进行准确分类。为此, 现有研究为全局模型重新训练一个平衡的分类器, 以缓解长尾数据带来的影响, 但其未考虑平衡模型的特征提取器以及如何让模型的特征提取器学习高质量的图像特征, 导致全局模型性能不佳。为了使模型在特征学习阶段没有偏见地学习高质量的图像特征, 提出一种结合旋转自监督和对比语言-图像预训练 (CLIP) 指导的联邦学习方法, 通过使用旋转自监督学习来指导本地客户端模型的训练, 减少长尾数据对客户端模型造成的影响, 并使模型高质量地学习图像中的特征。同时, 利用 CLIP 对模型的正常训练以及旋转后的图片进行指导, 将 CLIP 中丰富的知识转移到客户端模型中, 进一步提升特征提取器的性能。在不同长尾分布下的 CIFAR-10 和 CIFAR-100 数据集上进行测试, 并与其他联邦学习方法进行对比, 实验结果表明, 与现有方法相比, 该方法可将全局模型的分类准确率提升 2.35~4.72 个百分点。

关键词: 联邦学习; 长尾分布; 异构数据; 自监督学习; 对比语言-图像预训练

中图分类号: TP391

文献标志码: A

DOI: 10.19678/j.issn.1000-3428.0070288

Federated Learning on Long-Tail Data Combining Rotational Self-Supervision and CLIP Guidance

LIU Haijun¹, FU Xiaodong^{1,2}

(1. Faculty of Information Engineering and Automation,
Kunming University of Science and Technology, Kunming 650500, Yunnan, China;
2. Yunnan Key Laboratory of Computer Technology Application,
Kunming University of Science and Technology, Kunming 650500, Yunnan, China)

【Abstract】 Real-world data often follow a long-tail distribution. Federated learning methods that assume a balanced global data distribution struggle to classify tail-class data within long-tail data accurately. Researchers typically focus on retraining a balanced classifier for the global model, to mitigate the impact of long-tail data. However, this approach does not consider the feature extractor of the balanced model or how the model's feature extractor can be enabled to learn high-quality image features, leading to the poor performance of the global model. To enable the model to learn high-quality image features without bias during the feature learning stage, this study proposes a federated learning method combining rotational self-supervision and Contrastive Language-Image Pre-training (CLIP) guidance. This method uses rotational self-supervision to guide the training of local client models, thereby reducing the impact of long-tail data on the client models and enabling the model to learn high-quality image features. Simultaneously, CLIP is utilized to guide both the normal training of the model and the rotated images, transferring rich knowledge from CLIP to the client model and further enhancing the performance of the feature extractor. In experiments on the CIFAR-10 and CIFAR-100 datasets under different long-tail distributions, the proposed approach improves the global model's classification accuracy by 2.35 to 4.72 percentage points, respectively, compared with other federated learning methods.

【Key words】 federated learning; long-tailed distribution; heterogeneous data; self-supervised learning; Contrastive Language-Image Pre-training (CLIP)

0 引言

联邦学习是一种分布式的机器学习, 可以在不

共享客户端私有数据的条件下协同训练一个各客户端共享的全局模型^[1]。首先, 各个客户端使用本地私有数据集训练本地模型; 然后, 将训练完成的本地

基金项目: 国家自然科学基金 (62362043); 云南省“兴滇英才支持计划”项目 (KKXY202203008); 云南省科技计划项目 (202205AF150003, 202204BQ040010, 202102AD080002)。

作者简介: 刘海军 (CCF 学生会员), 男, 硕士研究生, 主研方向为联邦学习、长尾数据; 付晓东 (通信作者), 教授、博士、博士生导师。

收稿日期: 2024-08-26

修回日期: 2024-11-21

E-mail: xiaodong_fu@hotmail.com

模型上传至中心服务器,中心服务器聚合各个本地模型后,将全局模型发送至本地客户端,进行下一轮训练,直至训练完成。现有的联邦学习方法多数假设数据的类别分布是平衡的。然而,在现实世界中,数据的类别通常呈长尾分布^[2],即数据的类别不平衡,头部类数据有着大量样本,而尾部类数据只拥有少量样本^[3],直接使用长尾数据来训练模型,通常会导致模型偏向头部类数据而忽略尾部类数据^[4],使得本地客户端模型的性能变差,进而导致难以训练出一个拥有良好性能的全局模型。

为了解决联邦学习中同时存在数据异构和长尾数据的问题,CRFF^[4]实施解耦策略,通过在服务器端生成类别均匀的联邦特征来重新训练全局模型的分类器,这可以达到与使用真实数据集重新训练全局模型相当的性能,然而,该研究没有关注长尾分布对特征提取器的影响。CLIP2FL^[5]在CRFF的基础上,将对对比语言-图像预训练(CLIP)丰富的知识转移到本地客户端的模型和中心服务器以生成联邦特征,进一步提高了全局模型的性能,但其只关注将 CLIP 的知识转移到本地客户端的模型,这将导致每个本地客户端训练出来的模型过于相似,在现实情况中,每个本地客户端拥有的数据是不一样的,该方法不能使本地客户端更好地提取本地数据集的特征。RUCR^[6]则通过融合全局统一原型和局部特征生成虚拟特征集,随后使用虚拟特征集重新训练模型的分类器,促使模型学习结构更好的特征空间并纠正有偏差的分类器。然而,该方法同样不能使本地客户端更好地提取本地数据集的特征。

为了平衡模型的特征提取器以及使模型能更好地提取图像中的特征,本文提出了一种基于旋转自监督学习和 CLIP 指导的联邦学习方法。通过在本地客户端的训练过程中使用旋转自监督学习,指导本地客户端的学习,以缓解长尾数据的影响,并增强客户端本地模型特征提取器的特征提取能力。同时,利用 CLIP 丰富的先验知识,对正常训练的图像和旋转后的图像进行指导,将 CLIP 的知识转移到本地客户端的模型中,进而增强本地客户端模型的性能。本文的主要贡献如下:

1)提出了一种使用旋转自监督学习和 CLIP 指导的模型训练方法,减少长尾分布对模型训练的影响,提升全局模型的性能。

2)针对联邦学习中的长尾数据问题,使用旋转自监督学习和 CLIP 指导本地客户端训练,促使本地客户端更好地提取图片特征以及减少长尾分布的

影响。

3)在不同长尾分布下的 CIFAR-10-LT 和 CIFAR-100-LT 数据集上进行实验,以验证本文方法的有效性。

1 相关工作

1.1 联邦学习

联邦学习包含一个中心服务器和 K 个客户端, K 个客户端分别拥有其本地数据集 $D^1, D^2, \dots, D^K$,本文的目标是在客户端之间不进行数据传输的前提下,各本地客户端在本地完成模型训练后,将本地模型传输到中心服务器,中心服务器聚合本地模型后得到全局模型,再将得到的全局模型下发到本地客户端进行进一步训练,直至得到最终的全局模型。简单来说,本文利用中心服务器在数据集 $D \triangleq \bigcup_k D^k$ 上训练一个全局模型,且在训练过程中没有数据传输。在独立同分布场景中, FedAvg 有着良好的表现,但该方法在非独立同分布场景中表现较差。SCAFFOLD^[7]认为,这是由于不同客户端本地数据分布差异较大,导致本地模型的最优目标与全局最优目标不一致,因此,使用加权平均的方法不能获得高性能的全局模型。

目前,有很多方法用于解决联邦学习中的数据异构问题。FedNova^[8]在更新全局模型前,对客户端模型进行局部更新规范化和缩放,消除了目标不一致性,保持了快速的误差收敛。文献[9]使得客户端在本地生成虚拟样本并与其他客户端共享,缓解了由于客户端数据分布差异导致的模型偏移问题。同时,也存在一些方法针对数据不平衡问题进行改进。Ratio loss^[10]在服务器端使用平衡的辅助数据来评估全局数据类分布,帮助局部模型实现优化,但是,使用辅助数据在实际应用中是一种不可行的方法。CBFL^[11]使用类别分布均衡器对本地客户端中的少数类样本进行补充,在本地客户端构造符合全局数据分布的均衡数据集进行训练。但是,以上方法不适用于长尾数据的联邦学习。

1.2 长尾学习

近年来,针对长尾分布的研究主要聚焦于解耦特征提取器和分类器。如 MiSLAS^[12]提出标签感知平滑的方法来处理不同类别过度自信的问题。SSD^[13]使用自监督蒸馏,利用软标签将头部的知识转移到尾部类,以进行长尾数据识别。AREA^[14]发现数据类的数量不能准确衡量其特征空间大小,通过分析每个类别的有效空间面积,实现了更加公平的训练。然而,以上方法是针对集

中训练中的长尾问题,由于联邦学习是分布式训练的,因此针对长尾分布集中训练的方法并不适用。

1.3 长尾数据联邦学习

目前,一些方法针对长尾数据下的联邦学习进行改进。CReFF^[4]实施解耦策略,通过在中心服务器端生成类别均匀的联邦特征来重新训练全局模型的分层器。CLIP2FL^[5]在 CReFF 的基础上,利用知识蒸馏将 CLIP 的知识转移到客户端的模型中,并在中心服务器端使用原型对比学习,使生成的联邦特征与 CLIP 文本编码器输出的特征相似。RUCR^[6]通过融合全局原型和局部特征生成虚拟特征集,随后使用虚拟特征集重新训练模型的分层器,使模型能学习到结构更好的特征空间并纠正模型有偏差的分层器。与这些方法相比,本文方法增强了特征提取器的性能,并缓解了长尾分布对模型特征

提取器造成的影响。

2 本文方法设计与实现

为了使模型更好地提取图像中的特征并减轻长尾分布对联邦学习的影响,本文针对本地客户端模型的特征提取器进行改进,提出了一种结合旋转自监督和 CLIP 指导的长尾数据联邦学习。本文方法框架如图 1 所示。自监督学习可以人为设计一些辅助任务,在无须人工标记的情况下使模型的特征提取器学习到图像的特征。受文献[13]的启发,本文利用旋转预测自监督学习^[15]指导本地客户端模型的训练,使本地客户端在训练过程中可以更好地学习到图像中的特征。由于自监督学习是在无标签的情况下进行训练的,所以其不受长尾分布的影响,可以没有偏见地学习每一类数据的特征(见第 3.6 节),以减轻模型在训练过程中所受长尾分布的影响。

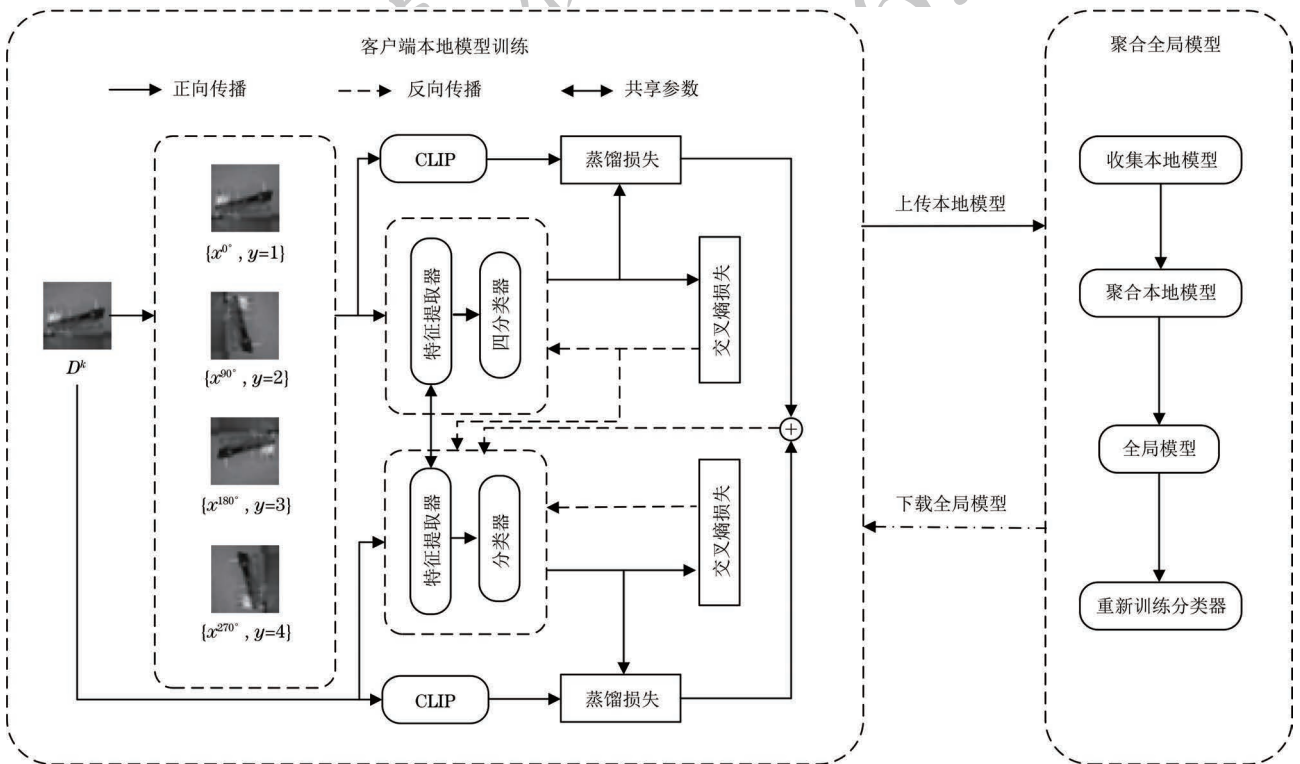


图 1 结合旋转自监督和 CLIP 指导的长尾数据联邦学习框架

Fig.1 Framework for rotational self-supervised and CLIP-guided federated learning on long-tail data

通过旋转自监督学习指导本地客户端模型的训练,使本地客户端模型的特征提取器学习高质量的特征并减轻长尾分布造成的影响。同时,CLIP^[16]含有丰富的知识,本文通过知识蒸馏的方法,将 CLIP 中丰富的知识转移到本地客户端以及旋转自监督学习中,为本地客户端模型的训练提供指导,丰富本地客户端模型的知识。最后,各本地客户端将本轮训练好的模型上传到中心服务器进行模型聚合。为了进一步提高模型的性能,本

文受文献[5]的启发,对聚合后的模型重新训练一个新的分类器。本文算法描述如算法 1 所示。在现实场景中,由于通信问题,不是所有客户端都能参加每一轮的训练,因此本文每轮选取一定比例的客户端参与训练,在每轮训练前,中心服务器按比例随机选择部分本地客户端,然后将全局模型下发至被选中的本地客户端,开始下一轮训练,以模拟现实场景。本地客户端的训练算法描述如算法 2 所示。

算法 1 结合旋转自监督学习和 CLIP 指导的联邦学习

输入 通信总轮数 T , 客户端总数 K , 本地数据集 D^k , 随机抽样客户端个数 M

输出 全局模型参数 $w^T = \{\theta^T, \varphi^T\}$

1. 初始化全局模型参数 $w^0 = \{\theta^0, \varphi^0\}$
2. for $t=1, 2, \dots, T$ do
3. $\varphi^t \leftarrow$ 从 K 个客户端中随机采样 M 个客户端
4. 发送全局模型 w^t 到客户端 φ^t 中的所有客户端
5. for $k \in \varphi^t$ do
6. 通过算法 2 获取第 k 个客户端的模型参数 s_k^{t+1}
7. end for
8. $w^{t+1} = \sum_{k \in \varphi^t} \frac{|D^k|}{\sum_{k \in \varphi^t} |D^k|} s_k^{t+1}$
9. 重新训练分类器
10. end for
11. return w^T

算法 2 客户端模型的训练

输入 全局模型参数 w^t , 本地数据集 D^k

输出 本地模型参数 s_k^{t+1}

1. 初始化本地模型参数 $s_k^t \leftarrow w^t$
2. $s^{t-\mu} = \{\theta^t, \mu^t\} \leftarrow$ 构造一个四分类器
3. for $i=1, 2, \dots, \text{Epoch}$ do
4. for each $b=(x, y)$ of D^t do
5. $(x^r, y^r) \leftarrow x$ 旋转后得到的合集
6. $p_k^t \leftarrow x$ 由 s_k^t 输出的 logit 向量
7. $p_k^{t-r} \leftarrow x^r$ 由 s_k^t 输出的 logit 向量
8. $q_k^t \leftarrow x$ 由 CLIP 输出的 logit 向量
9. $q_k^{t-r} \leftarrow x^r$ 由 CLIP 输出的 logit 向量
10. $L_1 = L_{\text{ce_sup}}(x, y, s^t) + \epsilon \cdot L_{\text{ce_self}}(x^r, y^r, s^{t-\mu})$
11. $L_2 = \beta \cdot \text{KL}(q_k^t \parallel p_k^t) + \delta \cdot \text{KL}(q_k^{t-r} \parallel p_k^{t-r})$
12. $L_{\text{local}} = L_1 + L_2$
13. $s_k^{t+1} \leftarrow s_k^t - \eta \nabla L_{\text{local}}$
14. end for
15. end for
16. return s_k^{t+1}

2.1 旋转自监督学习指导客户端模型训练

由于客户端数据高度不平衡, 稀有类别在训练过程中出现次数很少, 使用传统的交叉熵损失函数可能会忽略数量稀少的类, 模型缺少稀有类数据梯度的指导, 导致分类错误^[17], 而联邦学习旋转自监督学习的方法不会受到长尾分布的影响 (见第 3.6 节)。为此, 本文在客户端模型的训练过程中, 使用旋转自监督学习指导本地客户端模型的训练, 使客户端模型更好地提取图片特征同时平等地对待每一张图片, 减少训练过程中模型所受数据不均衡的影响。本文受 Gidaris^[15] 的启发, 采用了效果最好的图像旋转方式, 将每一张图

像分别旋转 $\{0^\circ, 90^\circ, 180^\circ, 270^\circ\}$, 旋转的角度标签分别为 $\{1, 2, 3, 4\}$, 创建一个四分类的分类层 v , 参数表示为 μ 。接着, 将构造好的图像和标签输入到客户端模型的特征提取器 f^k 中, 进一步将特征提取器输出的结果输入 v 中, 让模型预测图像旋转的角度, 以达到使模型学习图像特征的目的。由此, 组成一个四分类模型, 参数为 $s^{t-\mu} = \{\theta^t, \mu^t\}$, 相当于构造了一个四分类的有监督任务, 如图 2 所示, 则经过旋转自监督指导的客户端损失函数 $L_{\text{sup+self}}$ 为:

$$L_{\text{sup+self}} = L_{\text{ce_sup}}(x, y) + \epsilon \cdot L_{\text{ce_self}}(x^r, y^r) \quad (1)$$

式中: $L_{\text{ce_sup}}$ 表示客户端正常训练的有监督学习的损失函数; $L_{\text{ce_self}}$ 表示客户端使用旋转自监督学习指导的损失函数; L_{ce} 表示交叉熵损失函数; ϵ 表示旋转自监督学习指导的程度; $x^r = \{x^i\}_{i=0^\circ}^{270^\circ}, i \in \{0^\circ, 90^\circ, 180^\circ, 270^\circ\}$ 表示图像 x 经过旋转后的图像; y 和 y^r 分别表示 x 和 x^r 的标签, 且 $y^r = \{y^i\}_{i=1}^4, i \in \{1, 2, 3, 4\}$ 。

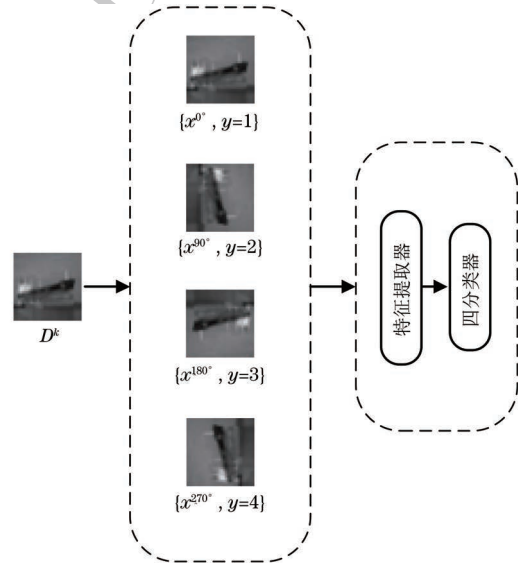


图 2 旋转自监督示意图

Fig.2 Schematic diagram of rotational self-supervision

2.2 CLIP 指导的客户端模型训练

CLIP^[16] 拥有丰富的先验知识, 为了进一步缓解客户端模型受类别不平衡数据异构的影响, 本文将 CLIP 作为教师模型, 将客户端作为学生模型, 让 CLIP 为客户端提供指导^[18], 从而将 CLIP 中丰富的知识转移到本地客户端的模型中。具体来说, 将正常训练的图像和旋转后的图像分别输入 CLIP 中, 利用知识蒸馏将 CLIP 的知识传输至本地客户端的模型中, 使 CLIP 指导正常训练的图像和旋转后的图像。其中, 正常训练的图像是未经旋转的图像, 符合现实世界大多数图像的规律, 起主导作用; 旋转后

的图像帮助客户端模型更好地获得 CLIP 的知识,这样可以充分指导客户端模型对当前图片进行特征提取,同时减少数据异构和长尾数据对模型造成的影响。需要注意的是,本文使用预训练的 CLIP,而不是重新训练的。经过 CLIP 指导的客户端损失函数 $L_{\text{KL_total}}$ 为:

$$L_{\text{KL_total}} = \beta \cdot \text{KL}(q_k^t \parallel p_k^t) + \delta \cdot \text{KL}(q_k^{t,r} \parallel p_k^{t,r}) \quad (2)$$

式中: p_k^t 和 $p_k^{t,r}$ 分别表示 x 和 x^r 由 s_k^t 输出的 logit 向量; q_k^t 和 $q_k^{t,r}$ 分别表示 x 和 x^r 由 CLIP 输出的 logit 向量; KL 是 KL 散度^[19]; β 和 δ 分别控制 CLIP 对正常图像和旋转后图像的指导程度。

综上,客户端总的损失函数为:

$$L_{\text{local}} = L_{\text{sup+self}} + L_{\text{KL_total}} \quad (3)$$

2.3 模型聚合

与多数联邦学习方法的聚合方式相同,本文参考 FedAvg^[20] 进行全局模型聚合,便于与其他联邦学习方法进行比较,聚合方式如图 3 所示。在第 t 轮,服务器端发送全局模型 w^t 至客户端,客户端接收全局模型,并使用其本地数据更新全局模型。在客户端模型更新后,被选中的客户端将他们更新的模型上传到服务器端。服务器端使用加权平均的方式聚合各客户端模型的参数来更新全局模型,客户端数据量的权重如式(4)所示:

$$w^{t+1} = \sum_{k \in \phi^t} \frac{|D^k|}{\sum_{k \in \phi^t} |D^k|} s_k^{t+1} \quad (4)$$

式中: D^k 为第 k 个客户端的私有数据。

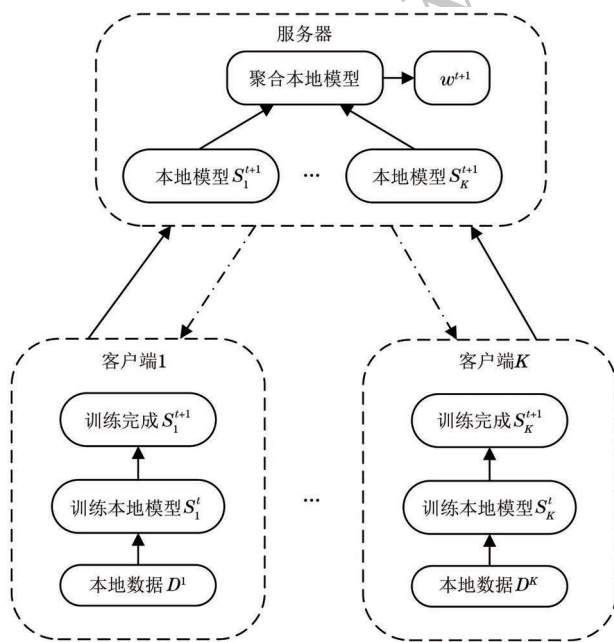


图 3 模型聚合示意图

Fig.3 Schematic diagram of model aggregation

3 实验验证

3.1 实验设置

3.1.1 数据集

通过对以往研究文献的调研,本文发现基于长尾问题的联邦学习主要是在 CIFAR-10 和 CIFAR-100^[21] 这 2 个数据集上进行验证的,因此,本文也选择这 2 个数据集作为实验的数据集。同时,本文采用与 CReFF^[4] 相同的数据集处理方式,对 2 个数据集分别设置不平衡因子为 100、50、10 的长尾分布和参数为 $\alpha=0.5$ 的狄利克雷分布,来构造客户端之间的数据异构程度,本文构造的长尾数据集称为 CIFAR-10-LT 和 CIFAR-100-LT。

1) CIFAR-10 由 60 000 张 32×32 像素的彩色图像组成,共有 10 个类别,每个类别有 6 000 张图像。其中,训练集为 50 000 张图像,测试集为 10 000 张图像。

2) CIFAR-100 的图像组成与 CIFAR-10 相同,但是其拥有的类别更多。CIFAR-100 共有 100 个类,每个类别的训练集和测试集分别为 500 和 100 张图像。

3.1.2 参数设置

为了方便与其他联邦学习方法进行比较,本文采用与文献[4-5]一样的方式,使用 ResNet-8 作为实验的基本模型,并且使用学习率为 0.1 的 SGD 优化器,客户端总数量为 20 个,通信轮数为 200 轮。每个轮次选取 40% (8 个) 客户端参加训练。本地模型在客户端的训练轮数为 10 轮,批次大小为 32。在 CIFAR-10-LT 数据集中, ϵ 和 δ 的取值均为 1; 在 CIFAR-100-LT 中,当 IF 分别为 100、50、10 时, ϵ 的取值分别为 0.3、0.3、1, δ 的取值均为 1。此外,本文使用 ViT-B/32 的 CLIP,并且在 ResNet-8 中添加了一层 MLP,以便使用 CLIP。本文实验环境是配置了 NVIDIA GeForce RTX 4090 显卡的 Ubuntu 系统。所使用的 Python 版本为 3.10.12, PyTorch 深度学习框架的版本为 2.0.1。

3.1.3 评价指标

目前,在联邦学习数据异构及长尾数据相关研究中,先前研究者所采用的评价指标均为不同长尾分布训练集上训练得到的全局模型对测试集进行分类的最高准确率。鉴于此,本文同样采用全局模型对测试集的分类最高准确率作为评价指标。

3.1.4 对比方法

为了验证本文所提方法的有效性,进行对比实

验分析。将对比方法分为 3 类:面向数据异构的联邦学习方法;面向类别不平衡的联邦学习方法;面向数据异构和长尾数据的联邦学习方法。其中几种典型对比方法具体如下:

1) 面向数据异构的联邦学习方法。

(1) FedAvg^[20]。FedAvg 是最经典的联邦学习基准算法。

(2) FedProx^[22]。FedProx 在 FedAvg 的基础上,在客户端的优化目标中加入 Proximal 项并动态调整客户端训练的迭代轮数,使得本地更新不会太过远离全局模型,提高了模型收敛的稳定性,同时保证了对系统异构的容忍性。

(3) CCVR^[23]。CCVR 是一种基于虚拟表示的分类层校准算法,通过从客户端额外收集的本地统计数据,在服务器端生成一组带有真实标签的虚拟特征,最后解耦特征提取器和分类器,使用生成的虚拟特征重新训练分类器,有效提高了全局模型的性能。

2) 面向类别不平衡的联邦学习方法。Fed-Focal Loss^[24]方法引入了 Fed-Focal 损失,降低了分配给分类良好的样本的损失,同时,通过一个可调节的采样框架,考虑到中央服务器上选择性客户端模型的贡献,在训练期间进一步关注检测器,进而提高其稳健性。

3.2 实验结果分析

实验结果如表 1 所示,最优结果加粗显示,并且使用箭头展示本文方法相比于 CLIP2FL(SOTA)提升的效果,排名第二的结果使用下划线显示。由表 1 可以得出,本文方法在所有数据集上都取得了最好结果。与 CLIP2FL(SOTA)相比,在 CIFAR-10-LT 上,当不平衡因子分别为 100、50、10 时,本文方法的最高分类准确率分别提升了 4.01、2.49、2.88 百分点,在 CIFAR-100-LT 上,当不平衡因子分别为 100、50、10 时,本文方法的最高分类准确率分别提升了 2.60、2.35、4.72 百分点,实验结果说明了本文方法的有效性。

表 1 CIFAR-10-LT 和 CIFAR-100-LT 数据集上最高分类准确率对比

Table 1 Comparison of the highest classification accuracy on the CIFAR-10-LT and CIFAR-100-LT datasets

方法类型	方法	CIFAR-10-LT 上的准确率/%			CIFAR-100-LT 上的准确率/%		
		IF 为 100	IF 为 50	IF 为 10	IF 为 100	IF 为 50	IF 为 10
面向数据异构的 联邦学习方法	FedAvg ^[20]	56.17	59.36	77.45	30.34	36.35	45.87
	FedAvgM ^[25]	52.03	57.11	70.81	30.80	35.33	44.66
	FedProx ^[22]	56.92	60.89	76.53	31.67	36.30	46.10
	FedDF ^[26]	55.15	58.74	76.51	31.43	36.22	46.19
	FedBE ^[27]	55.79	59.55	77.78	31.97	36.39	46.25
	CCVR ^[23]	69.53	71.89	78.48	33.43	36.98	46.88
	FedNova ^[8]	57.79	63.91	77.79	32.64	36.62	46.75
面向类别不平衡的 联邦学习方法	Fed-Focal Loss ^[24]	53.83	57.42	73.74	30.67	35.25	45.52
	Ratio loss ^[10]	59.75	64.77	78.14	32.95	36.88	46.79
	FedAvg+ τ -norm ^[28]	49.95	51.41	72.08	26.22	33.71	43.65
面向数据异构和 长尾数据的 联邦学习方法	CReFF ^[4]	70.55	73.08	80.71	34.67	37.64	47.08
	RUCR ^[6]	62.02	66.71	80.65	33.79	38.25	49.08
	CLIP2FL ^[5] (SOTA)	<u>73.37</u>	<u>75.35</u>	<u>81.18</u>	<u>37.56</u>	<u>41.29</u>	48.20
	本文方法	77.38	77.84	84.06	40.16	43.64	52.92
性能提升/百分点		↑ 4.01	↑ 2.49	↑ 2.88	↑ 2.60	↑ 2.35	↑ 4.72

3.3 消融分析

在 CIFAR-100-LT、CIFAR-10-LT 数据集上,且 IF 分别为 100、50、10, $\alpha=0.5$ 时,进行 2 个损失的消融实验:1) 客户端本地旋转自监督的指导损失($L_{\text{self}} = L_{\text{ce}}(x^r, y^r, p_k^{t-r})$); 2) CLIP 指导客户端旋转后的图像的知识蒸馏损失($L_{\text{KL-r}} = \text{KL}(q_k^{t-r} \parallel p_k^{t-r})$)。本文选择 CLIP2FL 作为基准模型。由表 2 可以得知,本文方

法的每个组件都发挥着重要作用。仅使用旋转自监督指导,准确率获得了 0.76~3.40 百分点的提升,这表明旋转自监督学习可以有效地帮助客户端提高模型特征提取器的能力。同时,在使用旋转自监督指导的基础上,使用 CLIP 指导旋转后的图像,准确率获得了 0.1~3.96 百分点的提升,这表明本文方法充分发挥了 CLIP 的潜力,使 CLIP 更好地指导本地模型的训练。

表 2 消融实验结果

Table 2 Ablation experiment results

方法	L_{self}	L_{KL_r}	CIFAR-10-LT 上的准确率/%			CIFAR-100-LT 上的准确率/%		
			IF 为 100	IF 为 50	IF 为 10	IF 为 100	IF 为 50	IF 为 10
Baseline	×	×	73.37	75.35	81.18	37.56	41.29	48.20
w/o L_{KL_r}	✓	×	76.77	77.74	83.82	38.76	43.43	48.96
本文方法	✓	✓	77.38	77.84	84.06	40.16	43.64	52.92

3.4 可视化结果

为了进一步验证本文方法能帮助本地客户端模型更好地提取图像中的特征,图 4 展示了在数据集为 CIFAR-10-LT、IF 为 100、 $\alpha=0.5$ 的客户端的本地模型和全局模型训练后,模型在测试数据集上提取的特征分布情况,使用 TSNE 方法进行可视化(显示其中类别数量最少的 5 类数据)。由图 4 可以看到,相比于 Baseline 方法客户端模型[图 4(a)左]提取到的特征分布,本文方法的本地客户端模型[图 4(a)右]提取到的特征呈现的菱形分布更加明

显,说明使用本文方法训练出来的本地客户端模型对数据的区分度更强,能有效提升本地客户端模型的特征提取能力。同时,本文方法训练出来的全局模型[图 4(b)右]在测试集上提取到的特征分布也比 Baseline 方法全局模型[图 4(b)左]在测试集上提取的特征分布聚集程度更高,不同类别数据的区分性也更明显,说明聚合后的全局模型也获得了更强的图像特征提取能力。TSNE 可视化结果说明了本文方法能够使训练出来的本地客户端模型以及全局模型更好地提取图像中的特征。

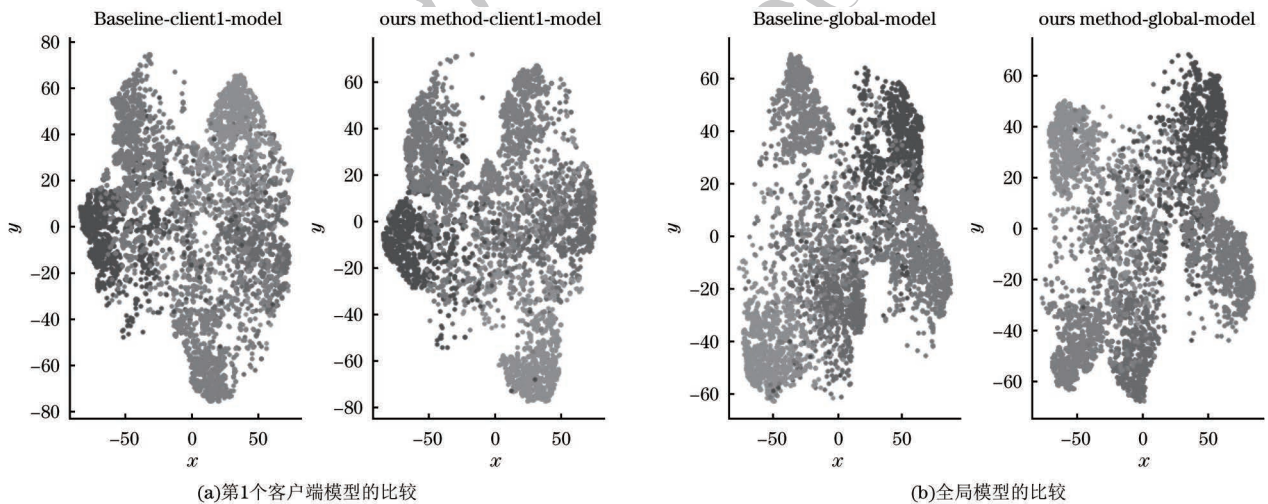


图 4 TSNE 可视化结果

Fig.4 TSNE visualization results

同时,本文还采用了与 CCVR^[23] 相同的方式,使用中心核对齐(CKA)^[29],在给定相同输入的情况下,比较客户端模型之间输出特征的相似性,CKA 输出 0(完全不相似)~1(完全相同)之间的相似性得分。该实验使用的数据集为 CIFAR-10-LT,将 IF 为 100、 $\alpha=0.5$ 时训练的全局模型与 Baseline 方法的全局模型进行比较。如图 5 所示,颜色越深表示客户端模型越不相似。与 Baseline 方法[图 5(a)]相比,本文方法[图 5(b)]显著地提升了客户端模型特征提取器的能力,进一步证明了本文方法能够使客户端模型更好地提取图像的特征,从而提升全局模型的

性能。

3.5 超参数分析

为了探究式(1)中超参数 ϵ 的不同取值对性能的影响,本文在数据集为 CIFAR-10-LT、IF 为 100、 $\alpha=0.5$ 的情况下进行实验。本实验固定 $\delta=1$,并参考 CLIP2FL^[5] 固定 $\beta=1$,选取 5 个不同的 ϵ 值,分别为 1、0.7、0.5、0.3 和 0.1。由于旋转自监督学习在本文中只起到辅助学习的作用,而不是主导作用,因此 $\epsilon \leq 1$ 。实验结果如表 3 所示,当 ϵ 逐渐减小时,模型精度逐渐降低,说明本地客户端受长尾分布的影响逐渐加重,且没有很好地提取到图像中的特征。

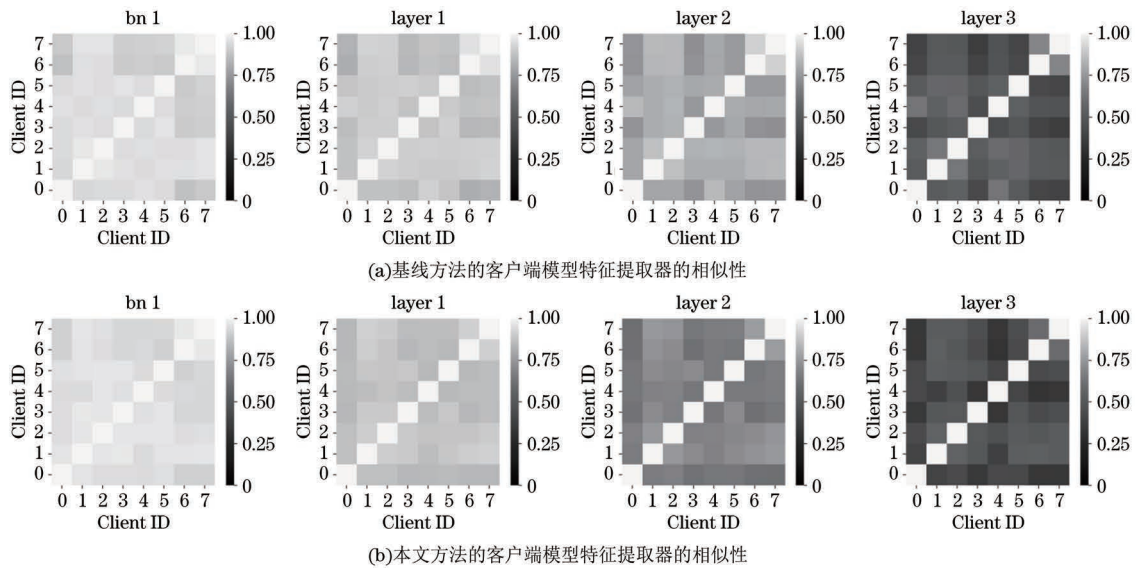


图 5 不同客户端模型特征提取器每一层的 CKA 相似性可视化

Fig.5 Visualization of CKA similarity for each layer of different client model feature extractors

表 3 不同的超参数 ϵ 取值对性能的影响Table 3 The impact of different hyperparameter ϵ values on performance

ϵ	准确率/%
1	77.38
0.7	76.22
0.5	76.60
0.3	76.12
0.1	75.56

为了探究式(2)中超参数 δ 的不同取值对性能的影响,本文在数据集为 CIFAR-10-LT、IF 为 100、 $\alpha=0.5$ 的情况下进行实验。本实验固定 $\epsilon=0.3$,并参考 CLIP2FL^[5] 固定 $\beta=1$,选取 5 个不同的 δ 值,分别为 1、0.7、0.5、0.3 和 0.1,由于 CLIP 指导的旋转自监督学习在本文中只起到辅助学习的作用,而不是主导作用,所以 $\delta \leq 1$ 。实验结果如表 4 所示,当 δ 逐渐减小时,模型精度逐渐降低,说明本地客户端受长尾分布的影响逐渐加重,且 CLIP 没有充分地对本地图客户端进行指导。

表 4 不同的超参数 δ 取值对性能的影响Table 4 The impact of different hyperparameter δ values on performance

δ	准确率/%
1	76.12
0.7	75.16
0.5	75.16
0.3	75.25
0.1	74.74

3.6 长尾分布的 FedAvg 和旋转自监督的联邦学习

为了验证旋转自监督学习是否能够缓解数据长尾分布造成的影响,本文在 CIFAR-100-LT (CIFAR-100 的长尾数据集)上进行实验,通过控制不平衡因子 IF 以及固定狄利克雷分布系数 α 构造长尾数据分布下的数据异构联邦学习,IF 越高,数据长尾程度越高, α 为控制数据异构的系数,与长尾分布无关,所以在该实验中 α 固定不变。在固定不平衡因子分别为 100、50、10,狄利克雷分布程度 $\alpha=0.5$ 时,分别对 FedAvg 算法和旋转自监督的联邦学习 (rotational-self-supervised-fedavg) 进行训练 (旋转自监督学习的联邦学习最后会使用一组包含长尾数据最少数量的类来训练一个新的分类器,从而对测试集进行分类,当 IF 分别为 100、50、10 时,该数量分别是 5、10、50),在训练结束后,通过比较 2 个不同方法得到的多数类、中数类、少数类的最高准确率,来探究 FedAvg 和旋转自监督联邦学习受长尾数据分布的影响。需要说明的是,当 IF 为 10 时,划分的长尾分布没有少数类。实验结果如图 6 所示,由图中可以看出,不论 IF 的值为多少, FedAvg 的准确率总会偏向数据多的类,而旋转自监督的联邦学习则对所有类一视同仁,不受数据量多少的影响。

4 结束语

本文在联邦学习中数据异构和长尾分布同时存在的情况下,针对模型特征提取器不平衡以及特征提取能力不足的问题,提出结合旋转自监督和 CLIP

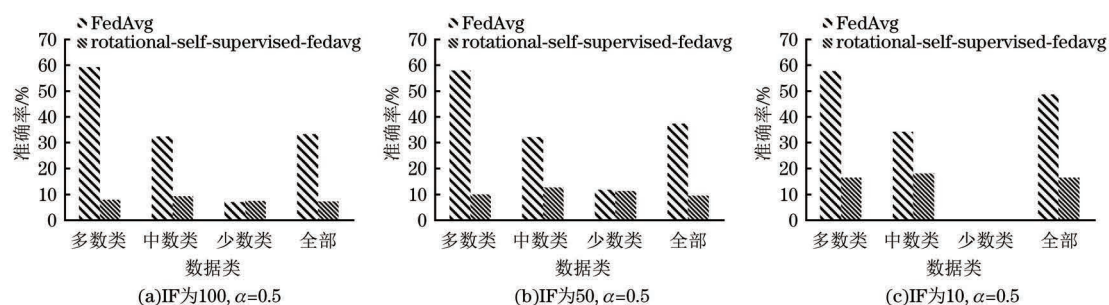


图 6 联邦学习和旋转自监督联邦学习准确率比较

Fig.6 Comparison of accuracy between federated learning and rotational self-supervised federated learning

指导的长尾数据联邦学习方法,并利用旋转自监督学习使得模型的特征提取能力更强,且能够更平等地对待每一张图像。同时,本文使用 CLIP 对本地客户端进行指导,充分发挥 CLIP 的潜力,提升模型的性能。最终,通过在不同长尾分布的数据集上与其他方法进行对比,验证了本文方法的有效性。在未来的研究中,本文将从原型的角度进一步平衡模型的特征提取器。

参考文献

- [1] 熊世强,何道敬,王振东,等. 联邦学习及其安全与隐私保护研究综述[J]. 计算机工程, 2024, 50(5): 1-15.
XIONG S Q, HE D J, WANG Z D, et al. Review of federated learning and its security and privacy protection[J]. Computer Engineering, 2024, 50(5): 1-15. (in Chinese)
- [2] 欧阳显中,韩锐,刘驰. 边缘侧领域自适应中长尾视觉识别技术研究[J]. 计算机工程, 2025, 51(7): 171-179.
OUYANG Y Z, HAN R, LIU C. Research on long-tail visual recognition technology with edge-side domain adaptation[J]. Computer Engineering, 2025, 51(7): 171-179. (in Chinese)
- [3] SHUAI X, SHEN Y L, JIANG S Y, et al. BalanceFL: addressing class imbalance in long-tail federated learning[C]//Proceedings of the 21st ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN). Washington D. C., USA: IEEE Press, 2022: 271-284.
- [4] SHANG X Y, LU Y, HUANG G, et al. Federated learning on heterogeneous and long-tailed data via classifier re-training with federated features[EB/OL]. [2024-06-05]. <https://arxiv.org/abs/2204.13399>.
- [5] SHI J M, ZHENG S S, YIN X B, et al. CLIP-guided federated learning on heterogeneity and long-tailed data[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2024, 38(13): 14955-14963.
- [6] HUANG W K, LIU Y X, YE M, et al. Federated learning with long-tailed data via representation unification and classifier rectification[J]. IEEE Transactions on Information Forensics and Security, 2024, 19: 5738-5750.
- [7] KARIMIREDDY S P, KALE S, MOHRI M, et al. SCAFFOLD: stochastic controlled averaging for federated learning[EB/OL]. [2024-06-05]. <https://arxiv.org/abs/1910.06378>.
- [8] WANG J Y, LIU Q H, LIANG H, et al. Tackling the objective inconsistency problem in heterogeneous federated optimization[EB/OL]. [2024-06-05]. <https://arxiv.org/abs/2007.07481>.
- [9] 汤凌韬,王迪,刘盛云. 面向非独立同分布数据的联邦学习数据增强方案[J]. 通信学报, 2023, 44(1): 164-176.
- [10] TANG L T, WANG D, LIU S Y. Data augmentation scheme for federated learning with non-IID data[J]. Journal on Communications, 2023, 44(1): 164-176. (in Chinese)
- [11] WANG L X, XU S C, WANG X, et al. Addressing class imbalance in federated learning[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2021, 35(11): 10165-10173.
- [12] 李志鹏,国雍,陈耀佛,等. 基于数据生成的类别均衡联邦学习[J]. 计算机学报, 2023, 46(3): 609-625.
LI Z P, GUO Y, CHEN Y F, et al. Class-balanced federated learning based on data generation[J]. Chinese Journal of Computers, 2023, 46(3): 609-625. (in Chinese)
- [13] ZHONG Z S, CUI J Q, LIU S, et al. Improving calibration for long-tailed recognition[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Washington D. C., USA: IEEE Press, 2021: 16484-16493.
- [14] LI T H, WANG L M, WU G S. Self supervision to distillation for long-tailed visual recognition[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). Washington D. C., USA: IEEE Press, 2022: 610-619.
- [15] CHEN X H, ZHOU Y C, WU D Y, et al. AREA: adaptive reweighting via effective area for long-tailed classification[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). Washington D. C., USA: IEEE Press, 2024: 19220-19230.
- [16] GIDARIS S, SINGH P, KOMODAKIS N. Unsupervised representation learning by predicting image rotations[EB/OL]. [2024-06-05]. <https://arxiv.org/abs/1803.07728>.
- [17] RADFORD A, KIM J W, HALLACY C, et al. Learning transferable visual models from natural language supervision[EB/OL]. [2024-06-05]. <https://arxiv.org/abs/2103.00020>.
- [18] TAN J R, WANG C B, LI B Y, et al. Equalization loss for long-tailed object recognition[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Washington D. C., USA: IEEE Press, 2020: 11659-11668.
- [19] 黄震华,杨顺志,林威,等. 知识蒸馏研究综述[J]. 计算机学报, 2022, 45(3): 624-653.
HUANG Z H, YANG S Z, LIN W, et al. Knowledge distillation: a survey[J]. Chinese Journal of Computers, 2022, 45(3): 624-653. (in Chinese)
- [20] JOYCE J M. Kullback-Leibler divergence[EB/OL]. [2024-06-05]. <https://statlect.com/fundamentals-of-probability/Kullback-Leibler-divergence>.
- [21] MCMAHAN H B, MOORE E, RAMAGE D, et al. Communication-efficient learning of deep networks from decentralized data[EB/OL]. [2024-06-05]. <https://arxiv.org/abs/1602.05629>.
- [22] KRIZHEVSKY A. Learning multiple layers of features from

- tiny images [EB/OL]. [2024-06-05]. <http://www.cs.toronto.edu/~kriz/learning-features-2009-TR.pdf>.
- [22] SAHU A K, LI T, SANJABI M, et al. Federated optimization in heterogeneous networks[EB/OL]. [2024-06-05]. <https://arxiv.org/abs/1812.06127>.
- [23] LUO M, CHEN F, HU D P, et al. No fear of heterogeneity: classifier calibration for federated learning with non-IID data[EB/OL]. [2024-06-05]. <https://arxiv.org/abs/2106.05001>.
- [24] SARKAR D, NARANG A, RAI S. Fed-Focal loss for imbalanced data classification in federated learning[EB/OL]. [2024-06-05]. <https://arxiv.org/abs/2011.06283>.
- [25] HSU T H, QI H, BROWN M. Measuring the effects of non-identical data distribution for federated visual classification [EB/OL]. [2024-06-05]. <https://arxiv.org/abs/1909.06335>.
- [26] LIN T, KONG L J, STICH S U, et al. Ensemble distillation for robust model fusion in federated learning [EB/OL]. [2024-06-05]. <https://arxiv.org/abs/2006.07242>.
- [27] CHEN H Y, CHAO W L. FedBE: making Bayesian model ensemble applicable to federated learning[EB/OL]. [2024-06-05]. <https://arxiv.org/abs/2009.01974>.
- [28] KANG B Y, XIE S N, ROHRBACH M, et al. Decoupling representation and classifier for long-tailed recognition [EB/OL]. [2024-06-05]. <https://arxiv.org/abs/1910.09217>.
- [29] KORNBLITH S, NOROUZI M, LEE H, et al. Similarity of neural network representations revisited[EB/OL]. [2024-06-05]. <https://arxiv.org/abs/1905.00414>.

文字编辑 吴云芳

栏目编辑 宋 圆