

# 基于异步分层联邦学习的数据异质性处理方法研究\*

郭昌昊, 唐湘云, 翁 彧

(中央民族大学信息工程学院, 北京 100081)

**摘要:**在物联网设备遍布的时代,时刻都在产生大量数据,数据分布和数据量各不相同,因此数据异质性普遍存在。针对物联网环境中智能设备的联邦学习挑战,传统联邦学习的同步机制解决数据异质性(NON-IID)问题并不理想,且面临着单点故障和维护全局时钟的复杂性问题,而异步机制则可能带来额外的通信开销和 NON-IID 数据分布导致的过时性问题。分层联邦学习结合异步机制在应对数据异质性的问题时更加灵活,为此,提出了一种基于分层联邦学习的异步分层联邦学习方法。首先,使用 BIRCH 算法分析物联网中各节点的数据分布并进行簇的划分;然后,对簇中的数据进行拆分与验证,目的是找到数据质量高的节点,然后将数据质量高的簇中的节点打散,重组到其他数据质量低的簇中,形成新的簇;最后,进行簇内聚合和全局聚合的两阶段模型训练。此外,基于 MNIST 数据集,对提出的方法进行了评估。结果表明,与经典方法相比,所提方法在 NON-IID 数据集上收敛速度提高,而且在模型精度上提高了 15% 以上。

**关键词:**物联网;联邦学习;异步联邦学习;分层联邦学习;数据异质性;数据分布

**中图分类号:**TP391

**文献标志码:**A

**doi:**10.3969/j.issn.1007-130X.2024.07.012

## A data heterogeneity processing method based on asynchronous hierarchical federated learning

GUO Chang-hao, TANG Xiang-yun, WENG Yu

(School of Information Engineering, Minzu University of China, Beijing 100081, China)

**Abstract:** In the era of ubiquitous Internet of Things devices, a vast amount of data with varying distributions and volumes is continuously generated, leading to pervasive data heterogeneity. Addressing the challenges of federated learning for intelligent devices in the IoT landscape, traditional synchronous federated learning mechanisms fall short in effectively tackling the NON-IID data distribution problem. Moreover, they are plagued by issues such as single-point failures and the complexity of maintaining a global clock. However, asynchronous mechanisms may introduce additional communication overhead and obsolescence due to NON-IID data distribution. To offer a more flexible solution to these challenges, an asynchronous hierarchical federated learning method is proposed. Initially, the BIRCH algorithm is employed to analyze the data distribution across various IoT nodes, leading to the formation of clusters. Subsequently, data within these clusters is dissected and validated to identify nodes with high data quality. Nodes from high-quality clusters are then disaggregated and reorganized into lower-quality clusters, forming new, optimized clusters. Finally, a two-stage model training is conducted, involving both intra-cluster and global aggregation. Additionally, our proposed approach is evaluated using the MNIST dataset. The results show that, compared to the baseline set by the classical FedAVG method, the proposed approach achieves faster convergence on NON-IID datasets and improves model accuracy by

\* 收稿日期:2023-10-20;修回日期:2023-11-23

基金项目:国家自然科学基金青年基金(62302539);中央民族大学国家安全研究院边疆少数民族地区国家安全研究项目(2023GJAJQ08)

通信作者:唐湘云(xiangyunt@muc.edu.cn)

通信地址:100081 北京市海淀区中央民族大学信息工程学院

Address: School of Information Engineering, Minzu University of China, Haidian District, Beijing 100081, P. R. China

more than 15%.

**Key words:** Internet of Things (IoT); federated learning; asynchronous federated learning; hierarchical federated learning; non-independent and identically distributed data; data distribution

## 1 引言

伴随着物联网的日益发展,智能设备的数量,如智能手机、可穿戴设备或其他物联网设备,在过去几年中一直在迅速增长,这些设备每秒钟都在产生大量数据<sup>[1]</sup>。传统上,这些大量的数据会被转发到远程云端进行处理和训练,但是因为长距离传输可能会造成严重的延迟,并且可能会泄露隐私。在这种背景下,联邦学习(Federated Learning)<sup>[2]</sup>显然是一种合适的解决方案。其允许在不共享数据的情况下进行模型的训练,通过在本地设备上进行的模型训练,然后将模型参数上传到中央服务器,在服务器上进行模型聚合。联邦学习的主要优点是:(1)实现了保护隐私的协作机器学习。(2)利用物联网设备的计算能力进行本地模型训练,从而减少了云的计算工作量。(3)模型参数通常比原始数据更小,从而减轻了骨干通信网络的负担。这已经实现了一些实际应用,例如,在车联网、医疗保健、无人机传感和移动边缘计算<sup>[3]</sup>中。

现有的大多数联邦学习方法都是同步的,但是在实际的运行环境中,同步联邦学习面临2个问题:第1个问题是单点故障问题;第2个问题是在大规模的物联网节点下维护一个完全同步的全局时钟非常复杂。异步联邦学习(Asynchronous Federated Learning)<sup>[4]</sup>旨在解决同步联邦学习的挑战。具体来说,在异步联邦学习中,服务器在接收到客户端的模型后可以立即更新全局模型,而无需等待其他客户端的模型。

然而,现有的异步联邦学习机制在物联网背景下的数据时代主要有2个缺点:首先,异步联邦学习机制可能会导致大量的通信资源消耗。其次,客户端采集数据的数据量、数据大小和数据类型也不同,导致客户端参与异步更新的相对频率也不同,且局部模型相对于全局模型有较大的过时性,进而导致训练精度较低。然而,目前的方法在提高精度和节约通信成本方面存在着两难问题<sup>[5]</sup>。

在这种情况下,Liu等人<sup>[6]</sup>提出了分层联邦学习(Hierarchical Federated Learning)框架,如图1所示。其中各节点不直接与全局服务器通信,而是通过分层后进行一次中间聚合。然后,由聚合在中

间层的节点进一步与参数服务器建立通信,实现全局聚合。如果节点间的数据分布具有较多的不同,那么通过简单的同步机制进行分层联邦学习可能会出现模型过拟合或欠拟合的情况,即在某些节点上产生的模型无法很好地泛化到其他节点上,或者在少数节点上产生的模型性能表现不佳。

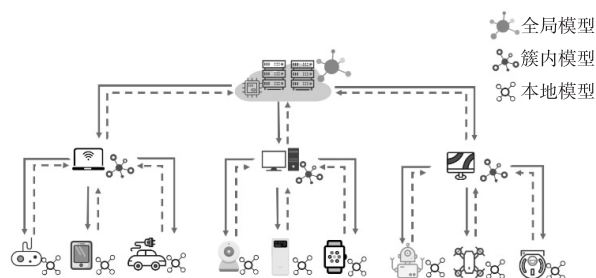


Figure 1 Framework of hierarchical federated learning

图1 分层联邦学习框架

将分层联邦学习与异步机制结合后可以根据各节点的处理能力和存储容量进行灵活的调整,具有更强的鲁棒性。

作为在物联网和大数据的背景下解决通信开销和数据异质性 NON-IID (NON-Independent and Identically Distributed) 的尝试,本文设计了一种根据数据异质性分布划分节点的异步分层联邦学习方法,使用 BIRCH (Balanced Iterative Reducing and Clustering using Hierarchies) 算法<sup>[7]</sup>对物联网环境下所有节点的数据进行计算后再划分成簇。

对部分簇内的节点进行数据拆分与验证,具体来说是对部分簇之间的模型进行快速评估,通过模型准确率泛化程度找到数据质量高的节点;将数据质量高的簇中的节点打散,重新划分到其他簇中,作为簇头;整个训练过程为两阶段提交,簇内和簇外为异步模式,整体划分为两阶段提交。

本文的主要工作总结如下:

(1)设计应用于物联网背景下包含大量数据节点的异步分层联邦学习方法,这是一种创新的方法,它根据数据的异质性来划分簇。

(2)提出了一种有效的机制来应对 NON-IID 数据分布的情况,从而提高模型训练的准确性。

(3)在测试平台上实现了提出的异步分层联邦学习机制,实验结果表明本文提出的方法可行有效。

2 相关工作

在物联网的大背景下,异步机制与分层联邦学习结合后相较传统的联邦学习更有优势,具体体现在异步分层联邦学习可以充分利用物联网设备的计算能力,且分层后可以根据各节点的处理能力和存储容量进行灵活的调整,减少了总体训练时间。其中,2个异步分层联邦学习的方向值得我们研究,一是减少异步分层联邦学习的通信开销,二是提高 NON-IID 数据下异步分层联邦学习模型的精度。

2.1 在异步分层联邦学习的通信开销方面

由于异步分层联邦学习工作在分布式环境中,需要在客户端和协作者之间进行高效的交互,以减少对边缘有限通信资源的消耗。一般来说,有3种方法被广泛讨论,即数据模式优化、频率降低和数据压缩。具体来说,可根据节点的本地计算资源和通信条件<sup>[8]</sup>,仅选择一部分节点参与全局,这种方法减少了需要传输到中央服务器的数据量,从而降低通信开销。Liu 等人<sup>[9]</sup>提出了 AAFL(Adaptive Asynchronous Federated Learning)机制,可以智能地根据网络情况在不同时期变化本地更新模型的数量以进行全局模型聚合,来应对通信资源发生的变化;Chen 等人<sup>[10]</sup>提出对训练的模型进行分割,调整不同部分的更新频率,减少需要在客户端和服务端之间传输的数据量,但是这种优化受到机器学习模型的限制。为了减少在异步联邦学习中异步通信所产生的通信成本,使用了一种称为“Polyline 编码”的压缩方法<sup>[11]</sup>,对在客户端和服务端之间传输的模型数据进行了压缩,但压缩通常需要增加额外的计算成本,可能会导致训练时间延长。

2.2 在异步分层联邦学习的 NON-IID 方面

当 NON-IID 数据存在时,联邦学习的性能下降是由异构数据分布引起的,因此,基于数据的方法旨在通过修改数据分布来解决这个问题。首先,可以通过预训练来提高模型的初始性能<sup>[12]</sup>,从而使模型具有更好的泛化能力,能够更好地适应 NON-IID 数据;其次,具有高质量数据的节点可以加速联邦学习训练,Hu 等人<sup>[13]</sup>提出了一种聚类数据共享方案,即通过聚类算法划分簇后,簇内的各节点之间分享一部分数据,以消除分布倾斜对联邦学习收敛性和目标精度的影响。但是,联邦学习的

目标是在不共享私有数据的情况下协作训练共享模型。

然而,联邦学习训练的局部模型在性能方面可能会受到全局模型的影响,并且可能因为数据异质性的影响而无法泛化。Liu 等人<sup>[14]</sup>提出使用局部微调应对这些问题,局部微调的目的是根据局部任务对模型进行调整或约束。还可以比较局部模型与全局模型之间的差异来确定模型最优更新<sup>[15]</sup>,Zhou 等人<sup>[16]</sup>提出增加额外的正则项来抑制局部模型和全局模型之间的差异,以减少数据异质性的影响,并使模型收敛更稳定。但是,频繁地对模型进行微调势必会增加性能消耗和通信开销。

在应对 NON-IID 数据的挑战时,可以考虑整体的系统设计方案<sup>[17]</sup>,也可以通过在每个客户端学习个性化的模型来提高准确率<sup>[18]</sup>,基于自动调整客户端学习率<sup>[19]</sup>和放弃一些贡献度较差的节点来加速模型收敛的方案<sup>[20]</sup>在应对 NON-IID 数据时也取得了不错的效果,但在实际的物联网背景下,可能会导致大量的节点被放弃。

3 预备知识

本节将介绍本文涉及的3种联邦学习方法:联邦学习、异步联邦学习和分层联邦学习。本文使用的符号及其含义在表1中给出。

Table 1 Notation

表1 符号含义

	变量符号	含义
联邦学习	$\mathcal{D}_i$	客户端节点的本地数据集
	$D_i$	客户端节点的本地数据大小
	$D$	总数据量的大小
	$\mathbf{w}$	模型参数向量
	$F_i(\mathbf{w})$	客户端节点损失函数
	$F(\mathbf{w})$	全局损失函数
异步 联邦学习	$\mathbf{w}^t$	第 $t$ 次迭代的全局模型
	$\mathbf{g}_i^t$	第 $t$ 次迭代中全局模型基于客户端节点的更新
	$\tau_i$	客户端节点更新的延迟
	$\eta$	学习率
分层 联邦学习	$\mathcal{J}$	所有簇的集合
	$j$	簇头
	$J$	划分簇的总数量
	$\mathcal{C}_j$	簇内可以与簇头通信的客户端节点的集合

3.1 联邦学习

联邦学习是一种分布式机器学习方法,旨在训



练一个全局模型,同时保护用户数据的隐私。在联邦学习中,每个客户端节点保留其本地数据,通过与中央服务器进行交互来更新模型参数。本文认定有一组设备  $\mathcal{V} = \{v_1, v_2, \dots, v_N\}$ , 每个客户端设备  $v_i$  在其本地数据集  $\mathcal{D}_i$  上训练一个模型,大小为  $D_i, D_i \triangleq |\mathcal{D}_{ii}|$ , 总数据量为  $D = \sum_{v_i \in \mathcal{V}} D_i$ 。假设  $\mathcal{D}_i$  表示训练数据  $d_{i,1}, d_{i,2}, \dots, d_{i,D_i}$ , 则客户端  $v_i$  的损失函数定义如式(1)所示:

$$F_i(\mathbf{w}) \triangleq \frac{1}{D_i} \sum f(\mathbf{w}; \mathcal{D}_i) \quad (1)$$

其中,  $\mathbf{w}$  表示参数向量;  $f(\cdot)$  表示用户自定义的损失函数(如线性回归、逻辑回归和支持向量机等)。

所有分布式数据集上的全局损失函数如式(2)所示:

$$F(\mathbf{w}) \triangleq \frac{\sum_{v_i \in \mathcal{V}} F_i(\mathbf{w})}{D} \quad (2)$$

最终的学习目标是求出使  $F(\mathbf{w})$  最小的最优向量  $\mathbf{w}^*$ , 如式(3)所示:

$$\mathbf{w}^* = \operatorname{argmin}_{\mathbf{w}} F(\mathbf{w}) \quad (3)$$

### 3.2 异步联邦学习

异步联邦学习是联邦学习的一种变体,允许客户端在不同的时间进行本地更新和参数同步。在异步联邦学习中,中央服务器不再等待所有客户端完成本地更新,而是在收到任何客户端节点的更新时立即进行聚合。形式上,本文将第  $t$  次迭代中的全局模型表示为  $\mathbf{w}^t$ 。此外,用  $g_i^t$  表示基于全局模型  $\mathbf{w}^t$  从客户端的模型更新。假设在第  $t$  次迭代中,服务器从客户端  $i$  中收到一个模型更新  $g_i^{t-\tau_i}$ , 该模型更新是基于之前的全局模型  $\mathbf{w}^{t-\tau_i}$  在之前的  $t-\tau_i$  次迭代,其中  $\tau_i$  是模型更新的延迟。服务器更新全局模型如式(4)所示:

$$\mathbf{w}^{t+1} = \mathbf{w}^t - \eta g_i^{t-\tau_i} \quad (4)$$

其中  $\eta$  表示学习率。

### 3.3 分层联邦学习

分层联邦学习是一种在多层次上进行模型参数聚合的方法。在分层联邦学习中,客户端节点被划分到不同的簇(Cluster),每个簇内的客户端节点先进行局部聚合,然后再将局部聚合结果发送给中央服务器进行全局聚合。这种方法旨在减少通信开销和提高模型训练速度。

本文在整个网络中划分了  $J$  个簇头(类似基站的作用),其集合表示为  $\mathcal{J} = \{1, \dots, j, \dots, J\}$ 。每个簇内的客户端节点都可以选择与任意一个

$j \in \mathcal{J}$  的簇头关联。不失一般性,每个簇头  $j \in \mathcal{J}$  一次只能为一个模型所有者服务。 $\mathcal{C}_j = \{1, 2, \dots, C_j\}$  表示某个簇内可以与簇头通信的客户端节点的集合,大小为  $|\mathcal{C}_j|$ 。在分层联邦学习中,簇头  $j$  首先从全局服务器接收初始全局模型,即由向量  $\mathbf{w}^0$  表示的参数。然后,将全局模型传递给它的簇内客户端节点。在模型经过第  $t$  次迭代训练后,以最小化全局损失  $F(\mathbf{w})$ 。本文的目标是最小化整个系统的目标函数  $F(\mathbf{w})$ , 其中  $\mathbf{w}$  是全局模型参数,其定义如式(5)所示:

$$F(\mathbf{w}) \triangleq \frac{1}{J} \sum_{j \in \mathcal{J}} \frac{1}{|\mathcal{C}_j|} \sum_{k \in \mathcal{C}_j} F_k(\mathbf{w}) \quad (5)$$

其中,  $F_i(\cdot)$  表示本地损失函数(一般来讲,  $F_i(\cdot)$  会根据特定应用设置不同的形式,例如交叉熵损失、均方误差损失等)。

## 4 系统总览

### 4.1 系统模型

考虑一个物联网下的计算场景,该场景由一个云服务器  $S$  和一组  $N$  个边缘客户端组成。服务器  $S$  借助客户端节点的数据服务和计算服务,聚合全局模型。但是,由于大量客户端全程参与所带来的高昂的通信成本,以及训练客户端的动态状态,服务器  $S$  通常在每次迭代中从  $N$  个客户端节点中异步聚合部分客户端节点的模型,并将  $N$  个客户端划分成了  $J$  个簇头,以促进各模型聚合任务的执行。在基于物联网的大数据时代,根据特定用户的使用情况,客户端节点拥有的数据集通常具有不同的类别和数量,且以往的一些研究表明, NON-IID 数据会导致使用联邦学习训练的模型的精度损失。文献[21]进一步分析了 NON-IID 数据分布与数据精度的关系。受到这一观察的启发,有以下 3 个问题:

(1) 如何根据客户端节点的 NON-IID 数据的分布来区分客户,并根据对应的数据分布划分不同的簇;

(2) 如何找到数据分布相对均匀的簇并根据相关的条件(通信时延、通信距离和资源等)重新打散到其他簇中;

(3) 如果让数据分布均匀的客户端节点比 NON-IID 数据程度较高的客户端节点作为簇头更多地参与训练,全局模型的准确率是否可以提高。

### 4.2 问题描述

在分层异步联邦学习中,全局以及簇内的节点

之间都会使用异步机制进行模型训练和参数聚合。具体实现如下:

簇内异步机制:对于第  $j$  个簇中的节点  $n \in \mathcal{C}_j$ , 设  $\mathbf{w}_{j,n}$  表示该节点  $n$  在簇  $\mathcal{C}_j$  上的本地模型参数,  $\bar{\mathbf{w}}_j$  表示簇  $\mathcal{C}_j$  的全局平均模型参数, 则该方法的具体实现可以表示如式(6)~式(8)所示:

$$\mathbf{w}_{j,n}^{t+1} = \text{Update\_Local}(\mathbf{w}_{j,n}^t, \bar{\mathbf{w}}_j) \quad (6)$$

$$\mathbf{w}_{j,n}^{t+1} = \mathbf{w}_{j,n}^t - \eta_t \nabla F_{j,n}(\mathbf{w}_{j,n}^t) \quad (7)$$

$$\bar{\mathbf{w}}_j^{t+1} = \frac{1}{|\mathcal{C}_j|} \sum_{n \in \mathcal{C}_j} \bar{\mathbf{w}}_{j,n}^{t+1} \quad (8)$$

其中,  $\text{Update\_Local}(\cdot, \cdot)$  表示本地更新函数,  $\nabla F_{j,n}(\cdot)$  表示簇  $\mathcal{C}_j$  中客户端节点  $n$  的本地损失函数相对于该客户端节点的本地模型参数  $\mathbf{w}_{j,n}^t$  的梯度向量。

全局异步机制:设  $\mathbf{w}_b$  表示第  $b$  个簇的全局平均模型参数,  $\bar{\mathbf{w}}$  表示所有簇的全局平均模型参数, 则该方法的具体实现可以表示如式(9)和式(10)所示:

$$\mathbf{w}_j^{t+1} = \text{Update}(\mathbf{w}_j^t, \bar{\mathbf{w}}) \quad (9)$$

$$\bar{\mathbf{w}}^{t+1} = \frac{1}{J} \sum_{b=1}^J \mathbf{w}_b^{t+1} \quad (10)$$

其中,  $\text{Update}(\cdot)$  表示全局的更新函数。

需要注意的是, 本文关注的是 NON-IID 数据集的模型训练, 本文根据数据集的分布将一个联邦学习任务划分为多个同时进行的任务。通过对簇内和簇外进行分层异步训练和参数聚合, 实现多个客户端节点的模型训练, 从而达到全局模型的优化。

## 5 系统设计

本节将详细介绍本文提出的异步分层联邦学习方法。它由 3 个步骤组成: (1) BIRCH 算法对所有设备节点上的数据进行计算后划分成簇; (2) 根据划分结果对数据质量进行评估, 并重新进行划分; (3) 两阶段聚合模型训练。

### 5.1 簇的划分

如果对大量的节点划分成簇的时候, 把数据质量高的节点划分到一起, 数据质量低的划分到一起后直接进行训练, 那么只有数据质量好的簇训练的模型效果表现好, 而从整体的表现来看可能并不会提高模型的准确度。因此, 在进行异步分层联邦学习的训练过程前, 需要使用 BIRCH 算法对所有设备上的节点数据进行分析后划分簇, 簇的划分过程如图 2 所示, 其目的是对划分簇后的各个簇中的节

点进行数据拆分与验证, 选出数据质量高的节点作为簇头。

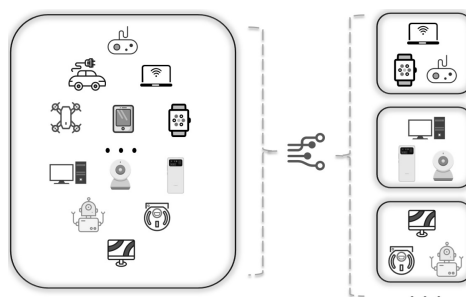


Figure 2 Partitioning clusters

图 2 划分簇

### 5.2 簇节点重组

根据 5.1 节划分的簇的输出信息, 对部分包含节点相对较少的簇进行数据集的拆分与验证, 因为考虑到真实物联网环境下的数据分布均匀的节点往往较少。具体做法是, 在选取的簇内选择任意个节点并将它们的数据集随机拆分成训练集和测试集, 并找到性能较好的节点暂时作为簇头对模型参数进行初始化或聚合; 然后将模型参数与其他簇内的模型参数进行交换; 最后, 通过各自簇内的测试集来评估模型的准确度和泛化能力。目的是为了找到数据分布均匀的节点, 即数据质量高的节点, 使其作为簇头并作为初次模型聚合的服务器。

接下来, 根据簇与簇之间的模型评估结果对簇进行标记, 划分出数据质量高的簇和数据质量低的簇。同时, 按照当前资源的上下文信息, 包括传输信息、带宽、距离和客户端可用计算资源, 将所有簇中的节点打散并重组, 如图 3 所示。旨在将数据质量良好的节点作为簇头, 领导数据质量低的节点进行模型训练, 减少数据异质性带来的影响。

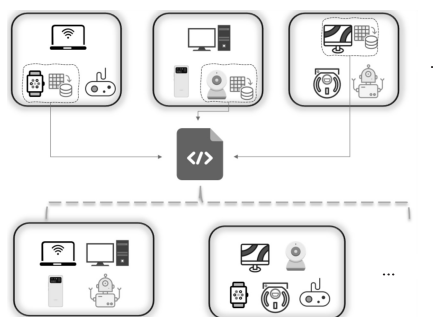


Figure 3 Re-shuffle clusters

图 3 重组簇

### 5.3 两阶段聚合

本文将整个模型训练划分为 2 个阶段, 分别是簇内聚合和全局聚合, 并且簇内和全局的聚合都是

基于异步机制。在异步环境下,客户端节点和参数服务器的发送和接收行为是无界的,因此,有3种可能的情况(如图4所示):(1)On-time received:成功接收,即服务器节点在第 $t$ 次通信轮的默认等待时间内接收到客户端节点的模型参数;(2)Late received:延时接收,即服务器节点在第 $t$ 次通信轮中,在默认等待时间之后接收客户端节点的模型参数,并在第 $t+1$ 次通信轮中使用;(3)Lost:未知故障,数据包丢失(对应的客户端节点无法对全局模型做出贡献)。

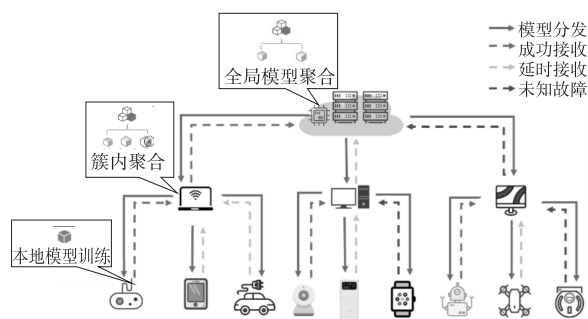


Figure 4 Two-stage aggregation

图4 两阶段聚合

### 5.3.1 第1阶段:簇内聚合

全局服务器在异步分层联邦学习中发布一个训练任务,把初始化模型分发给各个簇头节点,然后簇头节点再将初始化模型转发到各自簇内的客户端节点上开始训练。为了通过优化簇内客户端节点和簇头节点之间的交互来减少整个簇的资源消耗,并引入了模型相似度校验。具体来说,对于簇内各个客户端节点上传的模型参数,簇头节点会计算该模型参数与当前模型参数之间的余弦相似度。如果该相似度超过预设的阈值,就视为该簇内客户端节点训练得到的模型参数与当前的簇头节点的模型差异太大,如图4所示,可能会影响最终的聚合结果,因此可以将该模型参数从簇内聚合过程中剔除。

### 5.3.2 第2阶段:全局聚合

在将第1阶段各个簇内的节点所提交的模型汇总后,各簇头节点将更新后的模型参数提交至全局服务器中进行模型聚合。由于是异步提交,因此在全局聚合中,对在第 $t$ 次通信轮中默认等待时间内的簇头节点提交的模型参数先进行聚合,超出等待时间的模型参与第 $t+1$ 次通信轮的聚合。全局模型通过最小化所有簇头节点上的联合损失函数来达到最终的学习目标。然后将更新后的全局模型返回给各个簇头节点,各簇头节点再将最新的全

局模型转发给自己簇内的各客户端节点。

## 6 实验与结果分析

本节给出了数据样本在NON-IID数据集上的模拟结果,以验证本文所提出的异步联邦学习方法的有效性,并给出分析结果。

### 6.1 实验设置

为了验证本文方法的有效性,本文选择MNIST(Modified National Institute of Standards and Technology)数据集作为图像分类任务的数据集。MNIST数据集由60 000个用于训练的样本和10 000个用于测试的样本组成,其类型是10种数字。选择卷积神经网络CNN(Convolutional Neural Network)并基于MNIST数据集构建的NON-IID数据分布作为训练模型。实验设备为一台24 GB RAM, Intel® Core™ i7-9750H CPU和NVIDIA® GeForce GTX™-1650 GPU的笔记本电脑。实验环境基于Anaconda 22.9,环境配置为Python 3.7, PyTorch 1.13和TensorFlow 2.11。

实验过程中模型的部分参数如表2所示。其中 $frac$ 表示基于MNIST数据集所构建的新数据集中IID数据所占的比例,本次实验共构建3种NON-IID数据集:(1)10% IID和90% NON-IID;(2)20% IID和80% NON-IID;(3)100% NON-IID。本次实验模拟了20个客户端节点,在经过50次的全局迭代次数下,分别将模型在3种不同数据分布类型的数据集上进行训练,并查看输出结果。

Table 2 Experimental parameters

表2 实验参数

参数	参数含义	参数值
<i>epochs</i>	全局迭代次数	50
<i>batch size</i>	批次大小	128
<i>local_ep</i>	本地迭代次数	5
<i>local_bs</i>	本地批次大小	10
<i>lr</i>	学习率	0.01
<i>frac</i>	IID数据集比例	0.1/0.2

### 6.2 实验思路

为了模拟重组后的各个簇,本次实验分别按照10%和20%比率来模拟簇内包含高质量数据节点的占比。具体来说,当前模拟了20个客户端节点,假设选择了10%的高质量数据节点占比,即有2个客户端节点的数据分布为IID类型而其他18个客户端节点的数据分布为NON-IID类型。同理,

当选择 20% 的占比时,即有 4 个客户端节点为 IID 类型而其他 16 个为 NON-IID 类型。然后,通过对比数据分布完全为 NON-IID 类型簇内节点的模型准确率,以验证本文方法的有效性。

### 6.3 实验结果

首先,对 10% IID 和 90% NON-IID 构成的数据集进行了测试,测试结果如图 5 所示。图 5a 和图 5b 分别表示每一轮全局迭代的损失函数和精度函数。训练准确率达到 97.50%,共耗时 2 125.75 s。

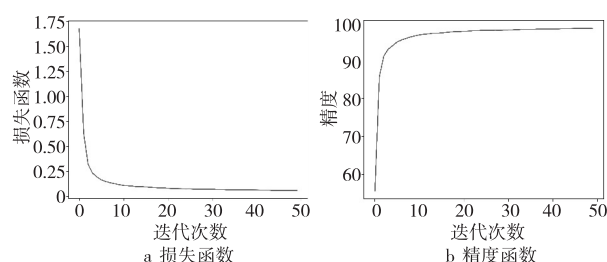


Figure 5 Loss function and precision

function when  $frac=0.1$

图 5  $frac=0.1$  时的损失函数和精度函数

图 6 对应 20% IID 和 80% NON-IID 组成的数据集的损失函数和精度函数,训练准确率达到 98.39%,执行耗时 2 587.55 s;图 7 对应完全 NON-IID 组成的数据集的损失函数和精度函数,训练准确率为 82.49%,执行耗时 1 637.41 s。

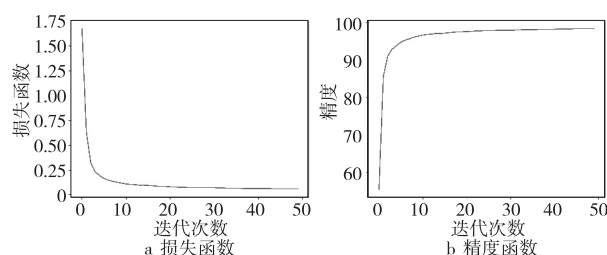


Figure 6 Loss function and precision

function when  $frac=0.2$

图 6  $frac=0.2$  时的损失函数和精度函数

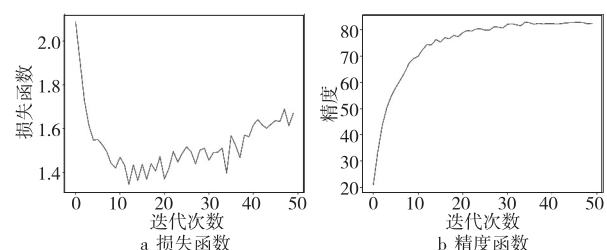


Figure 7 Full NON-IID loss function and precision function

图 7 Full NON-IID 损失函数和精度函数

实验结果显示,当簇内节点不包含数据质量高的节点时,模型的准确率较低;当簇内数据质量高的节点占比达到 10% 时,模型的准确率提高了

15%;当簇内数据质量高的节点占比达到 20% 时,模型的准确率提高了 15.9%。经过分析可知,本文提出的基于数据分布异步联邦学习方法有效,该方法通过将数据质量高的节点重新分发到数据质量低的簇中,提高了模型的准确率。

## 7 结束语

本文研究了物联网背景下,数据异质性导致联邦学习模型收敛速度慢和模型准确率低的问题,提出了基于数据分布异步联邦学习的方法,以解决整体训练耗时长和模型收敛差的问题,并通过构造不同类型的 NON-IID 数据集进行了测试。实验结果表明,本文方法具有优越性,可用于解决实际问题。

### 参考文献:

- [1] Zhu G X, Liu D Z, Du Y Q, et al. Toward an intelligent edge: Wireless communication meets machine learning[J]. IEEE Communications Magazine, 2020, 58(1): 19-25.
- [2] McMahan H B, Moore E, Ramage D. Communication-efficient learning of deep networks from decentralized data[C]//Proc of the 20th International Conference on Artificial Intelligence and Statistics, 2017: 1273-1282.
- [3] Nishio T, Yonetani R. Client selection for federated learning with heterogeneous resources in mobile edge[C]//Proc of 2019 IEEE International Conference on Communications, 2019: 1-7.
- [4] Konecny J, McMahan H B, Ramage D, et al. Federated optimization: Distributed machine learning for on-device intelligence[J]. arXiv: 1610. 02527v1, 2016.
- [5] Wahab O A, Mourad A, Otrok H, et al. Federated machine learning: Survey, multi-level classification, desirable criteria and future directions in communication and networking systems[J]. IEEE Communications Surveys and Tutorials, 2021, 23(2): 1342-1397.
- [6] Liu J C, Xu H L, Xu Y, et al. Communication-efficient asynchronous federated learning in resource-constrained edge computing[J]. Computer Networks, 2021, 199: 108429.
- [7] Zhang T, Ramakrishnan R, Livny M. BIRCH: An efficient data clustering method for very large databases[J]. ACM SIGMOD Record, 1996, 25(2): 103-114.
- [8] Qu Z, Duan R, Chen L X, et al. Context-aware online client selection for hierarchical federated learning[J]. IEEE Transactions on Parallel and Distributed Systems, 2022, 33(12): 4353-4367.
- [9] Liu J C, Xu H L, Wang L, et al. Adaptive asynchronous federated learning in resource-constrained edge computing[J]. IEEE Transactions on Mobile Computing, 2023, 22(2): 674-690.
- [10] Chen Y, Sun X Y, Jin Y C. Communication-efficient feder-



- ated 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.
- [11] Chai Z, Chen Y J, Anwar A, et al. FedAT: A high-performance and communication-efficient federated learning system with asynchronous tiers[C]//Proc of the International Conference for High Performance Computing, Networking, Storage and Analysis, 2021: 1-16.
- [12] Chen Y, Liao W X, Hua K, et al. Towards asynchronous federated learning for heterogeneous edge-powered Internet of Things[J]. Digital Communications and Networks, 2021, 7(3): 317-326.
- [13] Hu G, Teng Y L, Wang N, et al. Clustered data sharing for non-IID federated learning over wireless networks[J]. arXiv: 2302.10747v2, 2023.
- [14] Liu L M, Zhang J, Song S H, et al. Hierarchical federated learning with quantization: Convergence analysis and system design[J]. IEEE Transactions on Wireless Communications, 2023, 22(1): 2-18.
- [15] Tian P, Liao W X, Yu W, et al. WSCC: A weight similarity based client clustering approach for non-IID federated learning[J]. IEEE Internet of Things Journal, 2022, 9(20): 20243-20256.
- [16] Zhou C D, Tian H, Zhang H, et al. TEA-fed: Time-efficient asynchronous federated learning for edge computing[C]//Proc of the 18th ACM International Conference on Computing Frontiers, 2021: 30-37.
- [17] Liu S L, Yu G D, Chen X F, et al. Joint user association and resource allocation for wireless hierarchical federated learning with IID and non-IID data[J]. IEEE Transactions on Wireless Communications, 2022, 21(10): 7852-7866.
- [18] Liu Y, Zhao Y, Zhou G M, et al. FedPrune: Personalized and communication-efficient federated learning on non-IID data[J]. Neural Information Processing, 2021, 1516: 430-437.
- [19] Zhou Z H, Li Y N, Ren X B, et al. Towards efficient and stable  $K$ -asynchronous federated learning with unbounded stale gradients on non-IID data[J]. IEEE Transactions on Parallel and Distributed Systems, 2022, 33(12): 3291-3305.
- [20] Ma Q P, Xu Y, Xu H L, et al. FedSA: A semi-asynchronous federated learning mechanism in heterogeneous edge computing[J]. IEEE Journal on Selected Areas in Communications, 2021, 39(12): 3654-3672.
- [21] Harry Hsu T M, Qi H, Brown M. Measuring the effects of non-identical data distribution for federated visual classification[J]. arXiv: 1909.06335v1, 2019.

### 作者简介:



郭昌昊(1996-),男,河南安阳人,硕士生,研究方向为联邦学习。E-mail: 22302027@muc.edu.cn

GUO Chang-hao, born in 1996, MS candidate, his research interest includes federated learning.



唐湘云(1994-),女,湖南长沙人,博士,副教授,研究方向为人工智能安全。E-mail: xiangyunt@muc.edu.cn

TANG Xiang-yun, born in 1994, PhD, associate professor, her research interest includes artificial intelligence security.



翁彧(1980-),男,辽宁大连人,博士,教授,博士生导师,研究方向为云计算。E-mail: wengyu@muc.edu.cn

WENG Yu, born in 1980, PhD, professor, PhD supervisor, his research interest includes cloud computing.