基于数据生成的类别均衡联邦学习

李志鹏[1),2)] 国 雍[1)] 陈耀佛[1)] 王耀威[2)] 曾 炜[2),3)] 谭明奎[1)]

[1)](华南理工大学软件学院 广州 510006)

[2)](鹏城实验室人工智能研究中心 广东 深圳 518054)

[3)](北京大学信息科学技术学院 北京 100871)

摘 要 手机、可穿戴设备等终端设备每天产生海量数据,但这些数据往往涉及敏感隐私而不能直接公开并使用.为解决隐私保护下的机器学习问题,联邦学习应运而生,旨在通过构建协同训练机制,在不共享客户端数据条件下,训练高性能全局模型.然而,在实际应用中,现有联邦学习机制面临两大不足:(1)全局模型需考虑多个客户端的数据,但各客户端往往仅包含部分类别数据且类别间数据量严重不均衡,使得全局模型难以训练;(2)各客户端之间的数据分布往往存在较大差异,导致各客户端模型往往差异较大,使得传统通过模型参数加权平均以获得全局模型的方法难以奏效.为降低客户端类别不均衡和数据分布差异的影响,本文提出一种基于数据生成的类别均衡联邦学习(Class-Balanced Federated Learning,CBFL)方法.CBFL旨在通过数据生成技术,针对各客户端构造符合全局模型学习的类别均衡数据集.为此,CBFL设计了一个包含类别均衡采样器和数据生成器的类别分布均衡器.其中,类别均衡采样器对客户端数据量不足的类别以较高概率进行采样.然后,数据生成器则根据所采样的类别生成相应的虚拟数据以均衡客户端数据的类别分布并用于后续的模型训练.为验证所提出方法的有效性,本文在四个标准数据集上进行了大量实验.实验表明,本文方法可大幅提升联邦学习性能:如在CIFAR-100数据集上,CBFL训练的ResNet20模型与现有方法相比,分类准确率提高了5.82%.

关键词 联邦学习;数据生成;类别分布;类别不均衡;隐私保护

中图法分类号 TP18 **DOI号** 10.11897/SP.J.1016.2023.00609

Class-Balanced Federated Learning Based on Data Generation

LI Zhi-Peng[1),2)] GUO Yong[1)] CHEN Yao-Fo[1)] WANG Yao-Wei[2)] ZENG Wei[2),3)] TAN Ming-Kui[1)]

[1)](*School of Software Engineering, South China University of Technology, Guangzhou* 510006)

[2)](*Artificial Intelligence Research Center, Peng Cheng Laboratory, Shenzhen, guangdong* 518054)

[3)](*School of Electronics Engineering and Computer Science, Peking University, Beijing* 100871)

Abstract Modern terminal devices such as mobile phones and wearable devices produce massive amounts of data every day, but these data often involve sensitive privacy and thus cannot be directly disclosed and used. To solve this problem, Federated Learning (FL) has been developed as an important machine learning framework under privacy protection, which allows extensive terminal devices/clients to collaboratively learn a superior global model, without sharing the private data on the clients. However, in practical application, there are still two underlying limitations to existing FL mechanism. First, the global model needs to consider the data on multiple clients, but each client usually contains only partial classes of data and the data amount of different classes is severely imbalanced, making it difficult to train the global model. Specifically, most data

收稿日期:2021-06-30;在线发布日期:2022-02-09.本课题得到科技部青年项目(2020AAA0106900)、国家自然科学基金联合基金项目(U20B2052)、国家自然科学基金项目(62072190)、广东省重点领域研发计划项目(2018B010107001)、广东省珠江人才计划创新创业团队(2017ZT07X183)资助.**李志鹏**,硕士研究生,研究领域为深度学习和计算机视觉.Email:sezhipengli@mail.scut.edu.cn.**国 雍(共同第一作者)**,博士研究生,研究领域为深度学习.Email:guo.yong@mail.scut.edu.cn.**陈耀佛**,博士研究生,研究领域为深度学习.**王耀威**,博士,研究领域为计算机视觉.**曾 炜**,博士,研究领域为计算机视觉.**谭明奎(通信作者)**,博士,教授,研究领域为机器学习.Email:mingkuitan@scut.edu.cn.

on the client belong to a few classes, while other classes have few or no data. As a result, the trained local models tend to overfit the data on the clients and achieve poor performance on global data, which severely affects the training of the global model. Second, the data distribution is extremely different across the clients, which causes the trained models on each client to be quite different, making it hard to derive a promising global model. In fact, the training data on each client usually come from the usage of the terminal device by a particular user. Due to the differences in the functions of the terminal devices and the usage habits of users, different clients often produce different classes of data, leading to extremely different class distribution across the data on the clients. Consequently, there will be huge differences among the local models trained on such distribution, making it difficult to obtain a superior global model through the traditional approach of element-wise weighted averaging model parameters. To reduce the impact of class imbalance and distribution differences, in this paper, we propose a novel Class-Balanced Federated Learning (CBFL) method based on data generation, which aims to produce a class-balanced data set suitable for the training of global model for each client through data generation technique. To this end, CBFL designs a class distribution equalizer that consists of a class-balanced sampler and a data generator. First, the class-balanced sampler samples those classes that have insufficient data on the client with a higher sample probability. Then, the data generator generates corresponding dummy data according to the classes sampled by the class-balanced sampler. Finally, each client combines its original data and the generated data to produce a class-balanced data set for training. In this way, the performance of each local model can be greatly improved and the differences among local models are highly reduced, which contributes to obtaining a promising global model. Moreover, to obtain high-quality generated data, we exploit global data distribution information from the global model to train the data generator. Extensive experiments on four benchmark datasets demonstrate the superior performance of the proposed method over existing methods. For example, the ResNet20 model trained on CIFAR-100 dataset by the proposed CBFL outperforms existing methods by 5.82% in terms of accuracy.

Keywords federated learning; data generation; class distribution; class imbalance; privacy protection

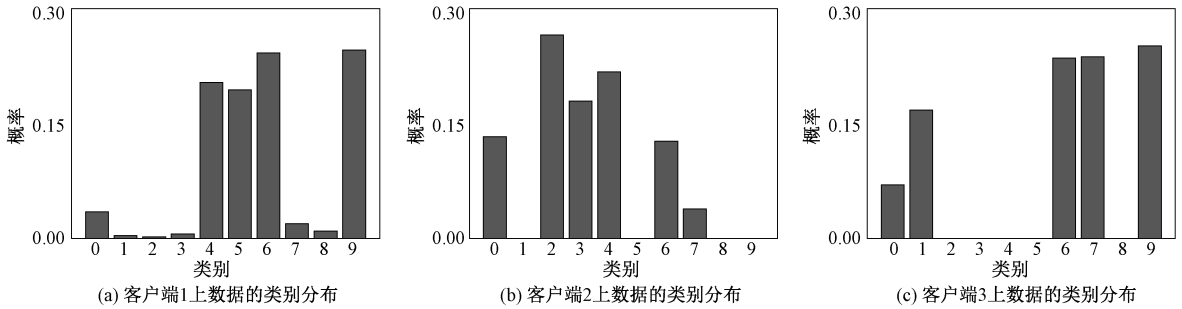
1 引 言

深度神经网络[1-2]在许多具有挑战性的任务中都取得了重大进展,如图像分类、人脸识别和目标检测等. 这一系列进展背后的关键因素之一是海量数据的收集与利用. 当前,海量数据分散于各种终端设备,如手机、可穿戴设备和传感器等. 为获得高性能深度神经网络模型,传统方法往往将各终端设备的数据收集并汇聚于同一数据中心后,对其进行中心化训练. 然而,收集并汇聚各设备的数据会导致数据隐私泄露问题. 在某些数据隐私敏感的场景下(如医疗诊断、人脸识别),传统的中心化训练因各设备的数据无法汇聚于数据中心而难以进行. 更重要的是,由于每台设备的本地数据量往往有限,各终端设备仅使用本地数据进行单独训练难以获得高性能深度神经网络模型.

为解决该问题,联邦学习[3-6]已被提出并引起广泛关注. 联邦学习是一种分布式机器学习方法,其利用一个中央服务器(也称为服务器端)协调各终端设备(也称为客户端),协同训练一个各客户端共享的全局模型. 与传统中心化训练方法不同,联邦学习不需要各设备发送自身隐私数据至数据中心,因此有利于保护数据隐私. 具体而言,联邦学习在客户端和服务器端之间通过多轮通信迭代优化模型. 每轮通信包含两个阶段:(1)各客户端从服务器端下载全局模型,并在本地数据上进行训练以获得本地模型;(2)服务器端接收并聚合各客户端的本地模型参数以获得性能更优的全局模型. 然而,现有联邦学习机制尚面临两大不足.

首先,全局模型需考虑多个客户端的数据,但各客户端往往仅包含部分类别数据且类别间数据量严重不均衡,使得全局模型难以训练. 如图 1 所示,每个客户端往往只拥有部分类别的数据. 特别地,客户端的少数类别占据了大部分的数据,而其它类别的数据则很少甚至没有. 因此,客户端数据的类别分布往往不均衡[3,5]. 值得一提的是,传统中心化训练方法从全局数据中随机采样各类别的数据进行训练,从而有利于获得在全局数据上性能优越的模型. 然而,现有联邦学习方法在各客户端上仅利用其本地数据来训练各自的本地模型[3,6-8],导致所训练的本地模型容易过拟合本地数据而在全局数据上往往取得较差性能. 更重要的是,这些性能较差的本地模型严重影响全局模型的训练,导致难以构建高性能全局模型. 因此,如何降低客户端数据类别不均衡的影响以利于构建高性能全局模型是一个重要问题.

其次,各客户端之间的数据分布往往存在较大差异,导致各客户端模型往往差异较大,使得传统通过模型参数加权平均以获得全局模型的方法难以奏效. 实际上,各客户端的数据来源取决于用户对该客户端的使用情况. 由于各客户端的功能和用户使用习惯不同,不同客户端往往产生不同类别的数据,导致各客户端数据之间的类别分布差异较大,如图 1 所示. 因此,各客户端基于自身数据所训练的本地模型之间往往存在较大差异. 更重要的是,服务器端难以通过传统的模型参数加权平均的方法聚合这些差异性大的本地模型以获得高性能全局模型[3,6,8]. 因此,如何减小本地模型之间的差异性以有效聚合本地模型并获得高性能全局模型仍然是一个开放性问题.



(a) 客户端1上数据的类别分布　(b) 客户端2上数据的类别分布　(c) 客户端3上数据的类别分布

图 1　部分客户端上数据的类别分布(以 CIFAR-10 数据集为例)

针对上述问题,本文提出一种基于数据生成的类别均衡联邦学习(Class-Balanced Federated Learning,CBFL)方法. CBFL 通过数据生成技术,针对各客户端构造符合全局模型学习的类别均衡数据集,以降低客户端类别不均衡和数据分布差异的影响. 具体而言,CBFL 设计了一个包括类别均衡采样器和数据生成器的类别分布均衡器. 其中,类别均衡采样器对客户端数据量不足的类别以较高概率进行采样. 然后,数据生成器从全局模型学习全局数据分布信息,从而根据类别均衡采样器得到的类别生成相应的虚拟数据. 结合本地数据和生成的虚拟数据,各客户端构造类别均衡的数据集进行训练,使得各本地模型的性能得以提高且各本地模型之间的差异性大大减少,进而有利于构建高性能全局模型. 本文在四个标准数据集上进行了大量实验,证明了所提出方法相对于现有方法的优越性.

本文的主要贡献总结如下:

(1)本文提出类别分布均衡器来均衡客户端的类别分布. 其中,类别分布均衡器由类别均衡采样器和数据生成器组成. 类别均衡采样器对客户端数据量不足的类别以较高概率进行采样,数据生成器则根据所采样的类别生成相应的虚拟数据;

(2)基于类别分布均衡器,本文提出一种新颖的类别均衡联邦学习方法(CBFL). CBFL 在客户端构造类别均衡的数据集进行训练,从而提高各本地模型的性能并减少各本地模型之间的差异性,进而构建高性能全局模型;

(3)本文在四个标准数据集上进行了大量实验,证明了所提出 CBFL 的有效性. 与最新方法相比,CBFL 在多种深度神经网络(如 ResNet20 和 MobileNetV2)取得更优越的性能. 尤其在客户端上类别分布高度不均衡且客户端之间类别分布差异巨大的情况下,CBFL 比现有方法具有更大的优势.

2　相关工作

2.1　联邦学习

联邦学习是一种面向数据隐私保护的分布式机器学习框架. 与传统的中心化训练方式不同,联邦学习允许客户端在不共享各自隐私数据的前提下协

同训练一个共享的全局模型. 联邦学习的经典算法是 FedAvg[3],该算法对本地模型的参数执行加权平均以获得性能更优的全局模型. 其中,各本地模型加权平均的权重与客户端上训练数据集的大小成正比. 然而,客户端本地数据的类别分布严重不均衡且客户端之间的数据分布往往差异巨大,难以通过模型参数简单加权平均的方法获得高性能全局模型. 为解决该问题,FedProx[6]在客户端训练本地模型的目标函数中引入一个 Proximal 项,使得本地模型尽量靠近全局模型以缓解本地模型之间差异性大带来的影响. SCAFFOLD[7]提出在服务器端和客户端分别维护一个全局控制变量和局部控制变量,并通过全局控制变量和局部控制变量的差值来修正各客户端本地模型的更新方向. FedNova[8]提出在更新全局模型之前对客户端本地模型的局部更新进行规范化和缩放. 此外,还有一些研究工作通过个性化客户端的本地模型[9-11]或者针对客户端数据分布的不同组合设计一种鲁棒的算法[12-14]来降低客户端之间数据分布差异带来的影响. 与现有方法不同,本文针对各客户端构造类别均衡的数据集进行训练,从而有效聚合各本地模型并获得高性能全局模型.

2.2 生成模型

生成模型旨在生成高质量图像,在图像生成领域受到越来越多的关注. 近年来,生成对抗网络(Generative Adversarial Network,GAN)[15-18]在图像生成领域取得了长足的进步. 然而,在客户端数据类别高度不均衡且客户端之间类别分布差异巨大的情况下,如何通过联邦学习获得一个高性能的 GAN 模型仍然是一个有待解决的问题. 此外,另一个研究方向旨在挖掘已训练模型的信息来生成图像. 具体而言,Fredrikson 等人[19]利用已训练模型的信息直接对输入噪声进行优化以获得生成图像. 然而,该方法尚未在深度神经网络取得良好的性能. 为利用深度神经网络信息,Mahendran 等人[20]采用激活值最大化技术从已训练模型中挖掘激活值信息以用于图像生成. Micaelli 等人[21]提出一种对抗学习的方式来训练生成模型. Naya 等人[22]利用已训练模型的类相似性来获得生成图像. 然而,如何在联邦学习的框架下挖掘深度神经网络的信息以生成高质量的图像仍然是一个开放性问题.

3 问题定义

本文研究联邦学习的训练机制. 假设联邦学习系统共有 K 个客户端,由一个中心服务器(服务器端)协调. 令 $\mathcal{D}_k=\{(x_k^i,y_k^i)\}_{i=1}^{N_k}$ 表示第 k 个客户端的数据及其类别,N_k 表示第 k 个客户端的数据量. $p(\mathcal{D}_k)$ 表示第 k 个客户端数据集所代表的经验分布. 各客户端在服务器端的协调下训练各自本地模型,然后服务器端聚合各本地模型参数以更新全局模型. 令 T 表示全局模型,S_k 表示第 k 个客户端的本地模型,$\boldsymbol{W}_T$ 和 $\boldsymbol{W}_{S_k}$ 分别表示全局模型 T 和第 k 个本地模型 S_k 的参数. 联邦学习的流程如图 2 所示.

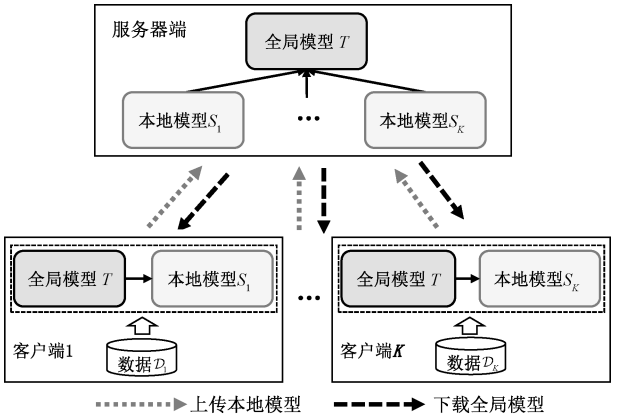


图 2 联邦学习流程

联邦学习旨在获得一个各客户端共享的全局模型 T. 令 $\mathcal{L}(\cdot;\cdot)$ 表示客户端训练本地模型所使用的损失函数(如交叉熵损失函数). 因此,联邦学习解决以下优化问题:

$$\min_{\boldsymbol{W}_T}\mathbb{E}_k[\mathbb{E}_{x_k\sim p(\mathcal{D}_k)}[\mathcal{L}(x_k;\boldsymbol{W}_T)]] \tag{1}$$

然而,现有联邦学习机制尚面临两大不足. 第一,全局模型需考虑多个客户端的数据,但各客户端往往仅包含部分类别数据且类别间数据量严重不均衡. 因此,各客户端训练的本地模型容易过拟合本地数据而在全局数据上往往取得较差性能,从而严重影响全局模型的训练. 第二,各客户端之间的数据分布往往存在较大差异,导致各客户端本地模型往往差异较大,使得传统通过模型参数加权平均以获得全局模型的方法难以奏效.

为解决上述难题,本文提出在客户端构造类别均衡的数据集进行训练的策略. 令$\mathcal{D}$表示所构造的类别均衡数据集,M 表示全局数据的类别数,则 $\mathcal{D}=\{(x,y)\,|\,P(y=i)=\dfrac{1}{M},\ i\in\{0,1,\ldots,M-1\}\}$. $p(\mathcal{D})$表示数据集$\mathcal{D}$所代表的经验分布. 因此,本文旨在解决如下优化问题:

$$\min_{\boldsymbol{W}_T}\mathbb{E}_k[\mathbb{E}_{x_k\sim p(\mathcal{D})}[\mathcal{L}(x_k;\boldsymbol{W}_T)]] \tag{2}$$

与问题(1)不同,问题(2)在客户端构造类别均衡的数据集来训练本地模型. 因此,对于第一个难题,本文在各客户端训练的本地模型能够在全局数据上取得更好性能,从而利于构建高性能全局模型. 对于第二个难题,与仅仅利用客户端本地数据进行训练的方式相比,基于类别均衡的数据集进行训练使得各客户端本地模型之间的差异大大减少. 更重要的是,服务器端更易于聚合这些性能良好且差异性小的本地模型以获得高性能全局模型.

然而,求解问题(2)仍然十分困难. 首先,如何在客户端获得类别均衡的数据集仍然未知;其次,如何利用所构造的类别均衡数据集来有效训练本地模型仍然是一个开放的问题.

4　基于数据生成的类别均衡联邦学习

4.1　算法概述

为求解问题(2),本文提出一种新颖的基于数据生成的类别均衡联邦学习算法(CBFL). CBFL 的整体方案如图 3 所示. 为构造类别均衡的数据集,本文提出一个类别分布均衡器. 该类别分布均衡器根据各客户端自身数据类别分布,针对客户端本地数据量不足的类别以较高概率生成所对应的虚拟数据. 通过结合客户端的本地数据和生成的虚拟数据,本文面向各客户端构造类别均衡的数据集. 为更有效训练本地模型,本文引入交叉熵损失函数和蒸馏损失函数进行训练. 最后,服务器端接收并聚合各客户端的本地模型参数以更新全局模型. 客户端数据生成的整体方案如图 4 所示. 其中,类别分布均衡器由一个类别均衡采样器和一个数据生成器组成. 首先,类别均衡采样器根据客户端的类别分布,以较高概率采样本地数据量不足的类别. 然后,数据生成器根据类别均衡采样器所采样的类别生成相应的虚拟数据.

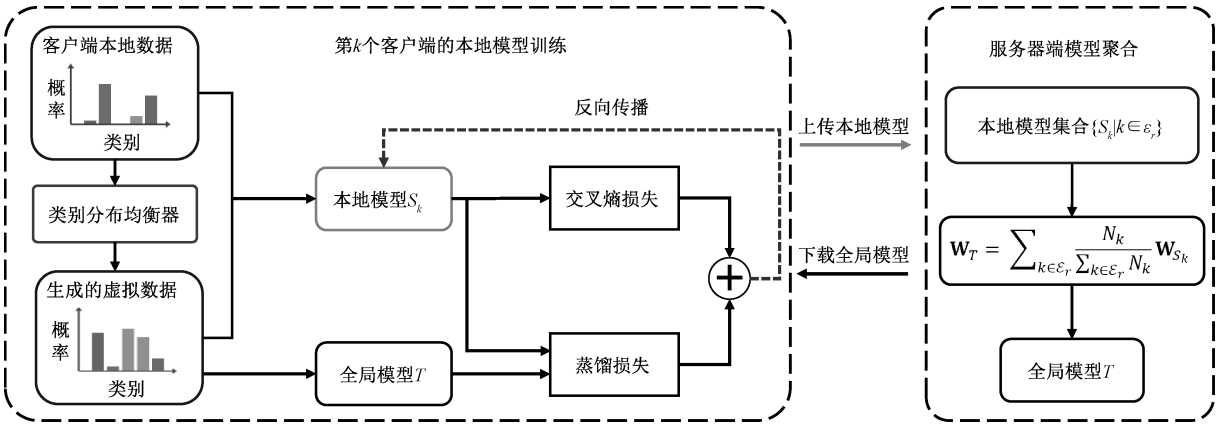


图 3　基于数据生成的类别均衡联邦学习方案

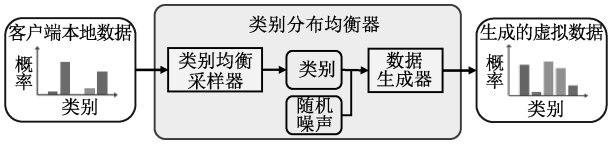


图 4　数据生成方案

本文提出的基于数据生成的类别均衡联邦学习算法如算法 1 所示. 在实际应用场景中,由于通信网络不稳定,并非所有客户端在每轮通信中都能被访问. 为了模拟该场景,在每轮通信中,服务器端随机采样部分客户端参与训练. 令 c 表示每轮通信中随机采样的客户端数量占客户端总数的比例,ε_r 表示第 r 轮通信参与训练的客户端集合. 在每轮通信中,服务器端先将当前的全局模型分发到被采样的客户端. 然后,被采样的客户端 $k \in \varepsilon_r$ 先利用全局模型初始化本地模型,再对本地模型进行训练. 最后,服务器端以加权平均的方式聚合各客户端的本地模型参数来获得性能优越的全局模型.

客户端本地模型的训练算法如算法 2 所示. 本文使用一种预热的方式来训练本地模型. 这是因为本文利用全局模型的全局数据分布信息来训练数据生成器. 在起始阶段,全局模型尚未经过充分的训练,不能获得准确的全局数据分布信息来训练数据生成器,从而影响本地模型的训练并损害全局模型性能(见第 6.5 节). 具体而言,本文将预热轮数设为 R_w. 当通信轮次 r 小于预热轮数 R_w 时,本文仅使用客户端本地数据来训练本地模型. 当通信轮次 r 大于预热轮数 R_w 时,本文先对数据生成器进行训练,然后通过类别均衡采样器采样客户端本地数据量不足的类别,并利用数据生成器生成所采样类别的虚拟数据. 最后,客户端结合本地数据和虚拟数

据构造类别均衡的数据集,用于训练本地模型.

算法 1. 基于数据生成的类别均衡联邦学习.

输入:通信总轮数 R,客户端总数 K,每轮通信采样客户端的比例 c

输出:全局模型参数 $\boldsymbol{W}_T$

1. 初始化全局模型参数 $\boldsymbol{W}_T$
2. FOR 通信轮次 $r=1,2,\ldots,R$ DO
3. $t \leftarrow \max(K \cdot c, 1)$
4. 随机采样 t 个客户端构成集合 ε_r
5. FOR 客户端 $k \in \varepsilon_r$ DO
6. 通过算法 2 获取第 k 个客户端的模型参数 $\boldsymbol{W}_{S_k}$
7. END FOR
8. $\boldsymbol{W}_T \leftarrow \sum_{k \in \varepsilon_r} \frac{N_k}{\sum_{k \in \varepsilon_r} N_k} \boldsymbol{W}_{S_k}$
9. END FOR

4.2 类别分布均衡器

4.2.1 类别均衡采样器

本文旨在额外采样客户端本地数据量不足的类别来均衡客户端的类别分布. 然而,给定客户端的类别分布,如何确定每个类别的采样概率是一个难题. 为解决该问题,本文设计了一个类别均衡采样器. 该类别均衡采样器尽可能多地采样客户端本地数据量不足的类别,从而实现客户端数据的类别均衡. 具体而言,每个类别在客户端的数据量越大,该类别的采样概率就越小. 令 M 表示全局数据的类别数,n_m 表示客户端上类别为 m 的数据量. 本文通过以下三步构造类别均衡采样器:

(1)统计客户端上类别 m 的概率:

$$P_m = \frac{n_m}{\sum_{m=0}^{M-1} n_m} \tag{3}$$

(2)计算与客户端上类别分布相反的采样概率,即类别 m 的采样概率为:

$$\bar{P}_m = 1 - P_m \tag{4}$$

(3)归一化采样概率:

$$\widetilde{P}_m = \frac{\bar{P}_m}{\sum_{m=0}^{M-1} \bar{P}_m} \tag{5}$$

类别均衡采样器构造完成后,各客户端根据采样概率 $\widetilde{P}_m$ 采样相应的类别 m. 值得注意的是,该过程在客户端进行,客户端的数据分布信息不会发送至服务器端.

算法 2. 客户端本地模型的训练.

输入:当前通信轮次 r,预热轮数 R_w,客户端索引 k,全局模型参数 $\boldsymbol{W}_T$

输出:本地模型参数 $\boldsymbol{W}_{S_k}$

// 运行在第 k 个客户端上

1. 初始化本地模型参数:$\boldsymbol{W}_{S_k} \leftarrow \boldsymbol{W}_T$
2. IF $r < R_w$ THEN
3. 利用客户端本地数据更新 $\boldsymbol{W}_{S_k}$
4. ELSE
5. 通过优化目标函数(9)训练数据生成器
6. 通过公式(5)采样类别
7. 根据公式(6)生成所采样类别对应的虚拟数据
8. 结合本地数据和虚拟数据通过公式(10)更新 $\boldsymbol{W}_{S_k}$
9. END IF

与现有均衡采样方法[23-24]的不同:(1) 现有均衡采样方法仅仅基于本地数据的类别来计算每个类别的采样概率. 直接将现有均衡采样方法应用于联邦学习中没有考虑客户端所缺乏的类别. 本文的类别均衡采样器基于全局数据的类别来计算每个类别的采样概率;(2) 现有均衡采样方法对本地数据进行重采样以提高数据量较少类别的比例. 如果某个类别在本地数据中不存在,那么上述重采样方法无法采样到该类别的数据. 本文通过类别均衡采样器对类别标签进行采样以获得客户端各类别的采样量,然后利用数据生成器生成相应类别的虚拟数据.

4.2.2 数据生成器

通过类别均衡采样器,客户端可获得各类别的采样量. 然而,如何根据所采样的类别获得相应的数据以用于本地模型的训练仍然是一个难题. 为解决该问题,本文引入了数据生成器 G. 该数据生成器的输入为类别标签 y 和噪声矢量 $\boldsymbol{z}$. 其中 $y \in \{0,1,\ldots,M-1\}$,$\boldsymbol{z}$ 服从高斯分布 $N(0,1)$. 数据生成器 G 根据噪声矢量 $\boldsymbol{z}$ 和类别标签 y 生成相应的数据 $\hat{x}$,即

$$\hat{x} = G(\boldsymbol{z} \mid y), \ \boldsymbol{z} \sim N(0,1). \tag{6}$$

数据生成器和类别均衡采样器构造完成后,客户端首先根据公式(5)采样类别 m,然后利用数据生成器生成类别为 m 的虚拟数据. 最后,客户端结合本地数据和生成的虚拟数据,从而构造类别均衡的数据集来训练本地模型.

4.3 模型训练

首先,客户端对数据生成器进行训练,从而根据类别均衡采样器获得的类别生成虚拟数据. 然后,客户端结合本地数据和虚拟数据训练本地模型.

4.3.1 数据生成器的训练

为训练数据生成器,一种常用的做法是采取对抗学习的方式[15-18]. 该方式需要额外训练一个判别

器,且训练判别器需要相应类别大量的真实数据样本. 然而,客户端上部分类别的数据样本很少甚至没有,导致所训练的数据生成器难以生成客户端本地数据量不足的类别所对应的数据. 为解决上述问题,本文提出利用全局模型的全局数据分布信息来训练生成器,从而无需训练判别器. 具体而言,该全局数据分布信息包含分类边界信息和统计信息. 数据生成器的训练方案如图 5 所示. 与传统训练生成器的方式不同,本方案不需要获得真实数据样本,故生成质量与真实数据的样本量无关.

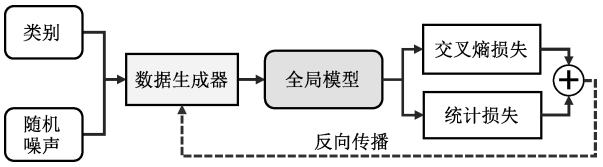


图 5　　数据生成器的训练方案

首先,全局模型包含全局数据的分类边界信息. 本文希望数据生成器学习该分类边界信息,从而能够生成给定类别对应的数据. 为此,给定噪声 $\boldsymbol{z}$ 和类别 y,生成的数据 $G(\boldsymbol{z}|y)$ 输入全局模型 T 后应当能被正确分类到相应类别 y. 因此,本文引入交叉熵损失函数 $CE(\cdot,\cdot)$ 来训练数据生成器 G,即

$$\mathcal{L}_{CE}(G)=\mathbb{E}_{\boldsymbol{z},y}[CE(T(G(\boldsymbol{z}|y)),y)] \tag{7}$$

此外,全局模型也包含全局数据的统计信息. 该统计信息可通过全局模型的 BN(Batch Normalization)层[25]获取. 具体而言,BN 层记录了当前输入的统计信息(记为均值 μ 和方差 σ^2). 为利用该统计信息,本文希望数据生成器所生成数据经过 BN 层所统计的信息(记为均值 $\hat{\mu}$ 和方差 $\hat{\sigma}^2$)与全局模型相同位置 BN 层记录的统计信息(即均值 μ 和方差 σ^2)相近. 因此本文引入统计损失函数 $\mathcal{L}_{BNS}$ 来训练数据生成器. 令 B 表示模型中 BN 层的数量,$I_b(\mu_b,{\sigma_b}^2)$ 表示全局模型中第 b 个 BN 层所记录的统计信息表示的分布,$\hat{I}_b(\hat{\mu}_b,\hat{\sigma}_b{}^2)$ 表示生成数据经过 BN 层所统计的信息表示的分布. 本文通过计算两个分布 I_b 与 $\hat{I}_b$ 的 KL 散度 $KL(\cdot)$ 来获得统计损失函数 $\mathcal{L}_{BNS}$,即

$$\mathcal{L}_{BNS}(G)=\sum_{b=1}^{B}KL(\hat{I}_b\,\|\,I_b) \tag{8}$$

因此,训练数据生成器的总损失函数 $\mathcal{L}(G)$ 为

$$\mathcal{L}(G)=\mathcal{L}_{CE}(G)+\gamma\,\mathcal{L}_{BNS}(G) \tag{9}$$

其中,γ 是一个平衡因子. 通过优化目标函数(9),数据生成器从全局模型中获取全局数据分布信息,进而生成虚拟数据来构造类别均衡的数据集.

数据隐私保护分析:本文所提出的方法在服务器端和客户端之间仅传输全局模型,这与现有联邦学习方法[6-8]保持一致,因此并未增加泄露客户端数据隐私的风险. 在本文方法中,我们通过有效挖掘全局模型的全局数据分布信息(包括分类边界信息和统计信息),从而生成虚拟数据来解决客户端数据类别不均衡问题. 尽管所训练的生成器可以生成具有类别特征的相似图片,但往往与客户端真实图片存在较大差异(见图 11).

4.3.2　本地模型的训练

通过结合本地数据和生成的虚拟数据,各客户端可以构造类别均衡的数据集. 然而,由于虚拟数据和真实数据的差异,直接在虚拟数据上使用传统的交叉熵损失函数难以获得性能良好的模型. 如何更有效利用虚拟数据的信息来训练本地模型仍是一个开放的问题. 为解决该问题,本文通过全局模型提供丰富的虚拟数据信息来训练本地模型. 首先,全局模型的输出可作为软标签[26]. 软标签可以提供丰富的类别相似性信息. 其次,全局模型的中间层可以提供丰富的特征信息. 因此,本文引入一个蒸馏损失函数 $\mathcal{H}(\cdot,\cdot)$,以将全局模型中关于虚拟数据的知识迁移至本地模型. 本文通过优化以下目标函数训练本地模型:

$$\mathbb{E}_{x_k}[\mathcal{L}(S(x_k);\boldsymbol{W}_{S_k})]+\lambda\mathbb{E}_{\hat{x}_k}[\mathcal{H}(S(\hat{x}_k),T(\hat{x}_k))] \tag{10}$$

其中,x_k 与 $\hat{x}_k$ 分别表示第 k 个客户端的本地数据和生成的虚拟数据. $\mathcal{L}(\cdot;\cdot)$ 为交叉熵损失函数,$\mathcal{H}(\cdot,\cdot)$ 为蒸馏损失函数,λ 是一个平衡因子.

本文从两个方面设计蒸馏损失函数 $\mathcal{H}(\cdot,\cdot)$. 一方面,给定相同的输入,本地模型和全局模型的输出应当足够接近,以确保本地模型能够取得与全局模型相近的性能. 故本文通过最小化本地模型和全局模型输出之间的 KL 散度 $KL(\cdot)$ 来训练本地模型. 因此,本文引入如下损失函数 $\mathcal{L}_{KL}(S,T)$:

$$\mathcal{L}_{KL}(S,T)=KL(S(\hat{x})\,\|\,T(\hat{x})) \tag{11}$$

另一方面,考虑到全局模型的中间层也能为本地模型的训练提供丰富的特征信息,因此,参考文献[27],本文引入如下损失函数 $\mathcal{L}_{AT}$:

$$\mathcal{L}_{AT}(S,T)=\sum_{l=1}^{L}\left\|\frac{A_l^T}{\|F(A_l^T)\|_2}-\frac{A_l^S}{\|F(A_l^S)\|_2}\right\|_2 \tag{12}$$

其中,A_l^T 和 A_l^S 分别表示全局模型和本地模型的第 l 层特征值,$F(A_l)$ 表示基于第 l 层特征值的注意

力图[27]. L 表示在本地模型和全局模型之间传递知识的层数. $\|\cdot\|_2$ 表示欧几里得范数.

总的蒸馏损失函数如下所示. 其中,β 是一个平衡因子.

$$\mathcal{H}(S(\hat{x}),T(\hat{x}))=\mathcal{L}_{KL}(S,T)+\beta\mathcal{L}_{AT}(S,T) \tag{13}$$

4.4 模型聚合

联邦学习通过聚合各客户端本地模型的参数获得一个高性能全局模型. 值得一提的是,本文所提出的 CBFL 在客户端构造类别均衡的数据集来训练本地模型,使得服务器端更易于聚合这些性能良好且差异性小的本地模型. 参考 FedAvg[3],CBFL 以加权平均的方式聚合各本地模型的参数来更新全局模型,即

$$\boldsymbol{W}_T=\mathbb{E}_k[\boldsymbol{W}_{S_k}]=\sum\nolimits_{k\in\epsilon_r}\frac{N_k}{\sum_{k\in\epsilon_r}N_k}\boldsymbol{W}_{S_k} \tag{14}$$

通过公式(14),全局模型汇聚各客户端本地模型的信息. 此外,更新后的全局模型也将帮助数据生成器获得更准确的全局数据分布信息.

5 实 验

5.1 数据集

本文在四个标准数据集对所提出的方法进行验证,即 CIFAR-10[28]、CIFAR-100[28]、CINIC-10[29] 和 iNaturalist 2019①.

(1)CIFAR-10 由 60000 张 32×32 像素的彩色图像组成,共包含 10 个类别,每个类别包含 6000 张图像. 训练集与测试集分别包含 50000 张和 10000 张图像;

(2)CIFAR-100 与 CIFAR-10 的组成相似,但 CIFAR-100 包含了更多的类别. CIFAR-100 共包含 100 个类别,每个类别包含 500 张训练图像和 100 张测试图像;

(3)CINIC-10 包含 270000 张图像. 值得一提的是,该数据集的图像不一定来自相同的分布. 这个特性非常适合联邦学习,因为实际场景中各个客户端的数据不一定服从相同分布. CINIC-10 具有相同大小的训练集、验证集和测试集. 本文在训练集上进行模型训练,并在测试集上进行测试,而不使用该数据集的验证集;

(4)iNaturalist 2019 是一个大规模的类别分布高度不均衡的真实数据集. 该数据集包含 1010 个类别,总计 268243 张彩色图片,由 iNaturalist 网站的 2295 个用户收集和上传. 每个用户上传的图片被视为一个客户端的数据集.

5.2 实验设置

如无特殊说明,本文使用如下统一的实验设置. 其中,客户端数量 K 设为 100,通信轮数 R 设为 1000. 在每轮通信中,随机采样客户端的数量占客户端总数的比例 c 设为 10%,本地模型在客户端的训练轮数设为 5. 遵循文献[4,5],本文对 CIFAR-10、CIFAR-100 和 CINIC-10 数据集通过狄利克雷分布 $Dir(\alpha)$ 来模拟各个客户端数据的类别分布情况,其中参数 α 控制客户端上类别分布不均衡和客户端之间类别分布差异的程度. 具体而言,本文采样 $q_m\sim Dir(\alpha)$,并将类别为 m 的数据按 $q_{m,k}$ 的比例分配给第 k 个客户端. 值得注意的是,模型性能对客户端数据的类别分布十分敏感. 因此,本文提出的方法与所有对比算法均使用相同的类别分布以进行公平的比较.

5.3 实验环境

本文在一个分布式集群上,通过模拟联邦学习的方式进行实验. 集群的一个节点被视为服务器端,其它节点被视为客户端. 客户端节点在英伟达 TITAN Xp GPU 上训练模型. 由于客户端的数量过多而 GPU 的资源有限,多个客户端会被分配到同一块 GPU 上训练模型,但客户端节点之间不会进行数据和模型信息传输. 客户端节点和服务器端节点之间通过网络进行通信.

5.4 实施细节

本文使用 SGD 优化器来训练本地模型,初始学习率设为 0.1,学习率衰减设为 0.996. 参考文献[4,5],本文在训练过程中对图像进行数据增强(随机裁剪和翻转)与归一化处理. 此外,本文参考 ACGAN[16] 构建数据生成器,并将随机噪声 $\boldsymbol{z}$ 的维度设为 100. 在每轮通信中,数据生成器采用 *Adam* 优化器进行训练,学习率设为 0.001,训练轮数设为 2000. 平衡因子 λ、γ、β 分别设为 1、10、400. 对于数据集 CIFAR-10、CIFAR-100、CINIC-10 和 iNaturalist 2019,预热轮数 R_w 分别设为 700、800、500 和 700. 所有实现均基于 *PyTorch* 深度学习框架.

5.5 与最新方法的比较

本文将 CBFL 与目前最新的方法进行比较,即 FedAvg[3]、FedProx[6]、SCAFFOLD[7] 和 FedNova[8]. 遵循 FedProx,本文将 FedProx 中 Proximal 项的权重设为 0.001. 本实验通过狄利克雷分布 $Dir(0.1)$ 模

① iNaturalist2019. https:sites. google. com/view/fgvc6/competitions/ina turalist-2019.

拟各客户端数据的类别分布. 所有方法基于两种常见的深度神经网络模型(即 ResNet20[1] 和 MobileNetV2[2])进行训练和性能比较.

由表 1 可知,SCAFFOLD 在所有数据集和神经网络模型上均表现最差. 可能的原因是每轮通信中服务器端随机采样客户端的比例比较低(仅为 10%),客户端的控制变量更新频率也就变得比较低. 此时使用客户端和服务器端的控制变量来估计本地模型更新方向的修正量非常不准确,导致最终性能严重受损. FedProx 通过 Proximal 项使本地模型尽量靠近全局模型,在某些情况下(如 CIFAR-10 数据集上训练的 MobileNetV2)略优于 FedAvg. 然而,该方法仍然只使用客户端不均衡的本地数据来训练本地模型,导致其取得的性能提升有限,甚至在某些情况下性能不如 FedAvg(如 CIFAR-100 数据集上训练的 ResNet20). FedNova 在 ResNet20 和 MobileNetV2 上的性能均不如 FedAvg. 相比之下,CBFL 在所有数据集和深度神经网络模型上均优于现有方法. 具体而言,在 CIFAR-10 与 CIFAR-100 数据集上训练 ResNet20,CBFL 所取得的准确率比其它方法所取得的最高准确率分别高出了 1.74% 和 1.11%. 类似地,CBFL 在 CINIC-10 数据集上训练的 ResNet20 和 MobileNetV2 模型的准确率比其它方法所取得的最高准确率仍高出了 1.80% 和 2.73%. 这些结果证明了所提出的 CBFL 相对于现有方法的优越性.

本节进一步展示训练过程中本地模型和全局模型在测试集上性能的变化曲线. 实验结果如图 6 和图 7 所示. 由于 SCAFFOLD 的表现欠佳,为方便实验与作图,本节不考虑与 SCAFFOLD 进行比较. 图

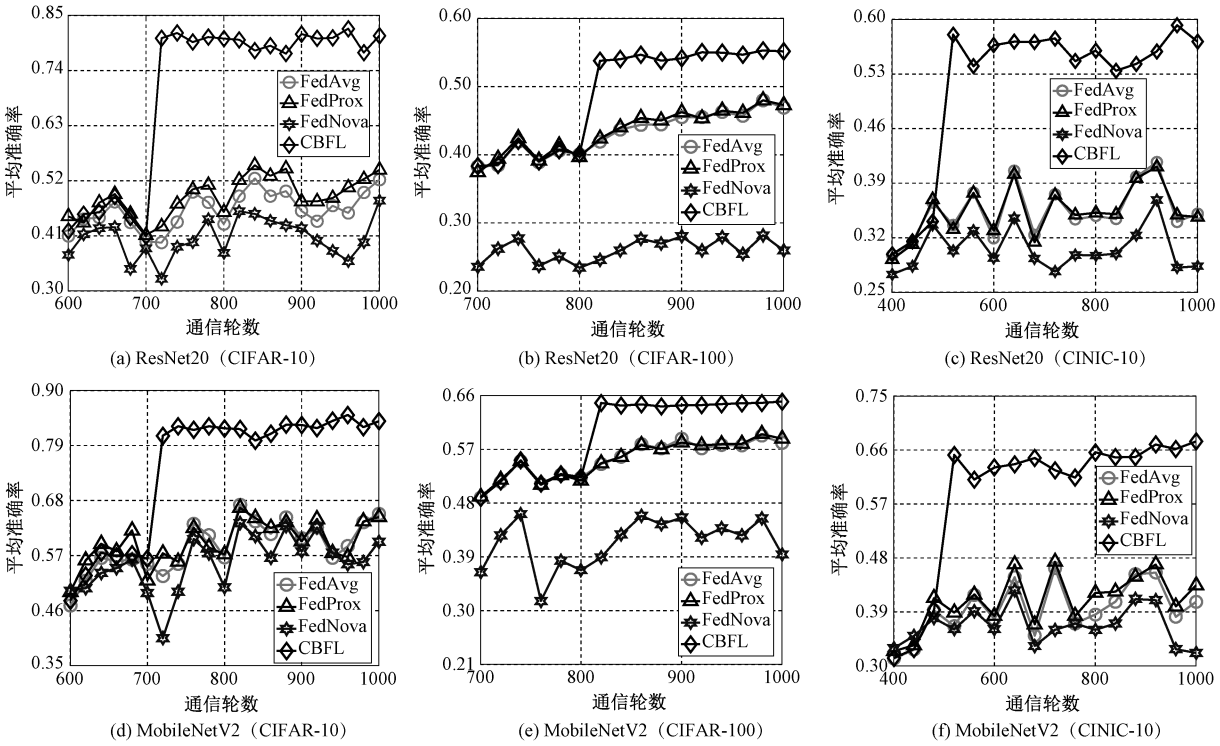


图 6　本地模型 ResNet20 和 MobileNetV2 在 CIFAR-10、CIFAR-100 和 CINIC-10 数据集的平均测试准确率曲线

6 展示了本地模型 ResNet20 和 MobileNetV2 在 CIFAR-10、CIFAR-100 和 CINIC-10 三个数据集的性能变化曲线. 该曲线的纵坐标为每轮通信中所有参与训练的本地模型在测试集的平均准确率. 与 FedAvg 相比,FedProx 可以提升本地模型的性能,但提升的性能有限. 然而,本文所提出的 CBFL 在所有数据集和神经网络模型上都相比于现有方法极大地提升了本地模型的性能. 如图 6 所示,当通信轮次大于预热轮数时,CBFL 在 CIFAR-10 数据集训练的本地模型 ResNet20 比现有方法高出了 20% 左右的平均准确率. 此外,在 CINIC-10 数据集,CBFL 训练的本地模型 ResNet20 和 MobileNetV2 均比现有方法高出了 15%左右的平均准确率. 值得注意的是,这些性能良好的本地模型有利于进一步提升全局模型的性能(见表 1 和图 7).

图 7 展示了全局模型 ResNet20 和 MobileNetV2 在 CIFAR-10、CIFAR-100 和 CINIC-10 三个数据集的测试准确率变化曲线. FedAvg、FedProx 和 FedNova 的全局模型准确率曲线在训练过程中非常不稳定. 相反,本文所提出的 CBFL 能较

好地稳定全局模型的训练. 当通信轮次大于预热轮数时,CBFL 的全局模型准确率曲线振荡幅度大大减小. 此外,CBFL 训练的全局模型收敛得更快且获得更高的性能. 这表明,本文所提出的 CBFL 能在客户端上类别分布高度不均衡且客户端之间类别分布差异巨大的情况下,有效改善全局模型的训练过程.

表 1 与最新方法在 CIFAR-10、CIFAR-100 与 CINIC-10 数据集上的性能比较

模型	数据集	测试准确率/%				
		FedAvg[3]	FedProx[6]	SCAFFOLD[7]	FedNova[8]	CBFL
ResNet20	CIFAR-10	84.13	84.15	19.63	82.55	**85.89**
	CIFAR-100	59.13	59.01	48.85	55.88	**60.24**
	CINIC-10	70.50	70.62	16.64	64.93	**72.42**
MobileNetV2	CIFAR-10	86.34	86.70	10.00	85.81	**87.99**
	CIFAR-100	66.18	66.22	45.18	66.60	**66.90**
	CINIC-10	75.35	75.19	10.06	75.26	**78.08**

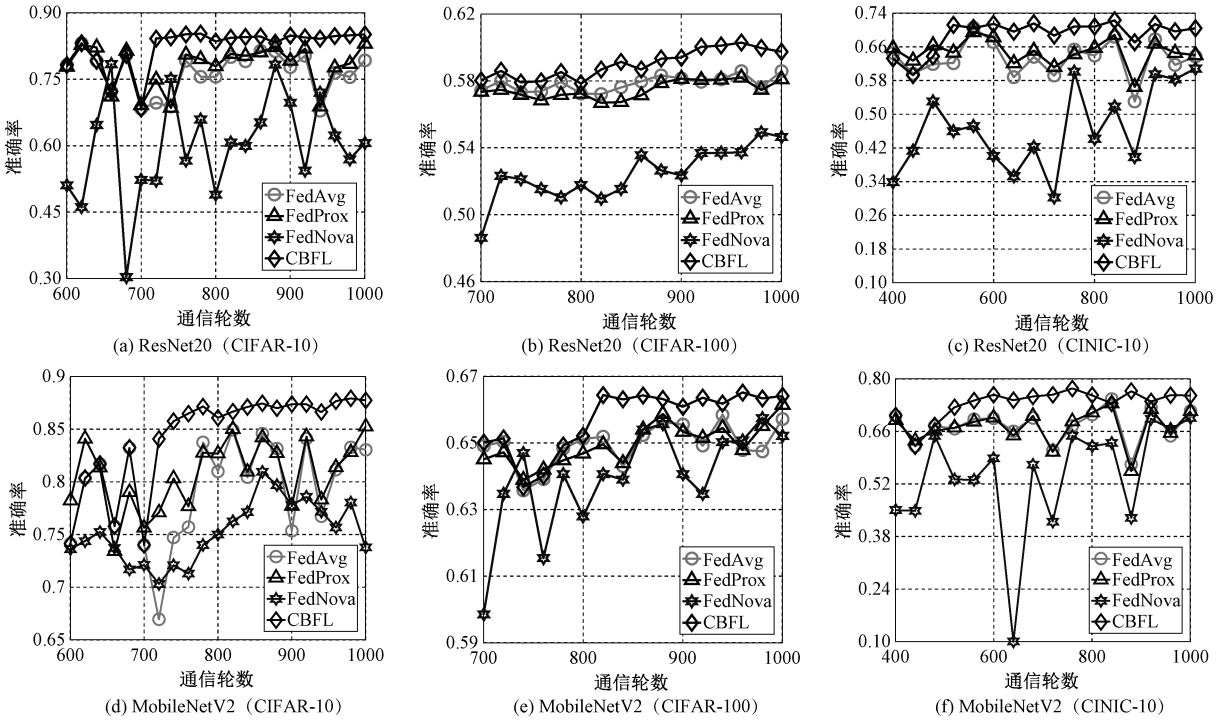


图 7 全局模型 ResNet20 和 MobileNetV2 在 CIFAR-10、CIFAR-100 和 CINIC-10 数据集的测试准确率曲线

5.6 在不同类别分布上的性能比较

本节研究客户端上不同类别分布对联邦学习算法性能的影响. 本实验使用狄利克雷分布 $Dir(\alpha)$ 模拟客户端数据的类别分布,其中参数 $\alpha > 0$ 控制客户端上类别分布不均衡和客户端之间类别分布差异的程度. 本节选取了 5 个不同的 α 值来进行实验,即 0.1、0.07、0.05、0.03 和 0.01. 随着 α 值的减小,客户端数据越来越集中于少数的几个类,故其类别分布变得越来越不均衡且客户端之间的类别分布差异变得越来越大,从而给联邦学习的模型训练造成更大的困难. 本实验基于 ResNet20 模型,在 CIFAR-10、CIFAR-100 和 CINIC-10 三个数据集上进行. 由表 1 可知,SCAFFOLD 在客户端采样频率很低的场景下表现欠佳,故本节不考虑与 SCAFFOLD 进行比较.

实验结果如表 2 所示. 本文所提出的 CBFL 在客户端不同的类别分布上始终优于现有方法. 更重要的是,当 α 变小时,CBFL 比现有方法取得更大的性能提升. 例如,当 α 设为 0.01 时,CBFL 在 CIFAR-100 数据集上比目前最好的方法高出了 5.82% 的准确率. 在 CINIC-10 数据集上,当 α 设为 0.03 和 0.01 时,CBFL 比目前最好的方法也分别高出了 4.39% 和 4.74% 的准确率. 这说明,当客户端数据的类别分布变得越来越不均衡且客户端之间的类别分布差异变得越来越大时,CBFL 比现有方法具有更大的优势. 这些结果证明了 CBFL 能有效缓解联邦学习中因客户端类别分布高度不均衡且客户端之间类别分布差异巨大造成的性能下降问题.

表 2　与最新方法在不同类别分布上的性能比较

数据集	方法	测试准确率/%				
		$\alpha = 0.1$	$\alpha = 0.07$	$\alpha = 0.05$	$\alpha = 0.03$	$\alpha = 0.01$
CIFAR-10	FedAvg[3]	84.13	81.60	79.00	72.87	41.21
	FedProx[6]	84.15	81.37	78.98	72.19	41.05
	FedNova[8]	82.55	78.58	77.04	69.08	48.25
	CBFL	**85.89**	**83.04**	**81.99**	**75.49**	**51.02**
CIFAR-100	FedAvg[3]	59.13	58.00	56.90	53.21	40.39
	FedProx[6]	59.01	58.08	56.45	53.92	41.21
	FedNova[8]	55.88	53.64	52.50	47.26	33.27
	CBFL	**60.24**	**58.47**	**58.03**	**55.29**	**47.03**
CINIC-10	FedAvg[3]	70.50	66.51	60.38	54.74	37.17
	FedProx[6]	70.62	66.26	61.11	53.79	38.69
	FedNova[8]	64.93	61.97	58.03	47.71	37.43
	CBFL	**72.42**	**68.74**	**65.28**	**59.13**	**43.43**

5.7　在真实数据集上的性能比较

本节在 iNaturalist 2019 数据集上探究所提出 CBFL 在真实的类别分布高度不均衡场景上的性能. 由于 SCAFFOLD 在客户端采样频率很低的场景下表现欠佳,故本节不考虑与 SCAFFOLD 进行比较. 本实验基于两种常见的深度神经网络模型(即 ResNet18 和 MobileNetV2)进行训练和性能比较.实验结果如表 3 所示. FedNova 训练的 ResNet18 和 MobileNetV2 模型均表现欠佳,这说明 FedNova 可能不适用于真实复杂的联邦学习场景. 本文提出的 CBFL 训练的 ResNet18 和 MobileNetV2 模型准确率比其它方法所取得的最高准确率仍分别高出了 1.19%和 0.99%. 这些结果证明了所提出的 CBFL 在真实场景上的有效性.

表 3　不同方法在 iNaturalist 2019 数据集上的性能比较

模型	测试准确率/%			
	FedAvg[3]	FedProx[6]	FedNova[8]	CBFL
ResNet18	15.24	15.30	0.21	**16.49**
MobileNetV2	20.25	20.01	0.33	**21.24**

5.8　算法适用场景分析

FedAvg 是联邦学习的经典算法,但该算法难以解决联邦学习面临的两个重要挑战,即客户端数据类别不均衡和客户端之间数据分布差异巨大. 现有方法(如 FedProx,SCAFFOLD,FedNova)对经典的 FedAvg 方案进行了改进. 然而,FedProx 和 SCAFFOLD 仅仅考虑在 EMNIST 数据集[30]上,采用浅层的网络结构进行实验. FedNova 虽然考虑在 CIFAR-10 数据集上进行实验,但也仅仅在 VGG-11[31]模型上被证明有效. 本文考虑了更复杂的数据集,包括 CIFAR-10, CIFAR-100, CINIC-10 以及 iNaturalist 2019, 并且在常见的深度神经网络结构(即 ResNet 和 MobileNetV2)上进行验证. 实验结果验证了所提出的 CBFL 相对于现有方法的优越性和在真实场景中的实用性.

6　进一步讨论与分析

6.1　数据生成器和类别均衡采样器对性能的影响

本节探究数据生成器和类别均衡采样器对算法性能的影响. 本实验基于 ResNet20 模型,在 CIFAR-10 数据集上进行. 当不使用所提出的类别均衡采样器时,生成的虚拟数据采取均匀采样的方式. 实验结果如表 4 所示. 当仅使用数据生成器时,由于所生成的虚拟数据使得客户端之间的数据分布差异变小,最终模型的准确率从 84.13%提升至 85.49%. 当进一步使用类别均衡采样器时,由于类别均衡采样器使得客户端的数据类别变得更均衡,最终模型的准确率从 85.49%进一步提升至 85.89%. 这些结果表明所提出的数据生成器和类别均衡采样器均能有效提升算法性能,且数据生成器对性能提升的贡献更大.

表 4　类别均衡采样器和数据生成器对性能的影响

数据生成器	类别均衡采样器	准确率/%
×	×	84.13
√	×	85.49
√	√	85.89

6.2　不同生成器对性能的影响

本节探究不同生成器对性能的影响. 本实验采用 MobileNetV2 模型,在 CIFAR-10、CIFAR-100 和 CINIC-10 三个数据集上进行. 同时,本节考虑了以下两种对比方法:(1)从均值为 0,方差为 1 的高斯分布中采样随机噪声,以替代所生成的虚拟数据;

(2)使用 DCGAN[15] 模型替代所提出的数据生成器. 具体而言,该方法使各客户端都基于本地数据训练一个 DCGAN 模型,并分享该模型给其它客户端以生成图像. 如表 5 所示,简单地使用随机噪声来训练本地模型并不能取得良好的效果. 例如,在 CIFAR-10 和 CIFAR-100 数据集上,与不使用数据生成器的方法(即 FedAvg)相比,利用随机噪声来辅助训练会损害模型性能. 这是因为,随机噪声与真实的图像分布相差较大,无法为本地模型的训练提供有用的信息. 使用 DCGAN 模型可以略微提升模型性能. 然而,由于客户端数据量有限且客户端数据的类别分布高度不均衡,训练一个性能良好的 DCGAN 模型极其困难. 这意味着 DCGAN 生成的图像质量可能很差,难以有效改善本地模型的训练. 相反,本文提出的数据生成器从全局模型中获取全局数据分布信息,从而生成可用于本地模型训练的高质量图像. 由表 5 可知,本文所提出的数据生成器在所有数据集上均取得最高的准确率. 这些结果证明了所提出数据生成器的有效性.

表 5 不同生成器对性能的影响

生成器	测试准确率/%		
	CIFAR-10	CIFAR-100	CINIC-10
无生成器(FedAvg)	86.34	66.18	75.35
随机高斯噪声	85.78	65.49	75.44
DCGAN[15]	86.59	66.39	75.95
数据生成器(本文)	**87.90**	**66.90**	**78.08**

6.3 不同类别采样器对性能的影响

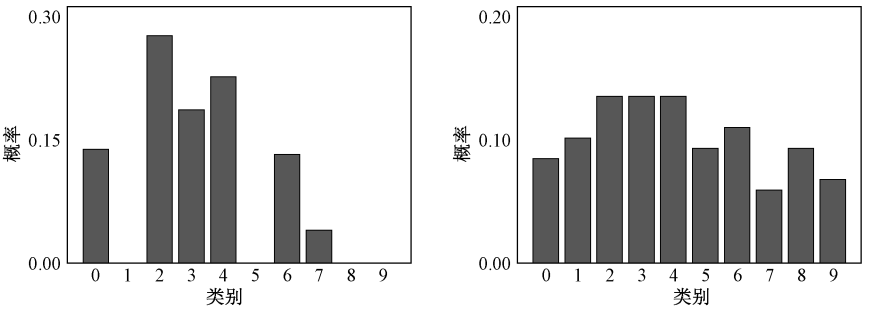
本节探究不同类别采样器对性能的影响. 本实验采用 ResNet20 模型,在 CINIC-10 数据集上进行. 同时,本节考虑以下基线类别采样器作为对比:只对客户端前 C 类本地数据量较少的类别进行采样. 实验结果如表 6 所示. 与仅使用客户端本地数据训练本地模型的 FedAvg 方法相比,通过采样所生成的虚拟数据来辅助本地模型训练的方式能获得更高的准确率. 此外,当越来越多类别的数据被采样,即当 C 变大时,客户端数据的类别分布将逐渐变得均衡,模型性能也逐渐变得更好. 本文所提出的类别均衡采样器始终取得最好的性能. 同时,本节也展示了经类别分布均衡器后客户端数据的类别分布,如图 8 所示. 可以看到,本文所提出的类别均衡采样器能够帮助客户端均衡其数据的类别分布. 这些结果证明了所提出类别均衡采样器的有效性.

6.4 数据生成和蒸馏损失函数对性能的影响

本节探讨数据生成和蒸馏损失函数对模型性能的影响. 本实验基于 ResNet20 模型,在 CIFAR-10 数据集上进行. 值得一提的是,本文仅在所生成的虚拟数据上使用蒸馏损失函数,当不采用数据生成技术时,无法使用蒸馏损失函数进行训练. 当仅采用数据生成技术但不添加蒸馏损失函数时,本实验使用交叉熵损失函数进行训练. 实验结果如表 7 所示. 当仅采用数据生成技术时,本地模型平均准确

表 6 不同类别采样器对性能的影响

类别采样器	无类别采样器(FedAvg)	$C=2$	$C=4$	$C=6$	$C=8$	$C=10$(类别均衡采样器)
准确率/%	70.50	71.18	72.17	72.28	72.32	**72.42**



(a) 客户端本地数据的类别分布 (b) 经类别均衡采样器后客户端数据的类别分布

图 8 客户端本地数据的类别分布和经类别均衡采样器后客户端数据的类别分布的比较

率从 54.67%提升至 73.63%,全局模型准确率从 84.13%提升至 84.97%. 这表明所生成数据能有效降低客户端数据类别不均衡的影响. 当进一步增加蒸馏损失函数时,本地模型平均准确率从 73.63%提升至 81.34%,全局模型准确率从 84.97%提升至 85.89%. 因此,与传统的交叉熵损失函数相比,使用蒸馏损失函数可以让全局模型提供更丰富的关于虚拟数据的知识来训练模型. 综上,所提出的数据生成技术和蒸馏损失函数均能有效提升全局模型性能,但蒸馏损失函数是为了更有效利用虚拟数据进

行训练,数据生成对性能提升的贡献更大.

表 7　数据生成和蒸馏损失函数对性能的影响

数据生成	蒸馏损失函数	本地模型平均准确率/%	全局模型准确率/%
×	×	54.67	84.13
√	×	73.63	84.97
√	√	81.34	85.89

6.5　不同超参数取值对性能的影响

为探究公式(10)中超参数 λ 的不同取值对性能的影响,我们采用 ResNet20 模型,在 CIFAR-100 数据集进行了实验. 本实验选取了 5 个不同的 λ 值,即 0.01、0.1、1、5 和 10. 实验结果如表 8 所示. 当 λ 从 0.01 增加到 1 时,蒸馏损失函数 $\mathcal{H}(\cdot,\cdot)$ 逐渐变得重要,从而能为本地模型提供更多关于虚拟数据的知识,使得模型取得更好性能. 然而,如果 λ 进一步增加到 5 甚至是 10 时,蒸馏损失函数 $\mathcal{H}(\cdot,\cdot)$ 将成为总损失函数的主导项,最终阻碍模型性能的提升. 因此,本文实验设定 $\lambda=1$.

为探究公式(13)中超参数 β 的不同取值对性能的影响,我们采用 ResNet20 模型,在 CIFAR-100 数据集进行了实验. 本实验选取了 5 个不同的 β 值,即 0、200、400、600 和 800. 实验结果如表 9 所示. 当 β 设为 400 时,损失函数 $\mathcal{L}_{KL}$ 和 $\mathcal{L}_{AT}$ 之间可以实现更好的权衡,从而取得更好的算法性能.

表 8　超参数 λ 的不同取值对性能的影响

λ	0.01	0.1	1	5	10
准确率/%	59.28	59.45	**60.24**	60.03	59.24

表 9　超参数 β 的不同取值对性能的影响

β	0	200	400	600	800
准确率/%	59.36	59.83	**60.24**	59.97	59.63

6.6　不同预热轮数取值对性能的影响

本节探究预热轮数 R_w 的不同取值对性能的影响. 本实验采用 ResNet20 模型,在 CIFAR-10 数据集进行. 本实验从[0, 900]区间均匀选取了 10 个不同的 R_w 值. 实验结果如图 9 所示. 当 R_w 取值为 0 时,此时全局模型尚未经过训练,不能获得准确的全局数据分布信息. 因此,利用该全局模型训练的数据生成器难以生成高质量数据,从而影响本地模型的训练并损害全局模型性能. 当 R_w 取值从 0 增加至 700 时,数据生成器能够从全局模型获得更准确的全局数据分布信息,从而生成更高质量的虚拟数据来训练本地模型,进而有利于提高全局模型的性能. 当 R_w 取值为 700 时,所训练的全局模型取得最好的性能. 当 R_w 的取值进一步增加到 800 和 900 时,模型性能的变化不明显.

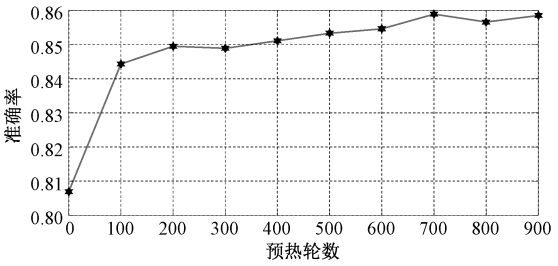


图 9　预热轮数 R_w 的不同取值对性能的影响

6.7　客户端数量对性能的影响

为探究客户端数量对联邦学习算法性能的影响,我们采用 ResNet20 模型,在 CIFAR-10 数据集上进行了实验. 本实验采用狄利克雷分布 $Dir(0.1)$ 来模拟客户端的类别分布,客户端数量的取值范围为 $K\in\{100,150,200,250,300\}$. 同时,本节也将所提出的 CBFL 与现有方法进行比较. 所有方法均采用相同的实验设置以进行公平的比较. 如表 10 所示,客户端数量的增加会损害联邦学习算法的性能. 然而,与现有方法相比,本文提出的 CBFL 对于所有的 K 值,始终能取得最好的性能. 这表明本文提出的 CBFL 方法在大规模联邦学习场景上仍然能比现有方法取得更好的性能. 此外,本节分别设置 5 和 16 个客户端,在真实环境上进行实际验证,每个客户端在一块独立的英伟达 TITAN Xp GPU 上进行训练. 所有客户端在每轮通信中都参与训练(即客户端采样率为 100%). 实验结果如表 11 所示. 此时,SCAFFOLD 可取得与 FedNova 相近的性能,这表明该方法在客户端采样率高的场景上能够取得较好的性能. 本文所提出的 CBFL 在两种场景上均取得最好性能.

表 10　客户端数量 K 对性能的影响

方法	测试准确率/%				
	$K=100$	$K=150$	$K=200$	$K=250$	$K=300$
FedAvg[3]	84.13	80.82	78.90	77.56	77.51
FedProx[6]	84.15	79.97	78.92	77.28	77.49
FedNova[8]	82.55	79.27	76.32	74.55	75.43
CBFL	**85.89**	**82.59**	**80.05**	**78.57**	**78.92**

表 11　不同方法在少量客户端场景上的准确率比较

客户端数量	FedAvg[3]	FedProx[6]	FedNova[8]	SCAFFOLD[7]	CBFL
$K=5$	82.61%	84.10%	82.15%	81.84%	**87.22%**
$K=16$	88.13%	88.47%	85.15%	86.37%	**89.31%**

6.8　通信轮数对性能的影响

如无特殊说明,本文所有实验的通信轮数 R 均

设为1000. 这种做法主要考虑到联邦学习系统中高昂的通信成本. 在实际应用场景中,特别是在全局模型参数量较大的情况下,联邦学习系统往往难以建立足够多的通信回合. 此外,由于各客户端数据严重不均衡且客户端之间的数据分布差异巨大,现有联邦学习算法往往难以收敛.

为探究通信轮数对算法性能的影响,我们在CIFAR-100数据集上进行了实验,通信轮数由1000延长至2000. 实验结果如图10所示. 当进一步延长通信轮数时,所有方法的全局模型和本地模型准确率会继续上升. 这是因为本文采用文献[3]的学习率下降方案,每轮通信的学习率衰减设为0.996. 即使到了第2000轮通信,学习率仍然不为零(约为0.000033),故此时模型的测试准确率有可能继续上升. 然而,本文所提出的CBFL始终比现有方法取得更高的准确率. 此外,CBFL训练的全局模型达到目标准确率所需要的通信轮数低于现有方法. 如表12所示,当ResNet20模型达到目标准确率(60%)时,CBFL仅需要902轮通信,远远低于FedAvg(1586)和FedProx(1765). 当MobileNetV2模型达到目标准确率(66%)时,CBFL仅需要807轮通信,远远低于FedProx(888)和FedNova(906). 这证明了CBFL比现有方法具有更快的收敛率.

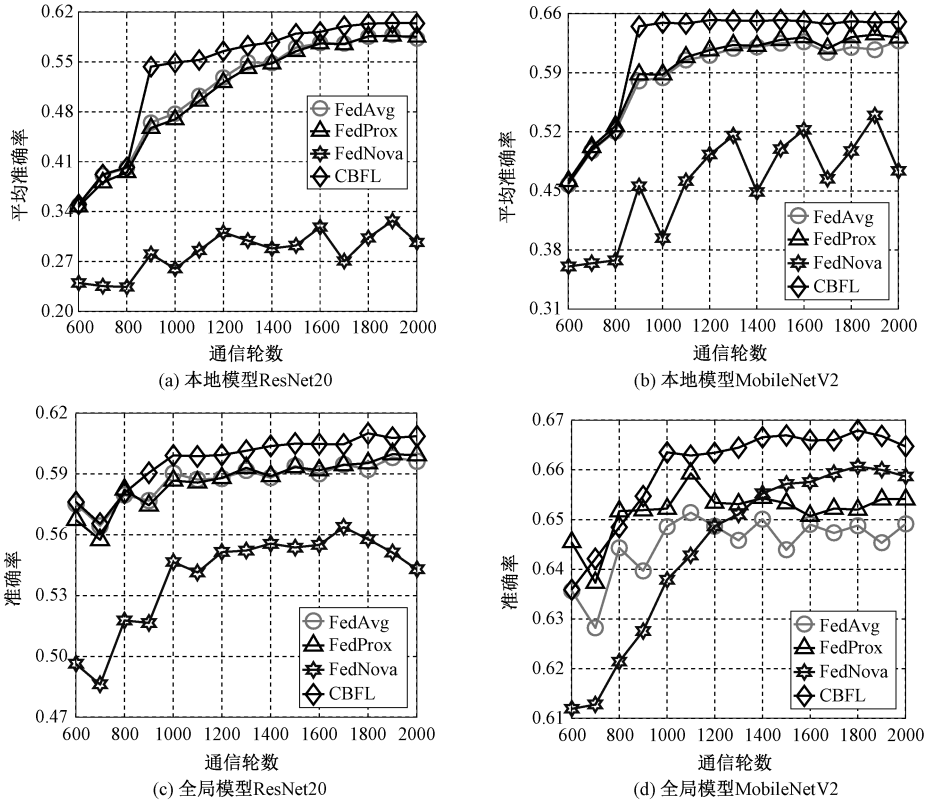


图10 本地模型和全局模型在CIFAR-100的测试准确率曲线

表12 不同方法达到目标准确率所需的通信轮数比较

模型	FedAvg[3]	FedProx[6]	FedNova[8]	CBFL
ResNet20	1586	1765	N/A*	**902**
MobileNetV2	N/A*	888	906	**807**

注:* 该方法无法达到目标准确率

6.9 实验效率分析

为测试不同联邦学习算法的实验效率,我们采用ResNet20(参数量约为0.3M),在CIFAR-100上进行实验. 本节使用训练时间和通信成本来衡量联邦学习算法的实验效率. 训练时间指模型训练过程的总耗时. 通信成本指模型取得最高准确率所需要的通信总开销,其包括客户端上传和下载模型两部分的通信量. 实验结果如表13所示. FedNova所需的通信成本在所有方法中最低,但其获得的准确率远低于其它方法. 本文所提出的CBFL所需的通信成本低于FedAvg和FedProx,这是因为CBFL具有更快的收敛率(见图7). 此外,由于需要额外对数据生成器进行训练,CBFL比其它方法消耗更长的训练时间. 然而,CBFL获得最高的准确率.

表 13　与最新方法的实验效率比较

方法	准确率/%	训练时间/h	通信成本/M
FedAvg[3]	59.13	4.1	584.4
FedProx[6]	59.01	6.0	582.0
FedNova[8]	55.88	4.1	466.8
CBFL	60.24	9.0	565.8

6.10　生成数据的可视化结果

本节可视化在 CINIC-10 数据集上的生成图片. 如图 11 所示，所提出的生成器确实可以生成给定类别对应的图片，但所生成图片与真实图片存在较大差异. 这是因为，本文训练的生成器仅利用全局模型的全局数据分布信息来辅助生成虚拟数据，而非恢复客户端数据，因此不存在隐私泄露.

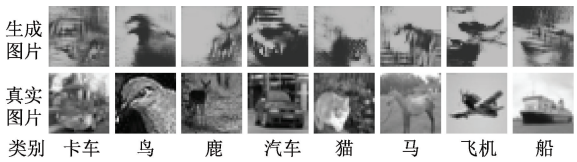


图 11　在 CINIC-10 数据集上生成图片的可视化

7　结　论

本文提出了一种基于数据生成的类别均衡联邦学习(CBFL)方法. CBFL 针对各客户端构造类别均衡的数据集，以降低客户端类别不均衡和客户端之间分布差异的影响. 具体而言，CBFL 设计了一个类别分布均衡器，其由一个类别均衡采样器和一个数据生成器组成. 其中，类别均衡采样器以较高概率采样客户端本地数据量不足的类别，然后，数据生成器根据所采样的类别生成相应的虚拟数据. 结合本地数据和虚拟数据，客户端构造类别均衡的数据集来进行模型训练，从而有利于构建高性能全局模型. 本文在四个标准数据集上进行了大量实验，证明了 CBFL 相对于现有方法的优越性.

致　谢　感谢鹏城云脑为本工作提供计算资源.

本文所有方法的实现代码开源于：https://github.com/lizhipengcs/CBFL.

参　考　文　献

[1] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA, 2016: 770-778

[2] Sandler M, Howard A, Zhu M, et al. Mobilenetv2: Inverted residuals and linear bottlenecks//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA, 2018: 4510-4520

[3] McMahan B, Moore E, Ramage D, et al. Communication-efficient learning of deep networks from decentralized data//Proceedings of the International Conference on Artificial Intelligence and Statistics. Fort Lauderdale, USA, 2017: 1273-1282

[4] Wang H, Yurochkin M, Sun Y, et al. Federated learning with matched averaging//Proceedings of the International Conference on Learning Representations. Addis Ababa, Ethiopia, 2020: 1-16

[5] Hsu T H, Qi H, Brown M. Measuring the effects of non-identical data distribution for federated visual classification. arXiv preprint arXiv: 1909.06335, 2019

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

[7] Karimireddy S P, Kale S, Mohri M, et al. Scaffold: Stochastic controlled averaging for federated learning//Proceedings of the International Conference on Machine Learning. Vienna, Austria, 2020: 5132-5143

[8] Wang J, Liu Q, Liang H, et al. Tackling the objective inconsistency problem in heterogeneous federated optimization//Proceedings of the Advances in Neural Information Processing Systems. Vancouver, Canada, 2020: 7611-7623

[9] Dinh C T, Tran N, Nguyen T D. Personalized federated learning with moreau envelopes//Proceedings of the Advances in Neural Information Processing Systems. Vancouver, Canada, 2020: 21394-21405

[10] Fallah A, Mokhtari A, Ozdaglar A. Personalized federated learning with theoretical guarantees: a model-agnostic meta-learning approach//Proceedings of the Advances in Ne-ural Information Processing Systems. Vancouver, Canada, 2020: 3557-3568

[11] Hanzely F, Hanzely S, Horvath S, et al. Lower bounds and optimal algorithms for personalized federated learning//Proceedings of the Advances in Neural Information Processing Systems. Vancouver, Canada, 2020: 2304-2315

[12] Mohri M, Sivek G, Suresh S. Agnostic federated learning//Proceedings of the International Conference on Machine Learning. Long Beach, USA, 2019: 4615-4625

[13] Reisizadeh A, Farnia F, Pedarsani R, et al. Robust federated learning: the case of affine distribution shifts//Proceedings of the Advances in Neural Information Processing Systems. Vancouver, Canada, 2020: 21554-21565

[14] Deng Y, Kamani M M, Mahdavi M. Distributionally robust federated averaging//Proceedings of the Advances in Neural Information Processing Systems. Vancouver, Canada, 2020: 15111-15122

[15] Radford A, Metz L, Chintala S. Unsupervised representation learning with deep convolutional generative adversarial networks

//Proceedings of the International Conference on Learning Representations. San Juan, Puerto Rico, 2016: 1-16

[16] Odena A, Olah C, Shlens J. Conditional image synthesis with auxiliary classifier gans//Proceedings of the International Conference on Machine Learning. Sydney, Australia, 2017: 2642-2651

[17] Feng Yong, Zhang Chun-Ping, Qiang Bao-Hua, et al. GP-WIRGAN: A novel image recurrent generative adversarial network model based on wasserstein and gradient penal. Chinese Journal of Computers, 2020, 43(2): 190-205 (in Chinese)
(冯永, 张春平, 强保华等. GP-WIRGAN:梯度惩罚优化的Wasserstein图像循环生成对抗网络模型. 计算机学报, 2020, 43(2): 190-205)

[18] Xiao Jin-Sheng, Shen Meng-Yao, Lei Jun-Feng, et al. Image conversion algorithm for haze scene based on generative adversarial networks. Chinese Journal of Computers, 2020, 43(1): 165-176 (in Chinese)
(肖进胜, 申梦瑶, 雷俊锋等. 基于生成对抗网络的雾霾场景图像转换算法. 计算机学报, 2020, 43(1): 165-176)

[19] Fredrikson M, Jha S, Ristenpart T. Model inversion attacks that exploit confidence information and basic countermeasures //Proceedings of the Conference on Computer and Communications Security. Denver, USA, 2015:1322-1333

[20] Mahendran A, edaldi A V. Visualizing deep convolutional neural networks using natural pre-images. International Journal of Computer Vision, 2016, 120(3): 233-255

[21] Micaelli P, Storkey A J. Zero-shot knowledge transfer via adversarial belief matching//Proceedings of the Advances in Neural Information Processing Systems. Vancouver, Canada, 2019: 9547-9557

[22] Nayak G K, Mopuri K R, Shaj V, et al. Zero-shot knowledge distillation in deep networks//Proceedings of the International Conference on Machine Learning. Long Beach, USA, 2019: 4743-4751

[23] Jonathon B, Zachary L. What is the effect of importance weighting in deep learning?//Proceedings of the International Conference on Machine Learning. Long Beach, USA, 2019: 872-881

[24] Zhou B, Cui Q, Wei X, et al. BBN: bilateral-branch network with cumulative learning for long-tailed visual recognition//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Seattle, USA, 2020: 9716-9725

[25] Ioffe S, Szegedy C. Batch normalization: Accelerating deep network training by reducing internal covariate shift//Proceedings of the International Conference on Machine Learning. Lille, France, 2015:448-456

[26] Hinton G, Vinyals O, Dean J. Distilling the knowledge in a neural network. arXiv preprint arXiv: 1503.02531, 2015

[27] Zagoruyko S, Komodakis N. Paying more attention to attention: Improving the performance of convolutional neural networks via attention transfer//Proceedings of the International Conference on Learning Representations. Toulon, France, 2017: 1-13

[28] Krizhevsky A, Hinton G. Learning multiple layers of features from tiny images. Handbook of Systemic Autoimmune Diseases, 2009, 1(4): 1-58

[29] Darlow L N, Crowley E J, Antoniou A, et al. CINIC-10 is not imagenet or CIFAR-10. arXiv preprint arXiv: 1810.03505, 2018

[30] Cohen G, Afshar S, Tapson J, et al. Emnist: extending mnist to handwritten letters//Proceedings of the International Joint Conference on Neural Network, Anchorage, USA, 2017:2921-2926

[31] Karen S, Andrew Z. Very deep convolutional networks for large-scale image recognition//Proceedings of the International Conference on Learning Representations. San Diego, USA, 2015: 1-14

LI Zhi-Peng, M. S. candidate. His research interests include deep learning and computer vision.

GUO Yong, Ph. D. candidate. His research interest focuses on deep learning.

CHEN Yao-Fo, Ph. D. candidate. His research interest focuses on deep learning.

WANG Yao-Wei, Ph. D. His research interest focuses on computer vision.

ZENG Wei, Ph. D. His research interest focuses on computer vision.

TAN Ming-Kui, Ph. D., professor, Ph. D. supervisor. His research interest focuses on machine learning.

Background

Currently, an unprecedented amount of data is distributed on extensive terminal devices. However, due to privacy concerns, it is infeasible to perform conventional centralized training by gathering the whole data from the terminal devices. To solve this problem, Federated Learning (FL) has emerged as a new paradigm to learn a shared global model without sharing the data on the terminal devices (also called clients). However, there are still two limitations to existing FL mechanism.

First, the global model needs to consider the data on

multiple clients, but each client contains only partial classes of data and the data amount of different classes is severely imbalanced, making it difficult to train the global model. Specifically, most data on the client belong to a few classes, while other classes have few or no data. As a result, the trained local models tend to overfit the data on the client and achieve poor performance on global data, which seriously affects the training of the global model. Therefore, how to reduce the impact of the imbalanced class distribution on the client to obtain a superior global model becomes an important problem.

Second, the data distribution is extremely different across the clients, making it hard to derive a good-performance global model. In fact, due to the differences in the application scenarios of the terminal devices, the data distribution across the client is usually different. Hence, there will be huge differences among local models trained on such distribution, making it difficult to obtain a superior global model through the traditional approach of weighted averaging model parameters. Thus, how to reduce the impact of data distribution differences across the clients on the global model is still an open question.

To address the above issues, in this paper, we propose a novel Class-Balanced Federated Learning (CBFL) method, which produces a class-balanced data set for each client through data generation technique. Specifically, CBFL designs a class distribution equalizer that consists of a class-balanced sampler and a data generator. First, the class-balanced sampler samples those classes that have insufficient data on the client with a higher probability. Then, the data generator generates corresponding dummy data according to the classes sampled by the class-balanced sampler. Finally, each client combines its original data and the generated data to produce a class-balanced data set for training, which contributes to obtaining a promising global model. Extensive experiments on four benchmark datasets verify the effectiveness of the proposed methods.

This work was partially supported by the Ministry of Science and Technology Foundation Project (No. 2020AAA0106900), the Joint Funds of the National Natural Science Foundation of China (No. U20B2052), the National Natural Science Foundation of China (No. 62072190), the Key-Area Research and Development Program of Guangdong Province (No. 2018B010107001), and the Program for Guangdong Introducing Innovative and Entrepreneurial Teams (No. 2017ZT07X183).