融合通道个性标准化的本地自适应联邦学习研究

赵 宇 陈思光*
(南京邮电大学物联网学院 南京 210003)

摘 要：为了缓解联邦学习(FL)中客户端之间由于完全重叠特征偏移所带来的数据异构问题影响，该文提出一种融合通道个性标准化的本地自适应联邦学习算法。具体地，构建了一个面向数据特征偏移的联邦学习模型，在训练开始之前先对客户端中的图像数据集进行一系列随机增强操作。其次，客户端分别按颜色通道单独计算数据集的均值和标准差，实现通道个性标准化。进一步地，设计本地自适应更新联邦学习算法，即自适应地聚合全局模型和本地模型以进行本地初始化，该聚合方法的独特之处在于既保留了客户端模型的个性化特征，同时又能从全局模型中捕获必要信息，以提升模型的泛化性能。最后，实验结果表明，该文所提算法与现有相关算法相比，收敛速度更快，准确率提高了3%～19%。

关键词：边缘计算；联邦学习；标准化；模型聚合
中图分类号：TN919; TP393 **文献标识码**：A **文章编号**：1009-5896(2022)YU-0001-10
DOI: 10.11999/JEIT231165

Local Adaptive Federated Learning with Channel Personalized Normalization

ZHAO Yu CHEN Siguang
(*School of Internet of Things*, *Nanjing University of Posts and Telecommunications*, *Nanjing* 210003, *China*)

Abstract: To relieve the impact of data heterogeneity problems caused by full overlapping attribute skew between clients in Federated Learning (FL), a local adaptive FL algorithm that incorporates channel personalized normalization is proposed in this paper. Specifically, an FL model oriented to data attribute skew is constructed, and a series of random enhancement operations are performed on the images data set in the client before training begins. Next, the client calculates the mean and standard deviation of the data set separately by color channel to achieve channel personalized normalization. Furthermore, a local adaptive update FL algorithm is designed, that is, the global model and the local model are adaptively aggregated for local initialization. The uniqueness of this aggregation method is that it not only retains the personalized characteristics of the client model, but also can capture necessary information in the global model to improve the generalization performance of the model. Finally, the experimental results demonstrate that the proposed algorithm obtains competitive convergence speed compared with existing representative works and the accuracy is 3%～19% higher.

Key words: Edge computing; Federated Learning (FL); Normalization; Model aggregation

1 引言

在这个数字时代，随着智能手机和物联网感知设备等在现实场景中的持续应用和部署，来自分布式终端的数据快速增长。各大组织正在使用大数据和人工智能来优化其流程和性能，大规模数据的积累已经成为科技和商业发展的推动力，促进着以数据驱动为主的深度学习技术迅猛发展。大数据为人工智能提供数据支撑，人工智能为大数据提供更高效的处理和分析工具，二者的有效结合已经在许多领域产生了巨大的影响。与此同时，数据的隐私和安全性问题也日益突出，尤其是在医疗保健行业，因为医疗数据高度敏感，并且通常在不同的医疗保健机构中收集和保留[1]，这些数据中的大多数本质上都是高度敏感且独立的，它们以孤岛的形式存在，这导致大量数据的价值难以被传统人工智能所

收稿日期：2023-10-26；改回日期：2024-01-24；网络出版：2024-03-04
*通信作者： 陈思光 sgchen@njupt.edu.cn
基金项目：国家自然科学基金(61971235)，江苏省“333高层次人才培养工程”和南邮“1311”人才计划
Foundation Items: The National Natural Science Foundation of China (61971235), The 333 High-level Talents Training Project of Jiangsu Province, and the 1311 Talents Plan of NJUPT

充分利用。同时，传统的中心化数据往往集中存储和处理用户的敏感信息，这种集中式方法常常引发隐私泄露和数据滥用的危机，并且数据集中在一个地方也存在着风险，一旦数据中心受到攻击或发生故障，将导致严重的数据丢失和业务中断。此外，当传统人工智能系统涉及到收集大规模数据时，数据的传输和同步会导致大量的通信开销。这些问题都给人工智能的应用带来了巨大的挑战。

为了解决上述问题，Google于2016年提出一种分布式的机器学习框架称为联邦学习(Federated Learning, FL)[2]。联邦学习可以利用分布式用户数据，同时通过迭代下载模型、在客户端本地训练模型、上传模型和在服务器上聚合模型来保护隐私。联邦学习不再将所有用户数据集中到1个中心服务器上进行训练，而是将模型分发到用户设备上训练。每个用户在本地设备上训练模型，只共享模型参数的更新，而不是原始数据。虽然联邦学习提供了一种很有价值的隐私保护方法，但当联邦学习应用于现实世界时，与集中式学习相比，会出现许多问题[3]。这些问题包括在服务器和本地设备之间传输参数所需的通信成本高、本地设备所需算力和能耗大、本地设备异构性所导致的模型性能低，以及不能有效提供不同客户或群体的个性化方案。目前有大量研究以解决上述问题，主要包括降低通信成本的研究方案例如文献[4–7]，考虑算力和能量限制的联邦学习[8,9]，面向数据异构的联邦学习[10–12]，以及聚类多任务联邦学习[13,14]。上述方案虽然尝试了减少客户端数据差异性影响和增加基于客户相似性的个性化方案，但仍然面临着两个挑战：在高度异构的数据上收敛性差，以及忽视对客户端间相似特征的个性化考量[15]。在存在异构本地数据分布的情况下，这些问题会恶化全局模型在单个客户端上的性能，甚至可能抑制受影响的客户端加入联邦过程。因此，相关研究者提出了一系列解决方案，用于解决数据异构和忽视个性化考量所带来的问题。

针对数据异构性问题，McMahan等人[16]首次引入了基于数据并行的联邦概念，并提出了联邦平均(Federated Averaging, FedAvg)算法。作为众所周知的传统联邦学习方法，FedAvg允许多个设备协同训练机器学习模型，同时将用户数据存储在本地。FedAvg消除了将用户的敏感数据上传到集中式服务器的需要，并使边缘设备能够在自己的本地数据集中本地训练共享模型[17]。通过聚合本地模型的更新梯度，FedAvg满足了隐私保护和数据安全的基本要求。然而，FedAvg方法下得到的局部收敛点可能与全局模型的目标(即通过在中央服务器进行聚合来学习的模型)不太一致[18]。因此，客户端模型经常偏离理想的全局优化点，并过度拟合到其局部目标。当发生客户端漂移现象时，中央聚合模型的性能会受到严重阻碍[19]。为了解决这个问题，联邦近端项算法(Federated Proximal, FedProx)[20]通过近端项提高了联邦学习过程的稳定性。智能设备选择联邦学习控制框架(Favor)[21]在每次迭代时基于深度Q学习选择一个子集的客户，从而抵消非独立同分布数据带来的偏差。联邦匹配平均算法(Federated Matched Averaging, FedMA)[22]通过将全局模型分层，使用匹配平均的方法来适应数据异构性。这些改进方法的设计是为了减轻过拟合和客户端漂移现象，但这些问题仍然存在。如何更好地维护全局模型的性能，同时减少客户端模型的过拟合，仍然是一个挑战。

针对忽视个性化考量的问题，当前的相关研究试图通过个性化聚合生成特定客户端的模型，以便进行个性化设置或提高本地模型的适应性能。为实现这一目标，现有研究方案存在不同的技术和方法。一种方法是通过注意力诱导机制和个性化聚合为单个客户端生成聚合模型，例如联邦注意力信息传递机制(Federated Attentive Message Passing, FedAMP)[23]。另一种方法是使用基于规则的移动平均和预定义的权重(超参数)来聚合全局模型和本地模型，例如继承私有模型的联邦个性化(Federated Personalization with inHerited Private models, FedPHP)[24]。FedPHP引入了“私有模型”和“继承”的概念。在训练过程中，客户端的私有模型可以与全局模型进行信息共享，这使得全局模型可以在不同客户端之间进行个性化调整。联邦一阶模型优化算法(Federated First order model optimization, FedFomo)[25]提出了一种新的联邦学习框架，该框架可以有效地计算每个客户端与其他可用客户端模型的个性化加权组合，其性能优于现有模型。类似地，自适应个性化跨领域联邦学习(Adaptive Personalized cross-siLo fEderated learning, APPLE)[26]可以自适应地学习每个客户端从其他客户端模型中受益的程度，并灵活地控制APPLE训练的焦点，使其在全局和本地目标之间达到个性化平衡。这些方法虽然在提高模型性能和稳定性方面取得了一些进展，但仍缺乏对客户端间相似特征的个性化考量。

基于上述挑战，本文提出一种融合通道个性标准化的本地自适应联邦学习算法，主要贡献总结如下：

(1)设计了通道个性标准化方法，即对不同数据集的每个特征通道(通常是三原色光色彩模式

(RGB color mode, RGB)图像中的颜色通道)应用独立的均值和标准差，而非传统标准化方法对整个批次应用统一固定值的均值和标准差。通道个性标准化可以更好地保留通道特定的信息，适应不同特征通道的数据分布，加快模型收敛速度。

(2)提出本地自适应联邦学习方法，即通过自适应地聚合全局模型和本地模型以进行本地初始化，利用这种聚合方式初始化的本地模型可实现客户端以分层聚合的方式参与模型更新，达到保留其个性化特征的目的，同时使不同客户端从全局模型中获取有用的信息以提高模型的泛化性能。本方法有助于解决联邦学习中的数据异构性问题，使模型更具适应性并具有更高的准确性。

(3)通过大量的仿真与分析，验证了本文所提算法能较好地解决数据异构带来的负面影响，并与其它经典的联邦学习算法对比，本文算法在准确率和收敛速度上有较大的性能优势。

本文其余部分组织如下：第2节介绍了系统模型；第3节为融合通道个性标准化的本地自适应联邦学习算法的相关定义及详细方案；第4节为仿真与性能评估；最后，第5节总结全文。

2　系统模型

本节构建了一个面向数据特征偏移的联邦学习模型，如图1所示，该模型由客户层和边缘层组成，每层的功能定义如下：

(1)边缘层。此层由一个边缘服务器组成，服务器主要包括两个功能。(a)模型初始化：在训练开始前，边缘服务器将随机初始化全局模型并下发给所有客户端。(b)模型聚合：边缘服务器收集来自各方客户端的本地模型，其模型参数被执行联邦平均聚合，从而生成新的全局模型供下一轮训练使用。

(2)客户层。此层由N个客户端设备组成，每个客户端i分别具有数据集$\{D_1, D_2, \cdots, D_N\}, \forall i \in [N]$，假设这些数据集来自同一领域并具有相同的标签和标签分布，但客户端之间的数据具有异构的外观，符合完全重叠特征偏移。由于各方客户端在训练中其原始数据始终保留在本地而不共享，降低了信息泄露的风险，可满足客户端对于本地数据的隐私和安全需求。如图1所示是本地自适应更新联邦学习系统中各客户端的本地学习过程，具体来说包括3个功能。(a)数据增强：在训练开始之前客户端会对数据集中的图象进行一系列随机增强操作，扩充原有数据集，增强数据多样性。(b)通道个性标准化：客户端分别按颜色通道单独计算数据集的均值和标准差，实现个性标准化，用于初始化模型训练，加快模型收敛。(c)本地自适应更新：客户端使用边缘服务器下发的全局模型与上一轮的本地模型聚合以自适应更新，并基于本地数据以及均值和标准差训练更新后的初始化模型，最后将训练好的本地模型上传至边缘服务器。

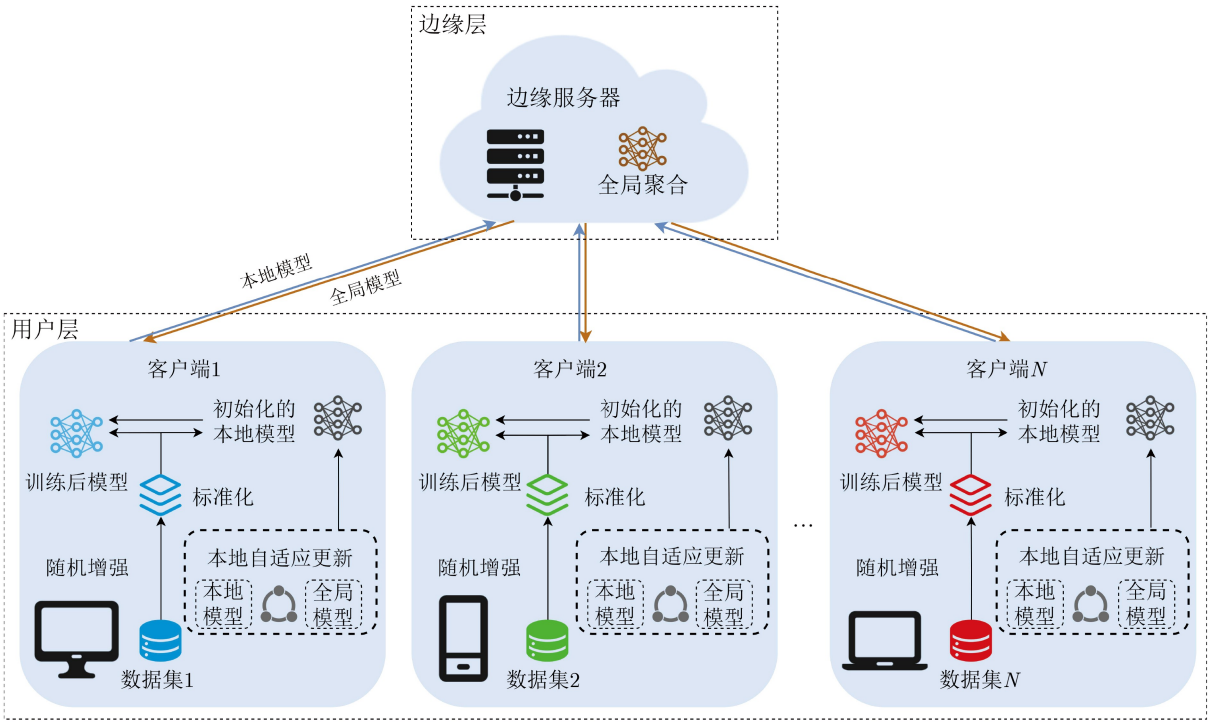

图 1 自适应更新联邦学习模型

3 融合通道个性标准化的本地自适应联邦学习

本文提出一种融合通道个性标准化的本地自适应联邦学习算法，即为解决客户端之间由于完全重叠特征偏移而带来的数据异构问题。本算法通过自适应地聚合全局模型和本地模型以进行本地初始化，利用这种聚合方式初始化的本地模型，一方面保留了客户端模型的个性化特征，另一方面从全局模型中捕获到的所需信息有助于提升模型的泛化性能。同时，采用数据增强和通道个性标准化的方法对各客户端数据集进行数据预处理，以达到避免过拟合和加快模型收敛速度的目的。

3.1 方案框架

本文提出的融合通道个性标准化的本地自适应联邦学习框架如图2所示。在边缘服务器的协调下，本方案的目标是使用每个客户端自己的$D_1, D_2, \cdots, D_N$协作学习单个初始化本地模型$\hat{\theta}_1, \hat{\theta}_2, \cdots, \hat{\theta}_N$，而不交换私有数据。其优化目标可表示为

$$\hat{\boldsymbol{\theta}}_i = \arg\min \sum_{i=1}^{N} r_i L_i \tag{1}$$

$$r_i = \frac{|D_i|}{\sum_{j=1}^{N} |D_j|} \tag{2}$$

其中，N为客户端总数，$|D_i|$为客户端i上的本地样本数据量，r_i为客户端i的本地样本数据量权重。$L_i = L\left(\hat{\boldsymbol{\theta}}_i, D_i; \boldsymbol{\theta}\right), \forall i \in [N]$，且$L(\cdot)$是表示客户端的损失函数，$\theta$是全局模型，由边缘服务器聚合生成。

如图2所示是第t次迭代中客户端的本地学习过程，客户端i首先对数据进行数据增强，然后求出各客户端数据集的平均值和标准差以实现通道个性标准化。数据预处理后，客户端i在本地进行自适应更新操作，结合全局模型$\boldsymbol{\theta}^{t-1}$计算出更新项，再利用上一轮的本地模型$\boldsymbol{\theta}_i^{t-1}$、更新项和旧的聚合权重学习出新的自适应聚合权重$\boldsymbol{W}_i^p$。在训练$\boldsymbol{W}_i^p$时，其他可训练参数均被冻结(图中透明遮盖部分)，包括全局模型$\boldsymbol{\theta}^{t-1}$和本地模型$\boldsymbol{\theta}_i^{t-1}$。在本地初始化完成之后，客户端$i$才会解冻模型，并正常进行本地模型训练。通过自适应聚合权重初始化本地模型$\hat{\boldsymbol{\theta}}_i^t$。最后利用本地数据集$D_i$以及均值和标准差对初始化的本地模型进行训练更新，训练完成后边缘服务器收集来自各方客户端的本地模型$\boldsymbol{\theta}_i^t$，其模型参数被执行联邦平均并聚合，从而生成新的全局模型θ^t供下一轮训练使用。详细的训练过程设计将在3.2节中进行描述。

3.2 详细方案

本节提出的融合通道个性标准化的本地自适应联邦学习算法主要包括4个部分内容：数据增强，通道个性标准化，自适应本地更新和全局模型聚合。

(1)数据增强。较早时期的数据增强方法往往包含30多个参数，为减少数据增强的参数空间且维持数据图像的多样性，本文采用随机增强(RandAugment)方法而不是自动增强(AutoAugment)方法，以省去搜索策略的过程。去除搜索阶段的原因是，单独的搜索阶段显著地使训练复杂化并且计算成本高昂。考虑从$K=14$个可用变换中选择应用哪些变化，具体包括Identity, AutoContrast, Equalize, Rotate, Solarize, Color, Posterize, Contrast, Brightness, Sharpness, Shear-X, Shear-Y, Translate-X, Translate-Y。应用每个变化的概率均匀一致为$1/K$，对给定训练图像做E次变换，RandAugment因此可以表示K^{E}个潜在策略。

本方案要考虑的参数是每个增强失真(augmentation distortion)的大小，使用相同的线性标

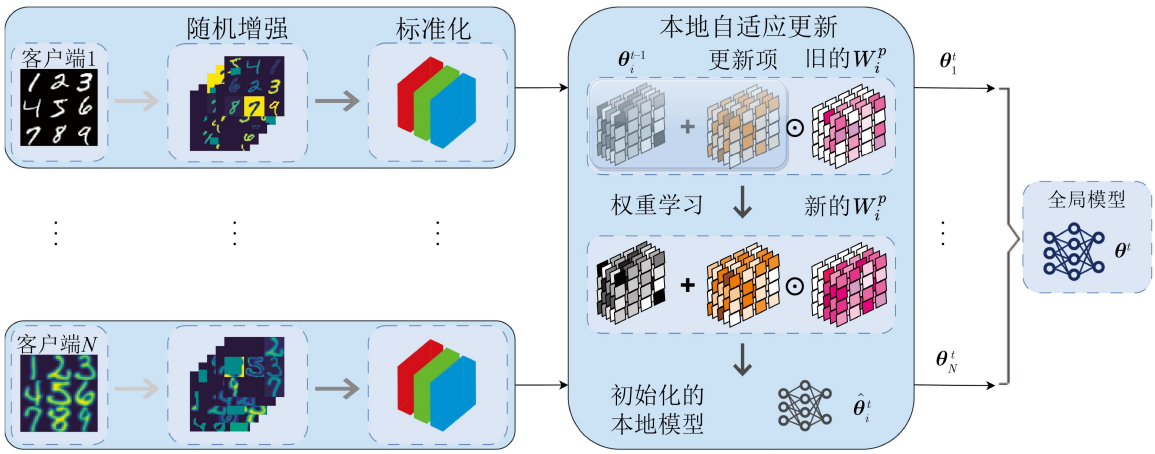

图 2 融合通道个性标准化的本地自适应联邦学习框架

度来指示每个变换的强度。即每个变换都位于0～10的整数范围内，其中10表示给定变换的最大强度，并假设一个单一的全局失真F(global distortion)对所有转换进行参数化。RandAugment包含两个参数E和F。本方案对训练数据集中的每个图像利用参数E和F进行一系列随机增强操作，从而扩充了原有数据集，增加了数据多样性，模型在训练期间将更好地适应不同的数据变化，可减少过拟合的风险，提升模型的鲁棒性和准确性。

(2)通道个性标准化。数据规范化是数据挖掘中的数据变换的一种方式，数据变换将数据变换或统一成适合于数据挖掘的形式，将被挖掘对象的属性数据按比例缩放，使其落入一个小的特定区间内，如[–1,1]或[0,1]。

对属性值进行规范化常用于涉及神经网络和距离度量的分类算法和聚类算法当中。比如使用神经网络反向传播算法进行分类挖掘时，对训练元组中度量每个属性的输入值进行规范化有利于加快学习阶段的速度。本文基于Z-Score标准化(Zero-Score Normalization)，在数据集内部使用式(3)取每个值x并将其转换为对应的z值

$$z=\frac{x-\text{mean}}{\text{std}} \tag{3}$$

其中，mean是平均值，std是标准差。

对于常规的标准化操作而言，式(3)中的mean值和std值通常默认设置为0.5，而本方案基于Z-Score设计了通道个性标准化方法，该方法会对特征偏移的数据集分别专门地计算其mean和std。这样做可以更好地适应每个数据集的数据分布特点。不同数据集可能有不同的数据统计特性，如果将它们都强制调整到相同的0.5均值和0.5标准差，可能会导致信息丢失或模型性能下降。因此，本方案个性化地计算均值和标准差可以更好地保留特征偏移数据集的数据分布特性，有助于提高模型的泛化性能。

需要注意的是，当对数据集进行标准化时，通常会将这些操作按照特征分组。这意味着mean和std是相对于每个待标准化的特征集而言的。如图3所示，对本文而言处理的是图像，即特征是RGB颜色通道，则对每个颜色通道分别进行标准化。考虑到图像中所有像素的平均值和标准差，计算出相应颜色通道的mean和std。

对数据集中的每个x值进行如上计算后，得到新的标准化数据集Z。其中mean和std是对于各颜色通道而言的。假设给定颜色通道S中有n个数据，S通道的mean和std定义为

$$\text{mean}=\frac{1}{n}\left(\sum_{i=1}^{n}x_i\right) \tag{4}$$

$$\text{std}=\sqrt{\frac{1}{n}\left(\sum_{i=1}^{n}(x_i-\text{mean})^2\right)} \tag{5}$$

经过通道个性标准化得到的结果是，对于每个属性来说所有数据都服从均值为0，标准差为1的标准正态分布。当数据分布接近标准正态分布时，梯度下降等优化算法通常更容易找到最优解，训练过程更加稳定，帮助模型更快收敛。

(3)本地自适应更新。在传统联邦学习中(例如FedAvg)，第t轮通信时，边缘服务器下发上一轮聚合的全局模型$\boldsymbol{\theta}^{t-1}$给客户端$i$，由全局模型$\theta^{t-1}$直接覆盖客户端$i$的上一轮本地模型$\boldsymbol{\theta}_i^{t-1}$，以获得用于这一轮本地训练的初始化本地模型$\hat{\boldsymbol{\theta}}_i^t$，即$\hat{\theta}_i^t=\theta^{t-1}$。而本方案中，本地客户端聚合全局模型和本地模型，而不是直接覆盖，从而保留了客户端的个性化特征。将$\left(\boldsymbol{\theta}^{t-1}-\boldsymbol{\theta}_i^{t-1}\right)$视为“更新项”，则初始化本地模型可以定义为

$$\hat{\boldsymbol{\theta}}_i^t=\boldsymbol{\theta}_i^{t-1}+\left(\boldsymbol{\theta}^{t-1}-\boldsymbol{\theta}_i^{t-1}\right)\odot\boldsymbol{W}_i \tag{6}$$

其中，$\boldsymbol{W}_i$是聚合权重，$\odot$为哈达玛积，是矩阵的一类运算，代表两个矩阵对应元素相乘，且两矩阵必须维度相等。

研究表明神经网络的底层学习更通用的信息，而高层学习更加具体和特定的信息。在完全重叠特征偏移的场景中，底层神经网络主要学习较为简单的特征，如边缘、角点、纹理等。例如对MNIST

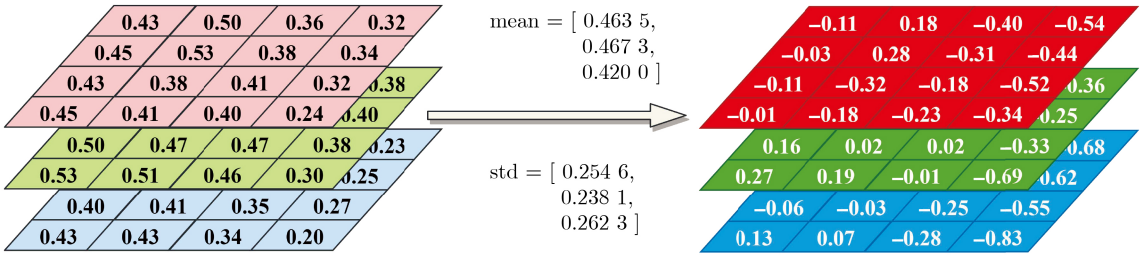

图 3 通道个性标准化

数据集而言，底层特征可能会对数字的一些基本形状和轮廓进行响应。这些特征对于数字识别来说是相对通用的，可以在不同数字之间共享。随着神经网络的深度增加，高层网络会逐渐学习更加抽象和复杂的特征。例如数字的连接部分、轮廓组合等。高层特征更加针对具体数字的识别，因此可能会在不同数据集中显示出更大的差异性。为此本文引入超参数p，其作用是控制哪些层的参数会被自适应本地聚合所影响。具体来说，p控制着本地聚合在网络的哪些高层进行。例如图2所示，如果$p=3$，那么本地聚合将仅应用于前3层(较高层)，而更低层的参数将像标准的联邦平均(FedAvg)一样进行本地初始化。

通过引入超参数p的本地自适应更新方法给联邦学习框架带来了两个优势。一是信息传递与特征重用。自适应本地聚合在高层使用客户端数据进行参数更新，这样的更新可以在一定程度上保留个性化特征。然而，低层参数仍然保留了更通用的特征，因此模型可以在不同客户端之间传递和重用有价值的信息。这种信息传递和特征重用有助于提升模型的泛化性能。二是减少计算开销。若对模型的每一层都进行自适应本地聚合可能会导致较大的计算开销。通过超参数p，可以选择只在较高层上应用本地聚合，这种分层的策略可以降低计算开销。减少计算开销意味着可以更快地进行模型训练，同时也减少了由于大量计算导致的过拟合风险，从而提升模型的泛化能力。具体定义为

$$\hat{\boldsymbol{\theta}}_i^t=\boldsymbol{\theta}_i^{t-1}+\left(\boldsymbol{\theta}^{t-1}-\boldsymbol{\theta}_i^{t-1}\right)\odot\left[1^{|\boldsymbol{\theta}_i|-p};\boldsymbol{W}_i^p\right] \tag{7}$$

其中，$|\boldsymbol{\theta}_i|$是上一轮本地模型θ_i^{t-1}中的层数，$1^{|\boldsymbol{\theta}_i|-p}$具有与$\boldsymbol{\theta}_i^{t-1}$中$p$层以下相同的形状，且元素均为1(常量)。聚合权重$W_i^p$则是具有与$\boldsymbol{\theta}_i^{t-1}$中$p$层以上相同的形状。本方案在开始时将$\boldsymbol{W}_i^p$中的每个元素值初始化为1，并在每轮通信$t$中基于旧的$\boldsymbol{W}_i^p$学习新的$\boldsymbol{W}_i^p$。为进一步减少开销，每轮通信$t$中随机采样$D_i$的$s\%$，并将其表示为$D_i^t(s)$。客户端$i$通过基于梯度的学习方法训练$\boldsymbol{W}_i^p$

$$\boldsymbol{W}_i^p\leftarrow\boldsymbol{W}_i^p-\eta\nabla\boldsymbol{W}_i^pL\left(\hat{\boldsymbol{\theta}}_i^t,D_i^t(s);\boldsymbol{\theta}^{t-1}\right) \tag{8}$$

其中，η是权重学习的学习率。在训练$\boldsymbol{W}_i^p$时，其他可训练参数均被冻结，包括整个全局模型和整个本地模型。在本地初始化完成之后，客户端i才会解冻模型，并基于数据集以及mean和std正常进行本地模型训练。

(4)全局模型聚合。联邦学习中全局模型聚合通常是指在各个客户端上进行局部训练后，将它们的本地模型参数汇总以生成全局模型。这个过程是联邦学习的核心部分，允许不同客户端的学习经验合并，以便全局模型能够从所有参与方的数据中受益。本文采用一种常见的联邦平均聚合方法，其中客户端i的本地模型参数按权重r_i进行平均。如式(2)所示，本地样本数据量权重r_i与客户端i的数据大小相关，拥有更多数据的客户端具有更大的权重。这种权重设计确保了对于数据更多的客户端，其贡献更大，以反映数据分布的不平衡。边缘服务器基于r_i和本地模型$\boldsymbol{\theta}_i^t$聚合生成全局模型$\boldsymbol{\theta}^t$，具体定义为

$$\boldsymbol{\theta}^t\leftarrow\sum_{i\in N}r_i\boldsymbol{\theta}_i^t \tag{9}$$

为便于理解上述执行流程，将此求解过程描述为算法1的形式。

算法1中由于客户端的聚合权重$\boldsymbol{W}_i^p$在第2轮训练直至收敛，并在随后的迭代中几乎没有变化，即W_i^p可以重复使用。因此在第2轮以后，客户端只为W_i^p训练1次，以适应不断变化的模型参数。需要注意的是，聚合权重在第1轮训练中是无意义的，因为此时$\boldsymbol{\theta}^0=\boldsymbol{\theta}_i^0$，客户端统一使用边缘服务器下发的初始化全局模型$\boldsymbol{\theta}^0$，所以无需聚合。

4 仿真与性能评估

本节通过仿真实验来评估融合通道个性标准化的本地自适应联邦学习算法的有效性，并将本文所提出的算法与其他经典基准方案进行对比，以突出本文算法的性能优势。

本仿真实验使用基准数字分类任务进行实证分析，该任务包含不同的数据集，其中每个数据集的数据具有异构的外观，但具有相同的标签和标签分布，符合完全重叠特征偏移。具体而言，使用了以下5个数据集：MNIST, MNIST-M, SVHN, USPS和SynthDigits。

MNIST数据集由70 000张像素为28×28的手写数字图像组成，包含0～9的10种数字，其中包含60 000张训练图像和10 000张测试图像；MNIST-M数据集是通过将MNIST数据集中的数字图像嵌入到不同的彩色背景中而生成，这个背景包含了各种颜色、纹理和噪声，使得图像的外观变得多样化；SVHN数据集由Google街景车辆拍摄的自然场景中的房屋门牌号码组成；USPS数据集由美国邮政服务(USPS)提供，包含大量的手写数字图像，这些图像包含了邮政编码中的数字，且都是16×16像素的灰度图像；SynthDigits数据集是通过在计算机上合成生成，而不是从真实世界中采集的手写数字图像，每个图像都是灰度图像，通常呈现在一个固定的背景上，以使数字字符突出。

算法 1　融合通道个性标准化的本地自适应联邦学习算法

输入：客户端数量N；
损失函数L；
客户端本地样本数据量权重r；
初始化全局模型$\boldsymbol{\theta}^0$；
本地模型学习率α；
聚合权重学习率η；
客户端数据采样率$s\%$；
自适应聚合层数p.

输出：训练后的本地模型$\boldsymbol{\theta}_1^t, \boldsymbol{\theta}_2^t, \cdots, \boldsymbol{\theta}_N^t$以及全局模型$\boldsymbol{\theta}^t$.

(1) **BEGIN**
(2) 边缘服务器向所有客户端发送$\boldsymbol{\theta}^0$以初始化本地模型；
(3) 所有客户端将聚合权重$\boldsymbol{W}_i^p, \forall i \in [N]$初始化为$\boldsymbol{1}$；
(4) FOR t IN 通信轮次T DO
(5)　　服务器将$\boldsymbol{\theta}^{t-1}$发送给所有客户端；
(6)　　FOR 所有客户端并行 DO
(7)　　客户端i采样$s\%$本地数据；
(8)　　客户端i对本地数据随机增强；
(9)　　客户端i基于式(3)通道个性标准化本地数据；
(10)　　IF $t = 2$ THEN
(11)　　WHILE $\boldsymbol{W}_i^p$不收敛 DO
(12)　　客户端i基于式(8)训练$\boldsymbol{W}_i^p$；
(13)　　ELSE IF $t > 2$ THEN
(14)　　客户端i基于式(8)训练$\boldsymbol{W}_i^p$；
(15)　　客户端i基于式(7)聚合出$\hat{\theta}_i^t$用以本地训练；
(16)　　客户端i基于本地训练获得
$\boldsymbol{\theta}_i^t \boldsymbol{\theta}_i^t \leftarrow \hat{\boldsymbol{\theta}}_i^t - \alpha \nabla \hat{\boldsymbol{\theta}}_i^t L\left(\hat{\boldsymbol{\theta}}_i^t, D_i; \boldsymbol{\theta}^{t-1}\right)$；
(17)　　客户端i上传$\boldsymbol{\theta}_i^t$给边缘服务器以聚合；
(18)　　END FOR
(19)　　服务器基于式(9)聚合全局模型$\boldsymbol{\theta}^t$；
(20) END FOR
(21) **END**

设定参与训练的联邦学习服务器个数为1，通信轮次(Communication rounds)为20，基于融合通道个性标准化的本地自适应联邦学习算法，聚合权重学习率为1.0；参与训练的客户端个数为5，且每个客户端分配不同的数据集数据，客户端数据采样率为80%，自适应聚合层数为5，本地训练使用随机梯度下降算法，其学习率为0.005，参数batch_size为32。

本文模型与现有的3种联邦学习模型：联邦批量归一化算法(Federated averaging with local Batch Normalization, FedBN) [27]、联邦近端项算法 (Federated oPtimization in heterogeneous networks, FedProx) [20]、联邦平均算法 (Federated averaging, Fedavg) [16]进行对比，其中FedBN在本地模型中加入批量归一化层(Batch Normalization)来解决联邦学习数据异构性中的特征偏移情况，与本文考虑联邦场景类似，具有对比实验价值。

本实验主要分为两个部分，旨在评估融合通道个性标准化的本地自适应联邦学习算法的性能，以下是实验设计的详细说明。

两部分实验都先对本地数据集进行随机增强操作。具体而言设计了包含14种可用变换的操作列表，变换次数E为2，即从操作列表中随机选择2个增强操作，并将全局失真F设为10。

(1)客户端训练与测试。这部分实验创建了5个客户端，每个客户端分别基于自己的本地数据集进行模型训练，并使用所有数据集的测试集进行性能评估。

图4展示了此部分实验的实验设置。客户端1～5分别分配数据集MNIST, MNIST-M, SVHN, SynthDigits和USPS。训练集由客户端本地数据集按80%随机采样生成，测试集由所有客户端的测试集拼接而成。图5绘制了5个客户端以及全局模型的网络模型准确率(Accuracy)随通信轮次(Communication rounds)的变化曲线。从中可以看出，本文算法的Accuracy高于另外3种算法且都达到了90%以上，表明其具有较高的准确率。同时可以看出，基于完全重叠特征偏移中任意一个数据集训练出来的网络模型，本文算法的模型收敛速度都优于其他的算法。这是因为通道个性标准化允许每个客户端根据其数据的特性个性地进行标准化，即对不同数据集的每个颜色通道分别计算其均值和标准差。这意味着在训练过程中，模型在每个通道上都会得到更接近于标准正态分布的数据分布。数据分布的对齐有助于加速模型的收敛，当数据分布接近标准正态分布时，梯度下降等优化算法通常更容易找到最优解。

(2)跨数据集测试。这部分实验中创建了3个客

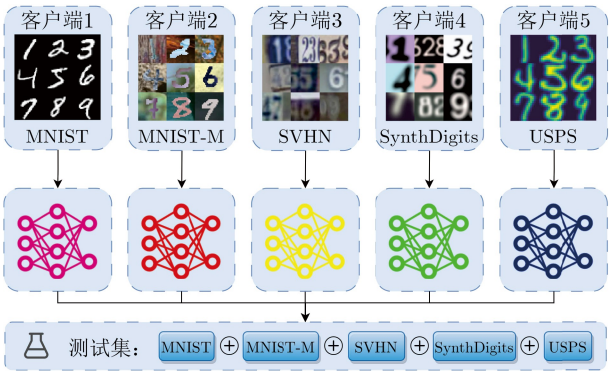

图 4 实验设置

户端，每个客户端分别基于自己的本地数据集进行模型训练，但在测试时将模型在另外两个数据集的测试集上进行了性能评估。客户端1～3分别分配数据集MNIST, SVHN和SynthDigits。训练集由客户端本地数据集按80%随机采样生成，测试集由数据集MNIST-M和USPS的测试集拼接而成。

图6绘制了3个客户端以及全局模型的网络模型准确率随Communication rounds的变化曲线。从

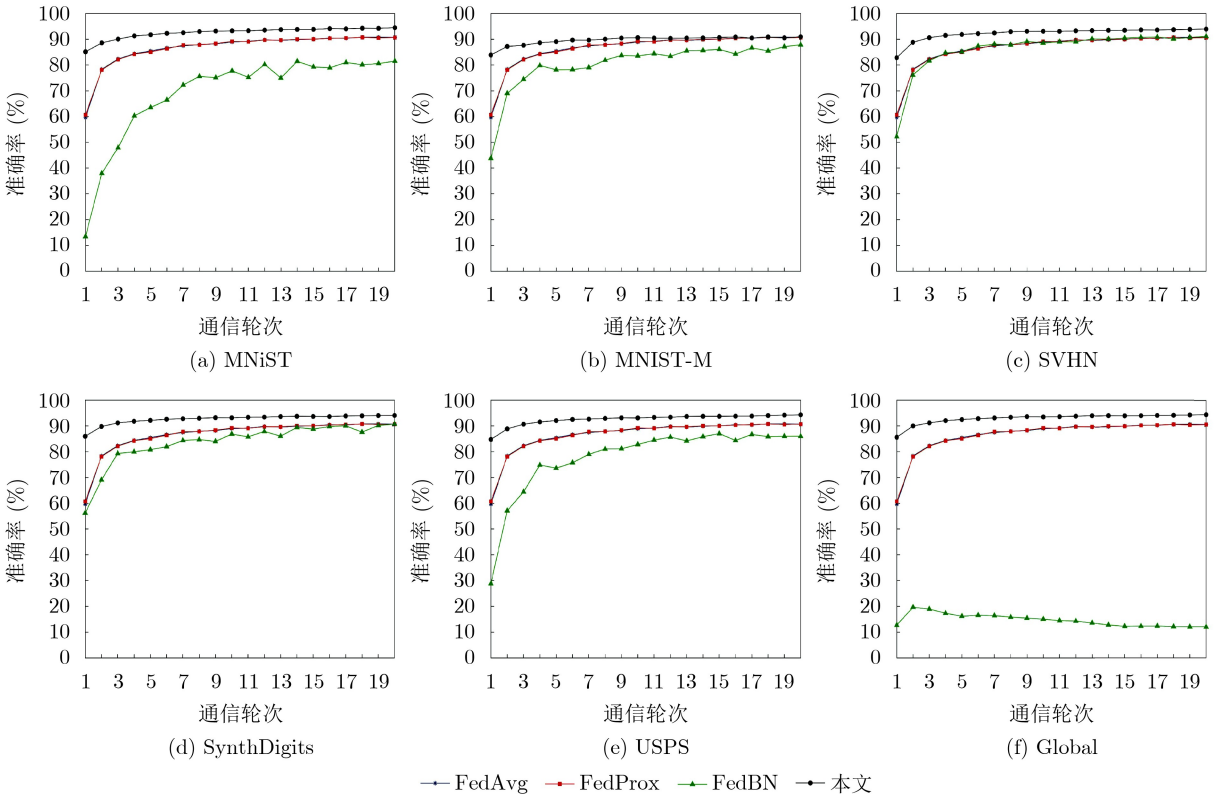

图 5 准确率变化曲线

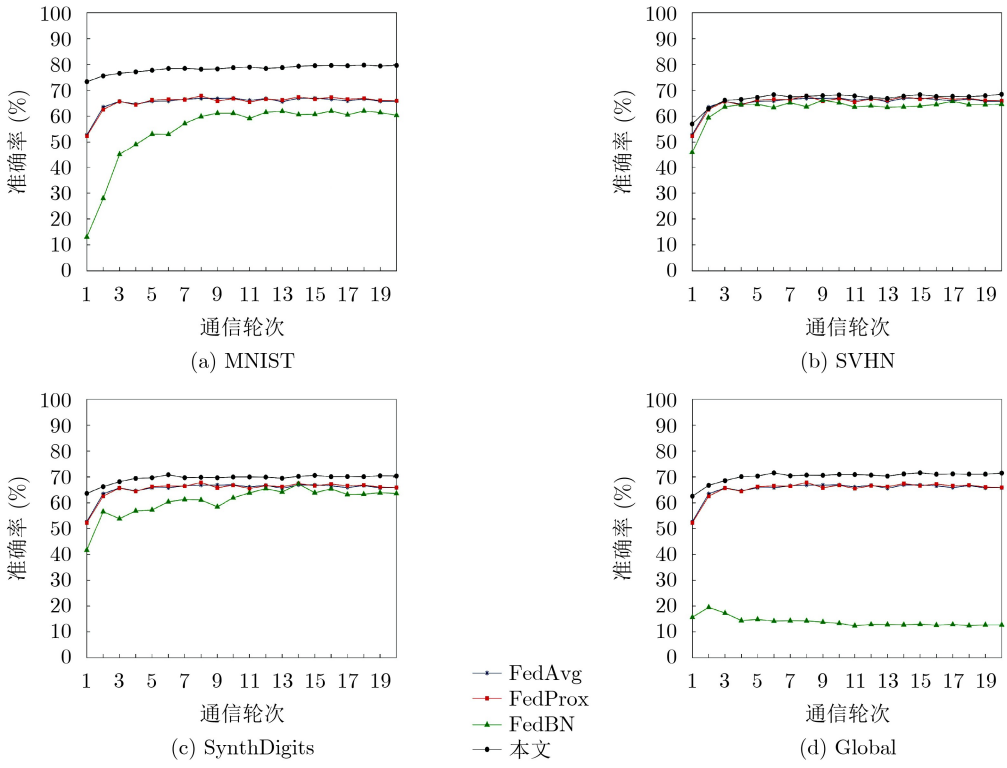

图 6 准确率变化曲线

该图可以看出，所有算法的模型准确率相较于前一部分实验均有所降低，这是因为测试集是由未参与训练全新数据集构成，客户端在测试前并未学习过这些知识，而本文算法依旧取得了最优的表现。这表明通过本文算法训练的模型在面对新的数据时的表现能力更好，模型的泛化性能更强。

这主要有两部分原因。一是在训练前，本文算法使用了RandAugment数据增强方法，扩充了原始数据集，增加了数据的多样性。这有助于模型更好地适应不同数据变化，同时减少了过拟合风险，从而在跨数据集测试中提高了性能。二是本地自适应更新策略，在高层网络上聚合全局模型和本地模型。通常高层网络更专注于抽象和复杂的特征学习。这些高层特征在不同客户端之间可能会有一定的共享性，尤其是对于完全重叠偏移的特征，如轮廓组合、连接部分等。通过自适应更新，模型可以在一定程度上传递有用的信息和重用特征。这有助于提高模型的泛化性能，因为模型可以更好地利用其他客户端的有用信息。

5　结束语

为了降低数据异构对联邦学习的负面影响，本文提出了一种融合通道个性标准化的本地自适应联邦学习算法。具体地，构建了一个面向数据特征偏移的联邦学习模型，通过自适应地聚合全局模型和本地模型以进行本地初始化，利用这种聚合方式初始化的本地模型，一方面保留了客户端模型的个性化特征，另一方面从全局模型中捕获到的所需信息有助于提升模型的泛化性能。同时，采用数据增强和通道个性标准化的方法对各客户端数据集进行数据预处理，以达到避免过拟合和加快模型收敛速度的目的。最后，与现有算法的对比结果表明，本文算法可以显著提高图像分类识别的准确性，且收敛速度更快。

参 考 文 献

[1] KAISSIS G A, MAKOWSKI M R, RÜCKERT D, *et al*. Secure, privacy-preserving and federated machine learning in medical imaging[J]. *Nature Machine Intelligence*, 2020, 2(6): 305–311. doi: 10.1038/s42256-020-0186-1.

[2] TANG Zhenheng, SHI Shaohuai, and CHU Xiaowen. Communication-efficient decentralized learning with sparsification and adaptive peer selection[C]. IEEE 40th International Conference on Distributed Computing Systems (ICDCS), Singapore, 2020: 1207–1208. doi: 10.1109/ICDCS47774.2020.00153.

[3] YANG Qiang, LIU Yang, CHEN Tianjian, *et al*. Federated machine learning: Concept and applications[J]. *ACM Transactions on Intelligent Systems and Technology*, 2019, 10(2): 12. doi: 10.1145/3298981.

[4] CHEN Yang, SUN Xiaoyan, and JIN Yaochu. Communication-efficient federated deep learning with layerwise asynchronous model update and temporally weighted aggregation[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2020, 31(10): 4229–4238. doi: 10.1109/TNNLS.2019.2953131.

[5] WU Donglei, ZOU Xiangyu, ZHANG Shuyu, *et al*. SmartIdx: Reducing communication cost in federated learning by exploiting the CNNs structures[C]. The 36th AAAI Conference on Artificial Intelligence, 2022: 4254–4262. doi: 10.1609/aaai.v36i4.20345.

[6] MILLS J, HU Jia, and MIN Geyong. Communication-efficient federated learning for wireless edge intelligence in IoT[J]. *IEEE Internet of Things Journal*, 2020, 7(7): 5986–5994. doi: 10.1109/JIOT.2019.2956615.

[7] SATTLER F, WIEDEMANN S, MÜLLER K R, *et al*. Robust and communication-efficient federated learning from non-I. I. D. data[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2020, 31(9): 3400–3413. doi: 10.1109/TNNLS.2019.2944481.

[8] DUAN Moming, LIU Duo, CHEN Xianzhang, *et al*. Astraea: Self-balancing federated learning for improving classification accuracy of mobile deep learning applications[C]. IEEE 37th International Conference on Computer Design (ICCD), Abu Dhabi, United Arab Emirates, 2019: 246–254. doi: 10.1109/ICCD46524.2019.00038.

[9] LIM W Y B, LUONG N C, HOANG D T, *et al*. Federated learning in mobile edge networks: A comprehensive survey[J]. *IEEE Communications Surveys & Tutorials*, 2020, 22(3): 2031–2063. doi: 10.1109/COMST.2020.2986024.

[10] MENDIETA M, YANG Taojiannan, WANG Pu, *et al*. Local learning matters: Rethinking data heterogeneity in federated learning[C]. The 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, USA, 2022: 8387–8396. doi: 10.1109/CVPR52688.2022.00821.

[11] WANG Jianyu, LIU Qinghua, LIANG Hao, *et al*. Tackling the objective inconsistency problem in heterogeneous federated optimization[C]. The 34th International Conference on Neural Information Processing Systems, Vancouver, Canada, 2020: 638. doi: 10.48550/arXiv.2007.07481.

[12] LI Qinbin, HE Bingsheng, and SONG D. Model-contrastive federated learning[C]. The 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, USA, 2021: 10708–10717. doi: 10.1109/CVPR46437.2021.01057.

[13] GHOSH A, CHUNG J, YIN Dong, *et al*. An efficient framework for clustered federated learning[J]. *IEEE Transactions on Information Theory*, 2022, 68(12): 8076–8091. doi: 10.1109/TIT.2022.3192506.

[14] SATTLER F, MÜLLER K R, and SAMEK W. Clustered federated learning: Model-agnostic distributed multitask optimization under privacy constraints[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2021, 32(8): 3710–3722. doi: 10.1109/TNNLS.2020.3015958.

[15] KULKARNI V, KULKARNI M, and PANT A. Survey of personalization techniques for federated learning[C]. The 2020 Fourth World Conference on Smart Trends in Systems, Security and Sustainability (WorldS4), London, UK, 2020: 794–797. doi: 10.1109/WorldS450073.2020.9210355.

[16] MCMAHAN B, MOORE E, RAMAGE D, *et al*. Communication-efficient learning of deep networks from decentralized data[C]. The 20th International Conference on Artificial Intelligence and Statistics (AISTATS), Fort Lauderdale, USA, 2017: 1273–1282.

[17] ZHAO Zhongyuan, FENG Chenyuan, HONG Wei, *et al*. Federated learning with non-iid data in wireless networks[J]. *IEEE Transactions on Wireless Communications*, 2021, 21(3): 1927–1942. doi: 10.1109/TWC.2021.3108197.

[18] KE Guolin, MENG Qi, FINLEY T, *et al*. LightGBM: A highly efficient gradient boosting decision tree[C]. The 31st International Conference on Neural Information Processing Systems, Long Beach, USA, 2017: 3149–3157.

[19] LI Qinbin, DIAO Yiqun, CHEN Quan, *et al*. Federated learning on non-IID data silos: An experimental study[C]. 2022 IEEE 38th International Conference on Data Engineering, Kuala Lumpur, Malaysia, 2022: 965–978. doi: 10.1109/ICDE53745.2022.00077.

[20] LI Tian, SAHU A K, ZAHEER M, *et al*. Federated optimization in heterogeneous networks[C]. Machine Learning and Systems, Austin, USA, 2020: 1–22. doi: 10.48550/arXiv.1812.06127.

[21] WANG Hao, KAPLAN Z, NIU Di, *et al*. Optimizing federated learning on non-iid data with reinforcement learning[C]. The IEEE INFOCOM 2020 - IEEE Conference on Computer Communications, Toronto, Canada, 2020: 1698–1707. doi: 10.1109/INFOCOM41043.2020.9155494.

[22] WANG Hongyi, YUROCHKIN M, SUN Yuekai, *et al*. Federated learning with matched averaging[C]. The 8th International Conference on Learning Representations, Addis Ababa, Ethiopia, 2020: 1–16.

[23] HUANG Yutao, CHU Lingyang, ZHOU Zirui, *et al*. Personalized cross-silo federated learning on non-IID data[C]. The 35th AAAI Conference on Artificial Intelligence, 2021: 7865–7873. doi: 10.1609/aaai.v35i9.16960.

[24] LI Xinchun, ZHAN Dechuan, SHAO Yunfeng, *et al*. FedPHP: Federated personalization with inherited private models[C]. Joint European Conference on Machine Learning and Knowledge Discovery in Databases, Bilbao, Spain, 2021: 587–602. doi: 10.1007/978-3-030-86486-6_36.

[25] ZHANG M, SAPRA K, FIDLER S, *et al*. Personalized federated learning with first order model optimization[C]. The 9th International Conference on Learning Representations (ICLR), 2021: 1–17.

[26] LUO Jun and WU Shandong. Adapt to adaptation: Learning personalization for cross-silo federated learning[C]. 31st International Joint Conference on Artificial Intelligence, Vienna, Austria, 2022: 2166–2173.

[27] LI Xiaoxiao, JIANG Meirui, ZHANG Xiaofei, *et al*. FedBN: Federated learning on non-IID features via local batch normalization[C]. The 9th International Conference on Learning Representations, 2021: 1–27.

赵 宇：男，博士生，研究方向为联邦学习与边缘智能.
陈思光：男，博士，教授，研究方向为边缘智能与安全.

责任编辑：余 蓉