

李靖宇, 李新宇, 黄瑞钧, 等. 基于余弦相似度的分层聚合个性化联邦学习[J]. 智能计算机与应用, 2025, 15(9): 185-191.
DOI: 10.20169/j.issn.2095-2163.250928

基于余弦相似度的分层聚合个性化联邦学习

李靖宇¹, 李新宇², 黄瑞钧¹, 姜新维¹, 林宇焱¹, 任慧慧³

(1 湖州师范学院 信息工程学院, 浙江 湖州 313000; 2 大连交通大学 机械工程学院, 辽宁 大连 116028;

3 晋中学院, 山西 晋中 030606)

摘要: 个性化联邦学习作为一种新的联邦学习范式, 为每一个客户端生成独立个性化模型, 可解决客户端数据高度异质性的情况下全局模型在各客户端上性能差异增大, 平均准确度降低的问题。为提高个性化模型的性能, 考虑相似客户端协作可优化个性化模型, 并在模型聚合时利用分层差异化聚合的方式进行优化, 提出了一个基于余弦相似度的分层聚合个性化联邦学习方法。每次迭代中, 聚合服务器通过计算各客户端模型层级之间的余弦相似度, 区分模型层级的聚合权重, 并在模型聚合时保留本地个性化模型层, 从而为每个客户端生成个性化的聚合模型。实验结果显示, 提出的方法比其他个性化联邦学习方法在 FashionMNIST, CIFAR10 数据集上模型精度准确率提升 3.5% 以上, 并具有较高的鲁棒性。

关键词: 联邦学习; 个性化; 相似度; 分层聚合; 隐私保护

中图分类号: TP391

文献标志码: A

文章编号: 2095-2163(2025)09-0185-07

Hierarchical aggregation personalized federated learning based on cosine similarity

LI Jingyu¹, LI Xinyu², HUANG Ruijun¹, JIANG Xinwei¹, LIN Yujun¹, REN Huihui³

(1 School of Information Engineering, Huzhou University, Huzhou 313000, Zhejiang, China;

2 School of Mechanical Engineering, Dalian Jiaotong University, Dalian 116028, Liaoning, China;

3 Jinzhong University, Jinzhong 030606, Shanxi, China)

Abstract: As a new paradigm of federated learning, personalized federated learning generates an independent personalized model for each client, which can solve the problem of increasing the performance difference and decreasing the average accuracy of the global model on each client under the condition of high heterogeneity of client data. In order to improve the performance of the personalization model, considering the similarity client collaboration to optimize the personalization model, and using the hierarchical differential aggregation method to optimize the model aggregation, a hierarchical aggregation personalization federated learning method based on cosine similarity is proposed. In each iteration, the aggregation server generates a personalized aggregation model for each client by calculating the cosine similarity between the model levels of each client, distinguishing the aggregation weights of the model levels, and retaining the local personalized model layer when the model is aggregated. Experimental results show that the proposed method has more than 3.5% improvement in model accuracy accuracy and high robustness on the FashionMNIST, CIFAR10 dataset compared with other personalized federated learning methods.

Key words: federated learning; personalize; similarity; hierarchical aggregation; privacy protection

0 引言

随着信息化技术的发展, 移动终端设备在逐年递增, 客户端上数以万计的信息为机器学习模型的训练提供支持, 机器学习成为发展的必然趋势^[1]。虽然这些数据有助于训练机器学习模型, 但是这些

数据包含大量的隐私信息, 用户不愿共享这些信息^[2], 所以随着机器学习的广泛应用, 如何利用客户端的海量数据并且保护隐私成为热点课题^[3]。

谷歌在 2016 年提出联邦学习 (Federated Learning, FL), 在进行机器学习时把训练数据保留在用户设备上而无需用户上传这些可能带有敏感信息的私有数

作者简介: 李靖宇 (1997—), 男, 硕士研究生, 主要研究方向: 联邦学习, 算法机制设计。Email: 1019442745@qq.com; 李新宇 (1997—), 男, 硕士研究生, 主要研究方向: 遥感图像处理。

收稿日期: 2023-12-15

哈尔滨工业大学主办 ◆ 科技创新与应用

据,并通过联邦平均(FedAvg)算法^[4-5]对权重分配后进行模型聚合。现在联邦学习的应用领域众多,可应用于医疗、物联网、边缘计算、机器人、推荐系统和网络安全等领域^[6-7]。当数据集为非独立同分布(non-IID)时,训练出的模型会偏离向局部数据多的一方,而且所有客户端共享一个全局模型可能会减缓收敛速度,模型准确度下降^[8-10]。特别是,对单个客户端的数据集来说,全局模型的识别准确度较低。在客户端需求不一致时,一个模型无法适应所有的客户端^[11]。但仅依赖客户端数据集训练出每个客户端的本地模型而不进行联邦训练,又会出现浪费训练资源、训练出的模型不够完善等问题,针对这些问题的解决方案便是个性化联邦学习^[12]。

个性化联邦学习是一种解决上述问题的优化方案,通过各种个性化解决方法,如优化客户端数据分布、对模型优化和对架构优化来解决模型对于本地任务或者数据集缺乏个性化等不适应问题,从而提高模型在客户端的精度,并且在训练过程中节约资源,减少重复训练,加快收敛速度,提高通信速度,减少通信费用,可以帮助更多的用户和客户端共同受益从而提升模型的性能。

最近几年,机器学习研究人员提出了多种个性化联邦学习方法。基于数据的方法^[13-17],通过客户端之间的数据进行传输,以此来消除数据中心间数据分布的不均衡问题,从而将非独立同分布(non-IID)转化为独立同分布(IID)的设定,之后使用已有的学习方法进行解决;基于模型的方法^[18-23],通过提升全局联邦学习模型的个性化表现来实现个性化联邦学习;基于结构的方法^[24-25],通过为每个数据中心设置个性化的结构达到模型个性化的目的;基于相似的方法^[13,26-27],通过相似的客户端来训练多个模型,使模型之间可以存在偏差,达到个性化的目的。

基于模型的方法有3种方式:正则化、元学习和迁移学习。其中,正则化是在优化的全局模型的过程中,同时训练了每个数据中心的本地个性化模型参数,再通过正则化项进行关联^[18-19]。元学习是通过已经学习完成的模型,再使用此模型针对新的目标任务进行少量的训练,从而解决目标任务。Hanzely等学者^[20]提出混合全局模型和本地化模型的自适应元学习方法。迁移学习(微调)是通过调整传输数据的权重和用一个模型解决另一个相关问题的方法来实现个性化,通过少量的训练就可提升模型的精度。Pratt^[22]提出新的算法来实现基于判

别性的传输(DBT),通过缩放传输的权重幅度使网络学习速度明显增快。个性化深度学习^[23]的研究表明,降低更通用、更广泛的数据集的性能为代价的数据增强都可以提高个人数据的性能。但是基于模型的方法都是单一的全局模型。

基于相似性的方法有3种,分别是:多任务学习、模型差值和聚类。多任务学习是多个相似的任务同时进行学习,Huang等学者^[13]提出了一种基于注意力的机制FedAMP,可以在具有相似数据分布的FL客户端之间加强联系,构成客户端的成对协作。但这种方法对客户的数据质量差很敏感。模型差值是每个数据中心学习一个本地模型,通过准许本地模型和全局模型存在一定偏差产生模型的个性化。Deng等学者^[26]提出了APFL算法,为每个客户端引入了一个混合参数来以一种高效通信的方式找到全局和局部模型的最优组合。聚类是通过把相似的客户端聚成一类,为每一类客户端训练一个全局模型。这种方法使用单一的全局模型作为个性化的基础。Briggs等学者^[27]使用一个凝聚层次聚类公式,将聚类减少到一个步骤,以降低计算和通信负载。但此方法计算和通信成本高。

本论文对上述方法进行研究,在基于模型和模型相似性的方法基础上,提出了基于余弦相似度分层聚合个性化联邦学习(SLAPFL)。方法中使用所有客户端的模型来互相配合,以分层的方式进行模型聚合,提高了客户端之间的互相协作,并使模型精度也得以提升(在FashionMNIST提升了3.5%),此后经过多次实验验证结论。

1 相关知识

1.1 联邦学习

联邦机器学习是一个机器学习框架,能有效帮助多个客户端在满足用户隐私保护、数据安全,并且遵从法规的要求下,进行数据使用和机器学习建模。联邦学习作为分布式的机器学习范式,可以有效解决数据孤岛问题,让参与方在不共享数据的基础上实现联合建模,在技术上打破数据孤岛,实现人工智能。

在联邦学习场景中,模型训练使用的训练数据是分散在各个边缘设备上的,全局服务器和参与方通过网络连接来交换数据信息,通过迭代的全局聚合和更新来实现模型的训练。联邦学习框架的训练过程为以下步骤:

首先,协调方建立基本模型,并将模型的基本结构与参数告知各参与方。

其次,各参与方利用本地数据进行模型训练,并将结果返回给协调方。

最后,协调方汇总各参与方的模型,通过聚合构建更精准的全局模型,以整体提升模型性能和效果。在独立同分布数据情况下,使用联邦平均方法可以使聚合出的模型和最佳全局模型一致,得到平均状态下的最优模型。但在非独立同分布情况下,实际最佳模型更趋于平均,从而偏移最佳模型,而且收敛速度较慢,如图1所示。

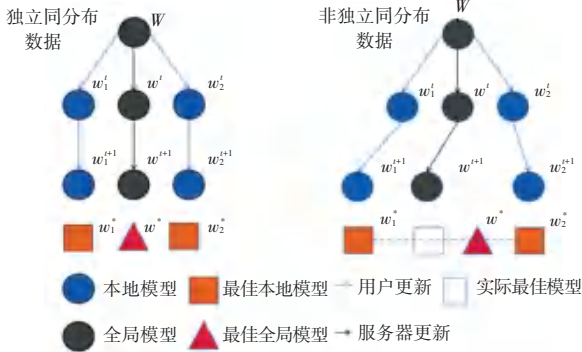


图1 联邦学习不同数据分布训练结果图

Fig. 1 Training results graph of different data distributions in federated learning

所以,需要在联邦学习的基础上进行个性化联邦学习,使客户端在非独立同分布数据情况下,解决收敛速度慢和缺少个性化模型的问题。

1.2 相似度

在人工智能领域,频繁出现距离、相似度概念,这些概念源于数学领域的度量、测度,现已在自然语言处理、计算机视觉等子领域有重要的应用。相似度是2个向量或事物相近程度的一种度量。其中,过滤算法和协同过滤算法使用特定的相似度度量来确定2个用户或商品的向量之间的相等程度,所以机器学习中,通过计算模型之间的距离评判模型的相似程度不够充分,应使用更有效的相似度方法来进行评估。经过实验验证,余弦相似度更能反映模型的相似程度,也可以计算2个向量之间的相似度,所以可以使用此方法来计算模型和模型层级之间的相似程度。

1.3 模型层级

基本的一般性网络模型有数据输入层、卷积层、激活层、池化层、全连接层等层级。其中,数据输入层对原始数据进行初步处理,实现数据灰度化和归一化,使卷积神经网络能有更好的效果。卷积层负责提取图像特征。激活层使非线性关系通过正价函数变成线性关系。池化层把输入的数据进行压缩,提取主要特征。全连接层通常是在卷积神经网络尾

部,与上一层之间的所有神经元都通过权重分配进行连接,是传统的神经网络(DNN)。

2 设计方法

2.1 问题定义

本文中设定为有 n 个客户端对本地的隐私数据 $\{D_1, D_2, \dots, D_n\}$ 进行训练,这些客户端的数据采用的分布方式是非独立同分布(non-IID)。具体来说,每个客户端都有与之数据分布相似的客户端,在不交换私有数据的情况下,每个客户端使用其本地数据 $\{D_1, D_2, \dots, D_n\}$,在中央服务器的帮助下互相协作,学习单个本地模型 M_1, M_2, \dots, M_n ,其中每个模型有 m 层。目标是最小化全局目标,获得最优的个性化客户端模型。客户端模型 M_1, M_2, \dots, M_n 通过如下公式计算后获取:

$$L = \arg \min \frac{1}{n} \sum_{i=0}^n L_i \quad (1)$$

其中, L 表示总体平均损失; L_i 表示第 i 个客户端的损失函数; n 表示客户端的总个数。此次训练目标为最小化每个客户端的损失函数,使平均总损失达到最小。

2.2 基于相似度的分层聚合

每个客户端在本地训练出模型后,上传到服务器,在服务器计算客户端模型相似度。在模型选择方面,使用的是多层次的一般性模型,把模型分为:基础层、本地个性化层。在聚合时,选择保留本地个性化层。通过大量实验,得出模型中本地个性化层学习到的是更具个性化客户端的特征,而其他层级学习到的是公共特征,保留与否对实验结果没有有效提升,为更好地提升模型精度,选择把本地个性化层直接保留到本地客户端进行使用。其中,聚合出的模型在服务器端,称之为云模型。

在评估选择计算相似度的方法上,经过实验研究,余弦相似度的综合相似准确度最高,所以选择计算相似度方法为余弦相似度,其定义公式为:

$$S_{ij} = \frac{M_i M_j}{|M_i| \cdot |M_j|} = \frac{x_i x_j + y_i y_j}{\sqrt{x_i^2 + y_i^2} \sqrt{x_j^2 + y_j^2}} \quad (2)$$

其中, x, y 分别表示客户端的横纵坐标; i, j 表示第几个客户端。本文的研究方法选择的是按层级计算模型的相似度,保留本地个性化层后,计算客户端层级之间的余弦相似度,根据相似度比例按层级进行模型权重分配,聚合出每个客户端的个性化云模型层,接下来进行模型层级相加,构成完整客户端云模型。分层聚合公式具体如下:

$$M_{ik} = \frac{s_{ij}}{\sum_{j=1}^n s_{ij}} M_{1k} + \frac{s_{ij}}{\sum_{j=1}^n s_{ij}} M_{2k} + \cdots + \frac{s_{ij}}{\sum_{j=1}^n s_{ij}} M_{nk}$$
$$(i, j \in \{1, 2, 3, \cdots, n\}) \quad (3)$$

其中, s_{ij} 表示客户端 i 和 j 的相似度, M_{ik} 表示第

i 个客户端的第 k 层模型。客户端整个云模型通过每个模型层相加构成, 数学定义公式为:

$$M_i = \sum_{k=1}^n M_{ik} \quad k \in \{1, 2, 3, \cdots, m\} \quad (4)$$

具体流程如图 2 所示。

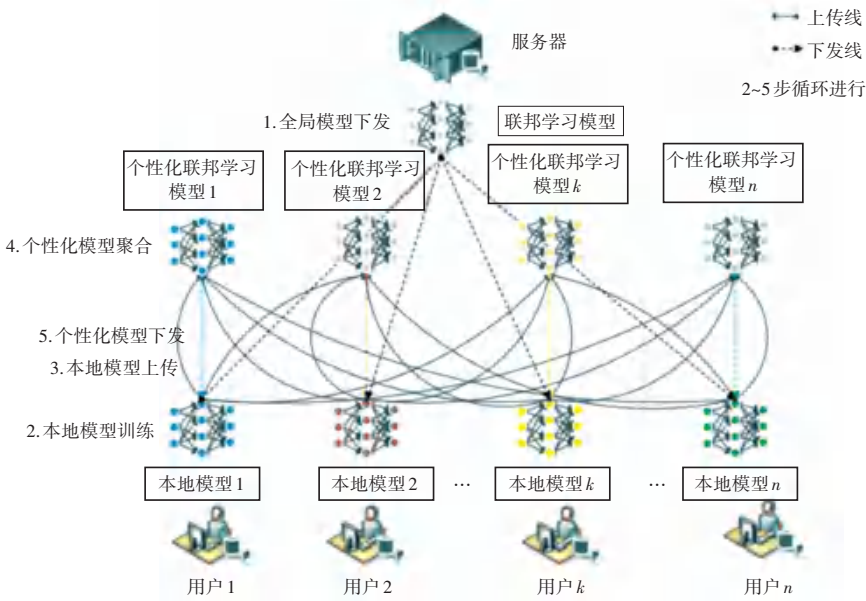


图 2 基于模型相似度的分层聚合流程

Fig. 2 Flowchart of hierarchical aggregation based on model similarity

由图 2 可知, 流程步骤为:

(1) 中央服务器构建初始网络模型及初始化模型对应的参数, 下发给所有参与训练的客户端。

(2) 每个客户端使用本地数据来训练本地模型。

(3) 客户端上传本地整体模型至中央服务器端。

(4) 服务器把模型基础层进行模型相似度分层聚合, 根据模型层级相似度比例分配权重进行聚合, 保留模型本地个性化层不进行聚合, 生成每个客户端在服务器端的个性化云模型层。通过层级相加, 聚合成每个客户端的个性化云模型。

(5) 服务器把聚合好的个性化云模型下发给每个客户端。

其中, 循环步骤 (3) ~ (5) 进行训练, 训练达到最小化平均客户端的损失后, 比较个性化模型在本地客户端测试集上的精度, 检验提升的精度。

2.3 算法

算法整体流程如下: 在服务器端, 服务器下发模型给每个客户端训练时, 需要判断是否是第一次开始训练, 如果是第一次, 则下发服务器的初始模型给每个客户端, 否则接收每个客户端传来的模型, 服务器把客户端上传的模型按式 (2) 计算出每 2 个客户

端之间的层级相似度, 按照式 (3) 聚合出每个客户端的个性化云模型层。其中, 保留本地个性化层不参与聚合, 通过式 (4) 相加模型层级得出每个客户端的云模型, 下发模型给每个客户端。在客户端, 客户端接收服务器传来的模型, 使用每个客户端的本地数据集进行训练, 训练完成后上传到服务器。以上步骤交叉循环进行, 直到到达指定轮数。算法描述具体如下。

算法 1 相似度分层聚合算法

输入 n 个客户端, 每个客户端都有一组私人的训练数据 $\{D_1, D_2, \cdots, D_n\}$; 一个服务器, 服务器拥有全局模型 M , 通讯轮次 T

输出 个性化模型 M_1, M_2, \cdots, M_n

1. 服务器执行:

2. for $t = 1$ to T do

3. if $t = 1$ then

4. 下发模型 M 给每个客户端

5. else

6. 接收每个客户端上传的个性化模型 M_1, M_2, \cdots, M_n

7. 根据上传的客户端模型按式 (2) 计算每

个客户端层级之间的余弦相似度 S_{ij}

8. CNN 模型保留本地个性化层后,根据式(3)按层级相似度比例进行分层聚合,再根据式(4)层级相加,得出每个客户端云模型 M'_1, M'_2, \dots, M'_n

9. 下发云模型 M'_1, M'_2, \dots, M'_n 给对应的每个客户端

10. end if
11. end for
12. 客户端执行:
13. for $t = 1$ to T do
14. if $t = 1$ then
15. 每个客户端接收服务器下发的模型 M
16. else

17. 每个客户端接收服务器下发的对应模型 M'_1, M'_2, \dots, M'_n
18. end if

19. 每个客户端接收的模型上用本地数据集 $\{D_1, D_2, \dots, D_n\}$ 训练本地模型 M_1, M_2, \dots, M_n

20. 客户端上传本地模型的个性化模型 M_1, M_2, \dots, M_n 给服务器

21. end for

算法 1 中,步骤 7 和步骤 8 为核心算法。步骤 7 中,每个客户端与所有的客户端(包括自己)通过式(2)计算余弦相似度。步骤 8 中,使用式(3)根据层级相似度完成层级聚合,每个客户端根据所有相似度和的百分比进行权重分配,模型中除本地个性化层外的基础层,每一层都根据此方法聚合,再通过式(4)相加模型层级得出每个客户端的云模型。

3 实验

在本小节,评估相似度分层聚合算法,对比了联邦平均算法、个性化联邦学习算法和相似度分层聚合算法,在标准数据集 FashionMNIST, CIFAR10 下进行训练实验,均在表 1 的硬件配置下进行实验验证。

表 1 实验硬件配置

Table 1 Experimental hardware configuration	
配置项	配置信息
CPU	AMD Ryzen7 4800H with Radeon Graphics@2.90 GHz
GPU	NVIDIA GeForce RTX 2060
RAM(CPU)	16.0 GB
RAM(GPU)	6.0 GB

3.1 数据集划分

文中使用了 2 个标准数据集 FashionMNIST, CIFAR10。客户端数据分布的划分方法是按照最接

近真实使用情况进行划分,使得每个客户端存在与之数据划分相似的客户端。所有方法均采用一般性模型卷积神经网络(CNN),含有输入层、卷积层、激活层、池化层、全连接层等层级。其中,全连接层学习到的是更偏向于本地的特征,并对应普通的神经网络做分类,所以把这层保留到本地进行训练,研究中将该层称为本地化个性化层。剩余层会提取图像基本、中等和高层特征,需要与其他模型进行聚合来学习公共特征,这些层被称为基础层。使用学习率为 0.01 的随机梯度下降优化器,设置联邦轮数为 1 000 或 1 500 轮,客户端本地训练轮数为 1, Batch Size 为 100。其他相关参数均与文献[13]实验中的设置相同。

因为客户端在具体使用情境中,不同的区域环境下所含有的数据是不同的,如一类客户端是学生,这类客户端所在的区域为学校,所含有的数据为文字资料;一类客户端为厨师,这类客户端所在的区域为餐厅,所含有的数据为美食;一类客户端为建筑师,这类客户端所在的区域为工地,所含有的数据为房子,但是其中每一类都会有一定数量的客户端,这些客户端的数据是相似的,训练的时候可以互相促进来训练个性化模型。

以 FashionMNIST 为例,设定为 20 个客户端,把客户端划分为 3 组,分别是:0~6 为第一组,7~13 为第二组,14~19 为第三组。第一组中把 FashionMNIST 的 10 类数据中的 2 类设置为主类、占 80%,剩下的 8 类数据设置为非主类数据、占 20%。第二组把 10 类数据中的 3 类设置为主类、占 80%,剩下的 7 类数据设置为非主类数据、占 20%。第三组把 10 类数据中的 5 类设置为主类、占 80%,剩下的 5 类数据设置为非主类数据、占 20%。划分客户端完毕之后,训练集和数据集划分一致,用客户端的总体 80%为训练集,其中主类占客户端总训练集的 80%,非主类占总训练集的 20%;20%为测试集,其中主类占客户端总训练集的 80%,非主类占总训练集的 20%。所以每个客户端都有与其相似的客户端进行协作,最后来共同聚合出客户端的个性化本地模型。

3.2 实验结果

图 3 是 CIFAR10 数据集下 1 500 轮训练后的精度比较结果,图 4 是 FashionMNIST 数据集下 1 000 轮训练后的精度比较结果,蓝色的点状线是联邦平均方法(Fedavg^[4])、橙色的实线是自适应局部聚合个性化方法(FedALA^[21])、绿色的虚线是本文的相似度分层

聚合个性化联邦学习算法 (SLAPFL)。所有方法评估的是最优的平均测试精度 (BMTA) 百分比。平均测试精度是各个客户端测试精度的平均值,计算精度时,按照的是个性化模型在本地数据集下来测试出个性化模型的精度。对比的指标为客户端平均模型精度、轮数平均值和轮数方差,详见图 3、图 4。

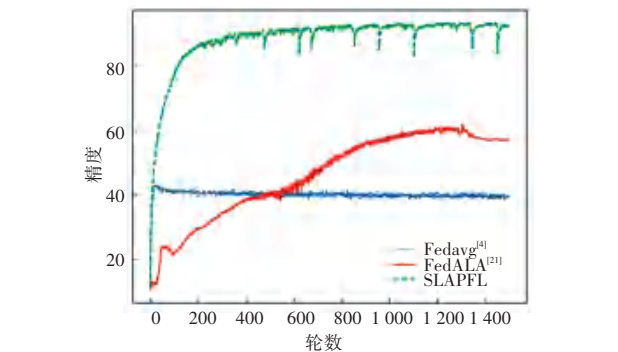


图 3 CIFAR10 数据集 1 500 轮训练的精度比较图

Fig. 3 Comparison chart of the accuracy of 1 500 rounds of training in the CIFAR10 dataset

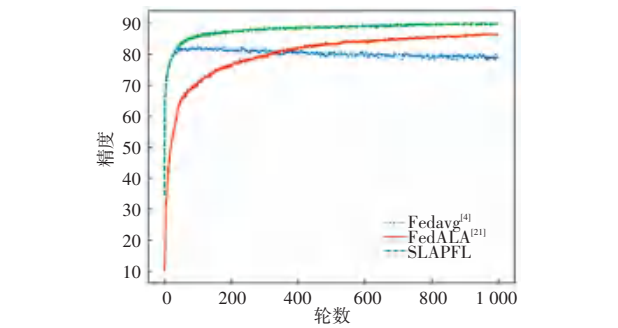


图 4 FashionMNIST 数据集 1 000 轮训练的精度比较图

Fig. 4 Comparison of the accuracy of 1 000 rounds of training in the FashionMNIST dataset

实验结果表明,图 3 中,在 CIFAR10 数据集下,与自适应局部聚合个性化方法相比、本文方法的模型精度显著提高,高达 30% 以上,并且模型训练速度有明显提升。图 4 中,在 FashionMNIST 数据集下,本文的方法比自适应局部聚合个性化方法模型精度提升 3.5% 以上,此方法能更好地使相似度高的模型进行聚合提高模型精度。综上所述,本文的方法使用相似度高的模型进行权重分配再聚合是可行的。

表 2 给出了 3 种方法对比的模型损失收敛后精度结果比较。可以看出本方法在非独立同分布数据分布情况下,在 FashionMNIST 中比其他方法高出 3.5% 以上,在 CIFAR10 中比其他方法要高 30% 以上。从方差来进行对比,2 个个性化方法比较结果显示,本文的方法比 FedALA 方法离散程度更低,训

练精度提升速度更快。

表 2 不同数据集下不同方法的准确率比较

Table 2 Comparison of the accuracy of different methods under the different dataset

方法	数据集					
	FashionMNIST			CIFAR10		
	模型精度	平均值	方差	模型精度	平均值	方差
Fedavg ^[4]	79.34	79.95	3.58	39.79	40.07	1.87
FedALA ^[21]	86.46	79.89	81.34	57.04	46.08	170.56
SLAPFL	89.91	87.83	11.06	92.67	88.45	73.32

在联邦平均方法中,每个客户端的权重比例都一致,而且聚合出的模型无法更好地适应所有客户端。自适应局部聚合方法中,是通过全局模型和本地模型进行聚合,而不是与其他客户端进行聚合得出模型。相似度分层聚合方法,使相似客户端之间进行协作,自适应地在每一轮迭代中按相似度来调整模型层级之间的权重,再进行聚合,而不是单纯地利用数据的数量进行聚合,同时还保留了本地个性化层不参与聚合,此方法可以快速聚合出本地个性化模型。

经过客户端非独立同分布数据分布的修改,进行 CIFAR10 数据集下的鲁棒性验证。结果证明,在产生一定偏差的数据下,此方法还具有一定的鲁棒性,实验结果如图 5 所示。

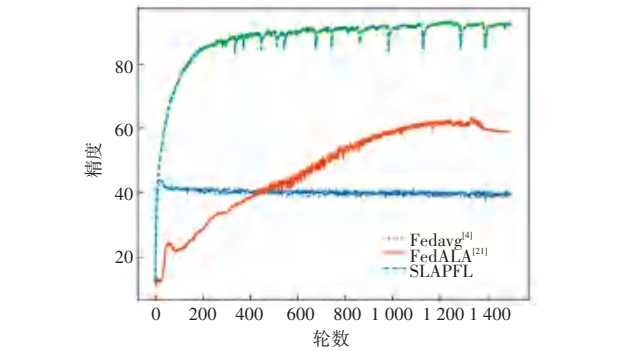


图 5 客户端数据分布偏差下 CIFAR10 数据集 1 500 轮训练的精度比较图

Fig. 5 Comparison of the accuracy of 1 500 rounds of training in the CIFAR10 dataset under the client-side data distribution bias

4 结束语

在本文中,提出了相似度分层聚合算法,使服务器在不侵犯客户端数据隐私的情况下促进客户端之间的互相协作。客户端模型保留本地个性化层后,按照相似度进行分层聚合,每一轮自适应地按照模型层级相似度调整每层模型权重,最后聚合出客户

端本地个性化模型。本文的方法解决了面对客户端数据高度异质性的情况下全局模型在各客户端上性能差异增大,平均准确度降低的问题,提高了个性化模型精度,并在此后进行广泛的实验证明了所提出的方法的优越的性能,得出的模型具有较高的普适性,在 FashionMNIST 和 CIFAR10 数据集下模型精度高于其他方法 3.5% 以上,还进一步证明了该方法的鲁棒性。未来,本文将考虑更多影响模型精度的因素,使本文的方法适应更多场景。

参考文献

[1] POUSHTER J. Smartphone ownership and internet usage continues to climb in emerging economies[J]. Pew Research Center, 2016, 22(1): 1-44.

[2] 杨强. AI 与数据隐私保护:“联邦学习”的破解之道[J]. 信息安全研究, 2019, 5(11): 961-965.

[3] 肖雄,唐卓,肖斌,等. 联邦学习的隐私保护与安全防护研究综述[J]. 计算机学报, 2023, 46(5): 1019-1044.

[4] MCMAHAN B, MOORE E, RAMAGE D, et al. Communication-efficient learning of deep networks from decentralized data[J]. arXiv preprint arXiv, 1602.05629, 2023.

[5] DINH C T, TRAN N H, NGUYEN M N H, et al. Federated learning over wireless networks: Convergence analysis and resource allocation[J]. IEEE/ACM Transactions on Networking, 2020, 29(1): 398-409.

[6] 王健宗,孔令炜,黄章成,等. 联邦学习算法综述[J]. 大数据, 2020, 6(6): 64-82.

[7] 周传鑫,孙奕,汪德刚,等. 联邦学习研究综述[J]. 网络与信息安全学报, 2021, 7(5): 77-92.

[8] ZHANG Jie, GUO Song, QU Zhihao, et al. Adaptive federated learning on non-IID data with resource constraint[J]. IEEE Transactions on Computers, 2021, 71(7): 1655-1667.

[9] HSIEH K, PHANISHAYEE A, MUTLU O, et al. The non-IID data quagmire of decentralized machine learning[J]. arXiv preprint arXiv, 1910.00189, 2019.

[10] ZHAO Yue, LI Meng, LAI Liangzhen, et al. Federated learning with non-IID data[J]. arXiv preprint arXiv, 1806.00582, 2018.

[11] TAN Ziyi, YU Han, CUI Lizhen, et al. Towards personalized federated learning[J]. IEEE Transactions on Neural Networks and Learning Systems, 2023, 34(12): 9587-9603.

[12] KULKARNI V, KULKARNI M, PANT A. Survey of personalization techniques for federated learning[C]//Proceedings of 2020 Fourth World Conference on Smart Trends in Systems, Security and Sustainability (WorldS4). Piscataway, NJ: IEEE, 2020: 794-797.

[13] HUANG Yutao, CHU Lingyang, ZHOU Zirui, et al.

Personalized cross-silo federated learning on non-IID data[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2021, 35(9): 7865-7873.

[14] ZHAO Yue, LI Meng, LAI Liangzhen, et al. Federated learning with non-IID data[J]. arXiv preprint arXiv, 1806.00582, 2018.

[15] JEONG E, OH S, KIM H, et al. Communication-efficient on-device machine learning: Federated distillation and augmentation under non-iid private data[J]. arXiv preprint arXiv, 1811.11479, 2018.

[16] CHAI Zheng, ALI A, ZAWAD S, et al. TiFL: A tier-based federated learning system [C]//Proceedings of the 29th International Symposium on High-performance Parallel and Distributed Computing. New York: ACM, 2020: 125-136.

[17] LI Li, DUAN Moming, LIU Duo, et al. FedSAE: A novel self-adaptive federated learning framework in heterogeneous systems [C]//Proceedings of 2021 International Joint Conference on Neural Networks (IJCNN). Piscataway, NJ: IEEE, 2021: 1-10.

[18] LI Tian, SAHU A K, ZAHEER M, et al. Federated optimization in heterogeneous networks[J]. Proceedings of Machine Learning and Systems, 2020, 2: 429-450.

[19] LI Tian, HU Shengyuan, BEIRAMI A, et al. Ditto: Fair and robust federated learning through personalization [J]. arXiv preprint arXiv, 2012.04221, 2020.

[20] HANZELY F, RICHTÁRIK P. Federated learning of a mixture of global and local models[J]. arXiv preprint arXiv, 2002.05516, 2020.

[21] ZHANG Jianqing, HUA Yang, WANG Hao, et al. FedALA: Adaptive local aggregation for personalized federated learning[J]. Proceedings of the AAAI Conference on Artificial Intelligence. 2023, 37(9): 11237-11244.

[22] PRATT L Y. Discriminability-based transfer between neural networks[C]//Proceedings of the 6th International Conference on Neural Information Processing Systems (NIPS'92). New York: ACM, 1992: 204-211.

[23] SCHNEIDER J, VLACHOS M. Mass personalization of deep learning[J]. arXiv preprint arXiv, 1909.02803, 2019.

[24] ARIVAZHAGAN M G, AGGARWAL V, SINGH A K, et al. Federated learning with personalization layers[J]. arXiv preprint arXiv, 1912.00818, 2019.

[25] HINTON G, VINYALS O, DEAN J. Distilling the knowledge in a neural network[J]. arXiv preprint arXiv, 1503.02531, 2015.

[26] DENG Y, KAMANI M M, MAHDAVI M. Adaptive personalized federated learning [J]. arXiv preprint arXiv, 2003.13461, 2020.

[27] BRIGGS C, FAN Zhong, ANDRAS P. Federated learning with hierarchical clustering of local updates to improve training on non-IID data[C]//Proceedings of 2020 International Joint Conference on Neural Networks (IJCNN). Piscataway, NJ: IEEE, 2020: 1-9.