计算机科学
COMPUTER SCIENCE

一种基于背景优化的高效联邦学习方案

郭桂娟, 田晖, 王田, 贾维嘉

引用本文

郭桂娟, 田晖, 王田, 贾维嘉.一种基于背景优化的高效联邦学习方案[J]. 计算机科学, 2022, 49(12): 40-45.

GUO Gui-juan, TIAN Hui, WANG Tian, JIA Wei-jia. Efficient Federated Learning Scheme Based on Background Optimization [J]. Computer Science, 2022, 49(12): 40-45.

计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.220600237

一种基于背景优化的高效联邦学习方案

郭桂娟[1]　田　晖[1]　王　田[2,3]　贾维嘉[2,3]
1 华侨大学计算机科学与技术学院　福建 厦门 361021
2 北京师范大学人工智能与未来网络研究院　广东 珠海 519000
3 北京师范大学-香港浸会大学联合国际学院人工智能与多模态数据处理广东省重点实验室　广东 珠海 519000
(guoguijuan727@163.com)

摘　要　联邦学习因其在客户端本地进行数据的训练,从而有效保证了数据的隐私性和安全性。对于联邦学习的研究虽然取得了很大的进展,但是,由于非独立同分布数据的存在以及数据量不平衡、数据类型不平衡等问题,客户端在利用本地数据进行训练时不可避免地存在精确度缺失、训练效率低下等问题。为了应对联邦学习背景环境的不同导致的联邦学习效率降低的问题,文中提出了一种基于背景优化的高效联邦学习方案,用于提高终端设备中本地模型的精确度,从而减小通信开销、提高整体模型的训练效率。具体来说,在不同的环境中根据精确度的差异性来选择第一设备和第二设备,将第一设备模型和全局模型的不相关性(下文统称为差异值)作为标准差异值;而第二设备是否上传本地模型则由第二设备和第一设备之间的差异值决定。实验结果表明,与传统的联邦学习相比,所提方案在普通联邦学习场景下的表现明显优于联邦平均算法,在 MINIST 数据集上,其精确度提高了约 7.5%;在 CIFAR-10 数据集上,其精确度提高了约 10%。

关键词:联邦学习;背景优化;设备分类;不相关性;差异值
中图法分类号　TP393.0

Efficient Federated Learning Scheme Based on Background Optimization

GUO Gui-juan[1],TIAN Hui[1],WANG Tian[2,3] and JIA Wei-jia[2,3]
1 College of Computer Science and Technology,Huaqiao University,Xiamen,Fujian 361021,China
2 Institute of Artificial Intelligence and Future Networks,Beijing Normal University,Zhuhai,Guangdong 519000,China
3 Guangdong Key Lab of AI and Multi-Modal Data Processing,BNU-HKBU United International College,Zhuhai,Guangdong 519000,China

Abstract　Federated learning can effectively ensure the privacy and security of data because it trains data locally on the client. The study of federal learning has made great progress. However,due to the existence of non-independent and identically distributed data,unbalanced data amount and data type,the client will inevitably have problems such as lack of accuracy and low training efficiency when using local data for training. In order to deal with the problem that the federal learning efficiency is reduced due to the difference of the federal learning background,this paper proposes an efficient federated learning scheme based on background optimization to improve the accuracy of the local model in the terminal device,so as to reduce the communication cost and improve the training efficiency of the whole model. Specifically,the first device and the second device are selected according to the difference in accuracy in different environments,and the irrelevance between the first device model and the global model (hereafter we collectively refer to as the difference value) is taken as the standard difference value. Whether the second device uploads the local model is determined by the value of the difference between the second device and the first device. Experimental results show that compared with the traditional federated learning,the proposed scheme performs better than the federated average algorithm in common federated learning scenarios,and improves the accuracy by about 7.5% in the MINIST data sets. In the CIFAR-10 data

到稿日期:2022-06-24　返修日期:2022-08-16
基金项目:国家科技重点研发计划(2022YFE0201400);国家自然科学基金(62172046);福建省自然科学基金杰出青年项目(2020J06023);广东省教育厅重点专项(2021ZDZX1063);珠海市产学研项目(ZH22017001210133PWC);广东省教育厅人工智能与多模态重点实验室项目(2020KSYS007);UIC 科研启动经费(R72021202)
This work was supported by the National Key R&D Plan of the Ministry of Science and Technology(2022YFE0201400),National Natural Science Foundation of China(62172046),Outstanding Youth Program of Fujian Natural Science Foundation(2020J06023),Key Special Projects of Guangdong Provincial Department of Education(2021ZDZX1063),Zhuhai Industry University Research Project(ZH22017001210133PWC),Artificial Intelligence and Multimodal Key Laboratory Project of Guangdong Provincial Department of Education(2020KSYS007) and UIC Scientific Research Startup Fund(R72021202).
通信作者:王田(cs_tianwang@163.com)

set，accuracy improves by about 10%.

Keywords Federated learning，Background optimization，Equipment classification，Irrelevance，Difference value

1 引言

随着云计算、边缘计算、人工智能以及联邦学习的发展，海量数据得到了充分的利用。而联邦学习的出现打破了数据孤岛问题，使参与方能在不共享数据的前提下联合建模，从而实现 AI 协作，如物联网、车联网中的应用[1-3]。但是，由于不同背景环境下移动终端设备间的非独立同分布数据、数据量不平衡以及数据类型不平衡等问题，不同背景下移动终端设备所训练的模型差异值较大[4]；此外，移动终端设备的数据质量和完整性会导致联邦学习整体训练的时长增加和通信延迟等问题[5-6]。在考虑数据对联邦学习训练效率的影响时，Zhao 等从非独立同分布角度出发，证明了当客户端数据存在高度倾斜的非独立同分布数据时，联邦学习的精确度明显降低，最高仅 55%[7]。为了缓和非独立同分布数据、数量类型不平衡引起的训练精度降低、通信成本增加等问题，Duan 等提出了一种不平衡数据的自平衡联邦学习方案，旨在通过建立一个边缘层来收集模型，以达到局部平衡的一种状态，从而提高联邦学习的训练效率[8]。该方案没有考虑到单个模型的优劣问题，而是采用贪婪算法对每个模型均进行边缘层聚合。上述研究都是从非独立同分布数据进行的，而本文则从数据类型着手，考虑到因移动终端设备的背景环境不同，移动终端设备训练的模型对联邦学习效率的影响。

2 相关工作

在联邦学习环境中，客户端存在的数据量大小不平衡、数据类不平衡、数据异构等问题，导致联邦学习效率低下、训练时间延长、精确度低等问题[9-10]。为了解决客户端数据导致的各种问题，国内外专家和学者进行了广泛的研究。Zhao 等从数据异构性出发，提出了一种共享数据子集的方法来平衡非独立同分布数据。实验结果表明，对于全局共享数据仅为 5%的 CIFAR-10 的数据集来说，这种方法可以提高 30%的准确率[7]。对于冗余数据，Sun 等提出了一种在线能量感知的动态调度策略，用于减小沟通成本[11]。在联邦学习中，客户端的数量是庞大的，联邦学习环境可能由数百万参与者组成[12]。因此，需要聚合的本地模型数量也是十分惊人的，如何高效训练模型并保证模型的健壮性便显得十分重要。对于大量的模型，最常用的方法就是模型压缩。模型压缩（稀疏化）后，更新的模型结构可以用更少的变量来刻画。从随机稀疏角度出发，Shi 等将训练算法与本地计算、梯度稀疏相结合，提出了柔性稀疏法：对参与者施加误差补偿，梯度稀疏允许参与方只上传小部分具有显著特征的梯度，从而减少每一轮的通信有效负荷[13]。上述研究主要从数据、算法和模型的角度进行优化，而本文更关注的是移动终端设备处于不同环境而导致的差异性问题。

由于移动终端设备所处的背景环境有所不同，因此每个环境背景下训练的模型与其他环境下的模型偏差较大。比如，在高原上，终端设备中关于狗的图像更多的是藏獒、牧德这些类型；而在城市里，终端设备中关于狗的图像更多的是柯基、二哈、贵宾等类型。对此，本文提出了一种基于背景优化的高效联邦学习方案，以减少移动终端设备与云端之间的通信数量，同时提高了联邦学习的整体训练效率。本文第 3 节介绍了问题描述；第 4 节详细阐述了本文方案、算法实现的实现细节；第 5 节评估了不同的实验方案；最后总结全文并展望未来。

3 问题描述

在联邦学习每次迭代训练的过程中，主要寻求一个最优化模型参数 w，使得每个终端设备在其本地数据集上能够最小化其损失函数。由于终端设备的异构性和非独立同分布特性，终端设备模型之间的差异以及终端设备模型和全局模型之间的差异非常大[14]，如果仅仅从全局模型的梯度着手进行处理是非常不理智的。因此，本文从另一个视角——背景环境来考虑。由于每个背景环境中数据类型可能存在某些特性，因此，本文考虑从每个环境中每轮挑选一个第一终端设备作为领队设备，该环境中的剩余设备则作为第二终端设备。那么，本文的关注点就转到第一终端设备模型和全局设备模型的差异值以及第一终端设备模型和第二终端设备模型的差异值上。

本文定义第一终端设备模型和全局终端设备模型的梯度向量差异值为标准差异值 σ，第一设备模型和第二设备模型的差异值设定为ϱ。对于给定的环境场景个数 C、每个场景的终端设备 N，假设 $f(w,x,y)$是客户端最终的目标损失函数，则本文寻求解决以下问题的方法：

$$\begin{aligned} &\min \sum_{i=1}^{C}\sum_{j=1}^{N}\varrho \\ &\text{s.t. } \min f(w,x,y) \end{aligned} \tag{1}$$

4 基于背景优化的高效联邦学习方案

4.1 本文方案

由于联邦学习参与者达到数百万，因此客户端上传模型的链路是比较拥挤的[15-16]，同时终端设备不可能只存在于某个特定的背景环境中，因此终端设备之间的数据异构性问题、数据不平衡性问题以及通信开销问题不可避免。但是，如果处于同一背景环境中，终端设备数据特性相对来说相似性极高[17-18]。鉴于不同背景环境中，各个终端设备模型表现不一样，本文提出了一种基于背景优化的联邦学习方案（An Efficient Federated Learning Scheme Based on Background Optimization，BOFL）。BOFL 方案的具体实现步骤如下：

（1）联邦服务器将初始全局模型 w_0 下发到各个边缘服务器中。

（2）在每个背景环境中，通过训练精确度对比挑选出训练精确度最高的终端设备作为第一终端设备，其余设备作为第二终端设备。

（3）边缘服务器将全局模型下发至第一终端设备，第一终

端设备上的模型与全局模型根据汉明距离得出标准差异值 σ，将其作为第一终端设备模型和第二终端设备模型的标准值。

(4)第二终端设备模型和第一终端设备模型根据汉明距离得出设备间的差异值ϱ。假设扩增的稳健外围百分比用 φ 表示，那么称所有满足 $\varrho<\varphi\sigma$ 的设备为 A 类设备，A 类设备可上传到边缘端进行聚合；假设扩增的摆动外围百分比用 ω 表示，那么称所有满足 $\varphi\sigma<\varrho<\omega\sigma$ 的设备为 B 类设备，B 类设备有 50%的概率能上传到边缘端进行聚合；称所有满足 $\omega\sigma<\varrho$ 的设备为 C 类设备，C 类设备在本轮次中被拒绝上传到边缘端进行聚合。

(5)所有的边缘端服务器对上传的模型进行聚合，称聚合后的模型为边缘模型。边缘模型同时被上传到联邦服务器上并下发给 C 类设备进行背景调整。

(6)联邦服务器对上传的边缘模型进行聚合，更新全局模型。

重复上述步骤，直到满足设定的阈值或者条件为止。

4.2 具体实现

由于每个环境中的设备模型有所差异，因此本文从设备的地理位置出发进行考虑，即每一个边缘服务器附近的设备为同一环境内的边缘设备；其余环境的划分也按照这种方法。对于处于两个环境中间的设备，则根据该设备与两个环境中的边缘服务器的最近距离进行划分。在进行全局模型的更新或者终端设备之间的比较时，我们考虑到模型参数的梯度向量反映的是模型的趋势。如文献[14]直接利用本地更新中参数正负值个数与全局更新之间的参数的相同符号个数来确定本地更新的重要性，即模型更新的参数符号代表的是模型参数在各个维度上的变化程度。因此，为了简化计算，本文先对模型参数进行二值化处理，二值化后参数符号相同的分别归为一类数值；然后由二值化后的全局模型参数和第一设备模型参数值不同的个数即汉明距离作为标准差异值 σ；根据二值化后的第一设备模型和第二设备模型间的汉明距离可以得到设备间的差异值 ρ。假设在第 t 轮的训练中，某个背景环境中共有 N 个客户端，且第一终端设备 C_{top} 的精确度为 $\varepsilon_{C_{\text{top}}}$，则必须满足式(2)：

$$\varepsilon_{C_{\text{top}}} > \forall_{\varepsilon_{C_i}} \quad \text{s.t. } i\neq\text{top}, i\in[1,n] \tag{2}$$

由于模型的梯度向量反映的是模型的趋势，并且模型参数大致可分为正负两个方向。为了简化计算量，本文先对所有模型参数进行二值化处理，把所有参数归为 0,1 这两个值。在信息论中，汉明距离原本用于数据传输差错控制编码，表示的是两个字对应位不同的数量。但当把两个字符看成向量空间中的元素时，汉明距离也就是一个 n 维的超立方体上两定点间的曼哈顿距离。因此，本文通过直接比较两个模型参数每位上不同值的个数来获得汉明距离，也就是说，本文通过汉明距离获得两个模型参数中对应位的不同数值的个数，即本文提到的模型差异值。由于计算汉明距离之前已经对模型参数进行了二值化处理，因此两个模型之间不同参数的个数反映的是两个模型之间的非相似程度，也就是差异程度。

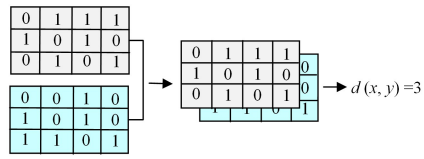

图 1 汉明距离

Fig. 1 Hamming distance

一般地，汉明距离用 $d(x,y)$ 表示，在本文中为了简化和区别，将第一终端设备模型和全局模型之间的差异值用 σ 表示，以 σ 作为标准差异值。第二终端设备模型和第一终端设备之间的差异值用ϱ表示。假设在第 t 轮的训练中，某个环境中共有 N 个客户端，第一终端设备 C_{top} 与其他第二终端设备 $C_i(i=1,2,\cdots,N)$ 的差异值若满足式(3)，则称其为 A 类设备，A 类设备在本轮可优先上传到边缘端进行聚合；为了实现环境变化的灵活性，本文设定了一个扩增中心外围百分比 φ，用于动态调整上传的比例。

$$\varrho_i < \varphi\sigma, i=1,2,\cdots,N \tag{3}$$

针对标准差异扩增后的中心外围，本文以此为基础又往外以摆动外围百分比 ω 划了一圈，在这个范围内的设备被称为 B 类设备。B 类设备的上传具有随机性，即每个设备都具有 50%的上传概率，B 类设备应满足式(4)：

$$\varphi\sigma<\varrho_i<\omega\sigma, i=1,2,\cdots,N \tag{4}$$

该环境中其他剩余的设备被称为 C 类设备，C 类设备在本次的训练中不具备上传的资格。为了减少第二设备一直占用资源或重复无用功，本文设置了一个阈值 δ，用于停止一直被拒绝上传的终端设备。当终端设备的被拒次数达到 δ 时，我们认为该设备对该联邦学习训练没有效果，即把该设备从联邦学习环境中删除，以减小额外的开销。实验结果证明，当 $\delta=8$ 时 BOFL 的表现具有鲁棒性。

4.3 模型差异值

对于异构性较大的设备，其模型之间的差异程度相对来说比较明显。文献[19]通过模型参数的余弦相似度来排除不相关的更新，相关的研究还包括文献[20]。虽然直观上看模型的相似度大致代表了维度的相似情况，但是余弦相似度是从空间矢量的角度反映方向(维数)的差异，而不是距离。考虑到设备间模型参数代表模型改变的趋势，本文从汉明距离进行研究。假设在同一背景环境中，第一终端设备的本地模型为 w_1，其他终端设备为 $w_i, i\in N$。那么根据汉明距离公式可以计算出第一终端设备和第二终端设备之间的汉明距离，具体如式(5)所示：

$$d(w_1,w_i)=\sum_{k=1}^{N} I(u_k=\bar{u}_k) \tag{5}$$

其中，N 代表参数总数，u_k 代表 w_1 在 k 位置的参数值，$\bar{u}_k$ 代表 w_i 在 k 位置的参数值。通过第一终端设备和全局模型的汉明距离可得到标准差异值 σ，σ 用作设备分类的判断中心，具体的设备分类方法已在 4.2 节详细介绍。

算法 1 描述了模型参数二值化后的汉明距离，即模型差异值求解算法。在求标准差异值 σ 和差异值ϱ时，由于均由模型参数着手，因此在求解汉明距离时，输入的是两个模型参数，输出的是汉明距离。

算法 1 模型差异值

输入：第一终端设备的本地模型 w_1，第二终端设备 i 的本地模型 w_i

```
1. binarizer=preprocessing.Binarizer()
2. x=binarizer.transform(w1) #对第一终端设备模型进行离散二值化
3. y=binarizer.transform(wi) #对第二终端设备模型进行离散二值化
4. for w1',w2' in zip (w1.parameters(),wi.parameters()):
5.   if (w1'!=w2')
6.       Hanmin++;
7. return Hanmin.
```

4.4 基于设备分类结果聚合

在 BOFL 方案中，其目标在于求解最小的ϱ，以提高联邦学习效率。鉴于背景环境的不同，假设共有 10 个边缘端，每个边缘端有若干个终端设备，假设每个环境中的第一终端设备模型的精确度用 acc^1 表示，其他第二终端设备模型的精确度用 acc_i 表示，其中 $i\in 20$。那么第一终端设备模型的精确度可以表示为：

$$acc^1=\max\{acc_1,acc_2,\cdots,acc_{20}\} \tag{6}$$

用 CID 作为参与训练的终端设备号，w_0 为初始化的全局模型，ES 为边缘服务器，δ 为第二终端设备被拒的阈值，φ 为标准差异扩增的百分比，ω 为标准差异扩增的另一百分比，A max 为第一终端设备的训练精确度，w_{max} 为第一终端设备的模型参数，σ 为标准差异值，p 用于累计参与边缘聚合的终端设备数量。当设备模型满足 $d(W_j,W_{max})<\sigma\varphi$ 时，认为该设备为 A 类设备，直接将其上传到云端进行聚合；当模型满足 $\sigma\varphi<d(W_j,W_{max})<\sigma\omega$ 时，认为该设备是 B 类设备，此时随机生成一个 0-1 数，若生成 1，则将该设备本地模型上传到云端进行聚合；当模型均不满足上述条件时，丢弃该本地模型。算法 2 详细介绍了 BOFL 方案。

算法 2 BOFL 算法

```
1. Procedure
2. Input:Client ID CID={cid1,cid2,…,cidn}
3. Initialize:the global model w0,edge servicer ES={es1,es2,…,esk},δ,φ,ω
4. For each round i=1,2,3,… do
5.    For each es∈ES do
6.      For each j∈CID do
7.        A max=acc1 #通过精确度来获取第一终端设备
8.      σ=d(Wmax,W0)
9.      For each j∈CID do
10.       If Cj<δ #该设备还没被删除
11.         Local update wj
            If d(Wj,Wmax)<σφ #第二终端|第一终端差异值和标准差异值比较
12.           Upload wj to cloud wes
13.           p++
14.         Else if d(Wj,Wmax)<σω
15.         Rand upload wj to cloud wes
16.         p++
17.       Else:
18.         Client list update Cj
19.     Else:
20.     Client list update Cj #记录该终端设备没有上传到边缘端进行聚合的次数
21.     W'es=(1/p)ΣWes
22.     For each j∈Cj do
23.       Local update Wj
24. Wi←Wi-1+(1/|ES|)ΣW'es
```

5 仿真结果与数据分析

5.1 数据集描述和实验准备

本文的模拟实验共有 10 个环境，每个环境中有一个边缘服务器；整个实验设置 100 个终端设备。为了模拟设备位置与边缘服务器的关系，实验中为每一个终端设备分配有 10 个随机数的时延列表，用于模拟该设备分别到 10 个边缘服务器的响应时延。我们认为，当时延达到最小时的位置 i 代表该设备离第 i 个边缘服务器最近，即该设备在第 i 个环境中。实验中，首先通过时延列表将所有设备归类到每一个环境中，然后使用 BOFL 方案进行性能评估。本文评估了 BOFL 模型在 MNIST 数据集和 CIFAR-10 数据集上联邦学习的训练效果。本文基于卷积神经网络模型（Convolutional Neural Network，CNN）训练模型。

MNIST 数据集是一个手写体数字的图片数据集，此数据集一共统计了来自 250 个不同人的手写数字图片。MNIST 的训练集一共包含了 60000 张图像和标签，而测试集一共包含了 10000 张图像和标签。MNIST 数据集是分类任务中最简单、最常用的数据集，每个样本都是一张 28 * 28 像素的灰度手写数字图片。

CIFAR-10 数据集是一个用于物体识别的计算机视觉数据集，它包含了 60000 张 32 * 32 的 RGB 彩色图片，总共有 10 个分类。其中，50000 张用作训练集，10000 张用作测试集。每个样本都是一张 32 * 32 像素的 3 通道彩色图片。

CNN 是一类包含卷积计算且具有深度结构的前馈神经网络[21]，CNN 由输入层、卷积成、激活函数、池化层、全连接层和输出层组成。本文使用的卷积核大小为 5×5，激活函数为 ReLU。

本文的实验均在 TensorFlow 框架上实现。本文在两种数据集上进行了测试，将训练样本平均分配到 100 个终端设备。本文对终端设备的停止上传阈值进行了测试调整，同时对扩增中心外围百分比、扩增摆动外围百分比进行了测试调整。

5.2 相关阈值设定

为了寻找到最佳的阈值条件，本文首先以 $\varphi=0.2$，$\varphi=0.5$ 为定量对 δ 进行变化测试。由于终端设备数据的异构特性，可能此次训练表现不佳，但是下次可能会因为新的数据而表现良好，因此 δ 取值不能太小，本文从 $\delta=5$ 开始进行对比实验。此外，如果 δ 取值太大，则可能会浪费资源，因为该设备做的训练对联邦环境没有任何贡献。因此，本文分别对 δ 取 5，6，8，10 进行实验。在 δ 取不同值时，训练的精确度表现如图 2 所示。从图 2 中可以看出，当 $\delta=8$ 时对其他变量而言图像上下波动的情况相对稳定。因此，本文设定停止上传的阈值$\delta=8$，此后其他实验均在 $\delta=8$ 的基础上进行。

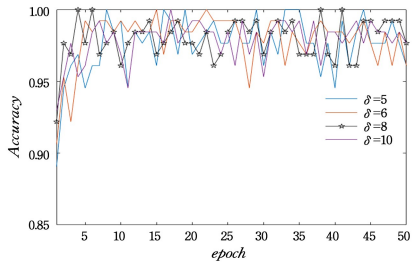

图 2 δ 取不同值时对应的精确度

Fig. 2 Accuracy of different values of δ

为了找到最优的 φ，本文分别取 $\varphi=0.1, 0.2, 0.3, 0.4$ 进行实验，0.5 以上的值，本文不再作比对实验，因为范围扩增太大，标准差异值将失去本身意义。φ 不同取值的实验结果如图 3 所示。

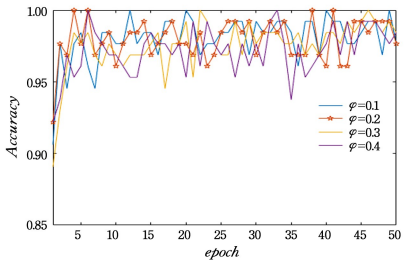

图 3 φ 取不同值时对应的精确度

Fig. 3 Accuracy of different values of φ

由图 3 可以看出，各线段差别不大，对比 φ 取不同值时的损失和精确度，当 $\varphi=0.2$ 时图像波动情况相对于其他值而言较为缓和，因此本文取 $\varphi=0.2$，此后其他实验均在 $\varphi=0.2$ 的条件下进行。

除了 A 类设备，考虑到某些终端设备可能因数据的异构性而此次表现不佳，但是这些设备下轮可能表现良好，即跳跃性较大。为了更加灵活地适应各种情况，本文在 $\varphi\sigma$ 中心处向外多划 $\omega\sigma$ 作为扩增的摆动外围。处于扩增摆动外围的终端设备，即满足式(4)的终端设备，有 50% 的随机上传概率可上传到边缘端进行聚合。为了获得最佳的 ω，本文分别取 $\omega=0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1$ 进行实验，实验结果如图 4 所示。

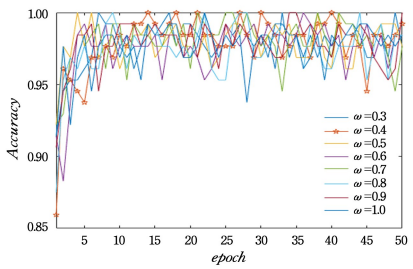

图 4 ω 取不同值时对应的精确度

Fig. 4 Accuracy of different values of ω

由图 4 可以看出，当 $\omega=0.4$ 时，相比其他取值的图像，该图像上下波动的情况更为缓和，因此本文取 $\omega=0.4$，此后的其他对比实验均在 $\omega=0.4$ 的情况下进行。

5.3 结果与分析

本节评估了 BOFL 在不同数据集上的性能结果。在本实验中，本文模拟共有 10 个不同的环境，每个环境中有若干个终端设备参与训练。其他参数值 φ，ω 分别取 0.2 和 0.4。

本文将学习率设为 $\lambda=0.001$，训练轮次为 50 轮。本文测试了传统 FedAvg 和 BOFL 在 CNN 模型上的表现。如图 5、图 6 所示，BOFL 方案在训练过程中表现良好。在实验进行过程中 BOFL 很快达到了收敛状态，并且精确度比传统的 FedAvg 更高。此外，BOFL 训练的损失值明显比传统 FedAvg 更小。

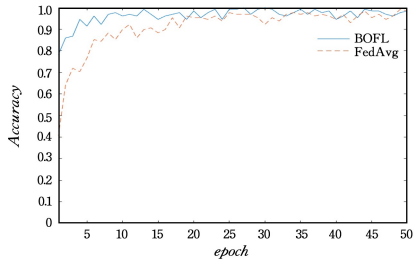

图 5 不同方法在 MINIST 上的精确度

Fig. 5 Accuracy of different methods at MINIST

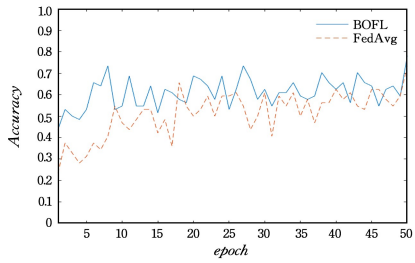

图 6 不同方法在 CIFAR-10 上的精确度

Fig. 6 Accuracy of different methods at CIFAR-10

当在 CIFAR-10 数据集上进行训练时，相比 MINIST 数据集而言，两种方法的精确度都不高，损失都较大，但是 BOFL 方法相对而言还是比传统的联邦学习效率更好，精确度整体上比传统联邦学习高 10%左右。

结束语 本文提出了一种基于背景优化的高效联邦学习方案 BOFL，旨在优化不同背景下联邦学习的通信效率及其精确度。BOFL 从背景不同数据异构导致模型有所偏差着手，在不同背景环境中挑选一个第一终端设备作为领队设备，其余设备只需要与本环境中的第一终端设备进行对比即可。在对比过程中，考虑到模型的趋势问题，本文从汉明距离着手计算终端设备之间的差异值，然后根据差异值来决定本地模型是否上传到边缘端进行聚合。实验结果表明，BOFL 在实验开始时就能达到很高的精确度，而 FedAvg 在跑到 9 轮左右精确度才达到 90%以上。总体上，对于 MINIST 数据集，BOFL 将精确度提高了约 7.5%；对于 CIFAR-10 数据集，精确度提高了约 10%。

在未来的研究中，我们将考虑联邦学习环境中的更多可变因素，如设备的对比的随机性、设备的不稳定性、可靠性等，以综合提高联邦学习的系统性能。

参 考 文 献

[1] WANG T, CAO Z, WANG S, et al. Privacy-enhanced data collection based on deep learning for Internet of vehicles[J]. IEEE

Transactions on Industrial Informatics, 2019, 16(10): 6663-6672.

[2] YANG Q, LIU Y, CHEN T, et al. Federated machine learning: Concept and applications[J]. ACM Transactions on Intelligent Systems and Technology(TIST), 2019, 10(2): 1-19.

[3] ZHANG Y L, LIANG Y Z, YIN M J, et al. A Review of Offloading Schemes in moving Edge Computing[J]. Chinese Journal of Computers, 2021, 44(12): 2406-2430.

[4] SATTLER F, WIEDEMANN S, MÜLLER K R, et al. Robust and communication-efficient federated learning from non-iid data [J]. IEEE Transactions on Neural Networks and Learning Systems, 2019, 31(9): 3400-3413.

[5] WANG T, BHUIYAN M Z A, WANG G, et al. Preserving balance between privacy and data integrity in edge-assisted Internet of Things[J]. IEEE Internet of Things Journal, 2019, 7(4): 2679-2689.

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

[7] ZHAO Y, LI M, LAI L, et al. Federated learning with non-iid data[J]. arXiv: 1806.00582, 2018.

[8] DUAN M, LIU D, CHEN X, et al. Self-balancing federated learning with global imbalanced data in mobile systems[J]. IEEE Transactions on Parallel and DistributedSystems, 2020, 32(1): 59-71.

[9] HUANG Y, CHU L, ZHOU Z, et al. Personalized cross-silo federated learning on non-iid data[C]// Proceedings of the AAAI Conference on Artificial Intelligence. 2021: 7865-7873.

[10] ZHANG L, LUO Y, BAI Y, et al. Federated Learning for Non-IID Data via Unified Feature Learning and Optimization Objective Alignment[C]// Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021: 4420-4428.

[11] SUN Y, ZHOU S, GÜNDÜZ D. Energy-aware analog aggregation for federated learning with redundant data[C]// ICC 2020-2020 IEEE International Conference on Communications(ICC). IEEE, 2020: 1-7.

[12] BONAWITZ K, EICHNER H, GRIESKAMP W, et al. Towards federated learning at scale: System design[J]. arXiv: 1902.01046, 2019.

[13] SHI D, LI L, CHEN R, et al. Towards Energy Efficient Federated Learning over 5G+ Mobile Devices [J]. IEEE Signal Processing Magazine, 2020, 37(3): 50-60.

[14] WANG L, WANG W, LI B. CMFL: Mitigating communication overhead for federated learning[C]// Proceedings of the IEEE 39th International Conference on Distributed Computing Systems. Texas, USA, 2019: 954-964.

[15] WANG T, WANG P, CAI S, et al. Mobile edge-enabled trust evaluation for the Internet of Things[J]. Information Fusion, 2021, 75(4): 90-100.

[16] SUN B, LIU Y, WANG T, et al. Review offederated learning Efficiency Optimization in Mobile Edge Networks [EB/OL]. http://kns.cnki.net/kcms/detail/11.1777.TP.20211104.2019.024.html. 2022-4-16.

[17] ZHANG C, XIE Y, BAI H, et al. A survey on federated learning [J]. Knowledge-Based Systems, 2021, 216(15): 1-11.

[18] KAIROUZ P, MCMAHAN H B, AVENT B, et al. Advances and open problems in federated learning[J]. arXiv: 1912.04977, 2019.

[19] WANG T, LIU Y, ZHENG X, et al. Edge-Based Communication Optimization for Distributed Federated Learning [J]. IEEE Transactions on Network Science and Engineering, 2022, 9(4): 2015-2024.

[20] LIU Y, WANG T, PENG S L, et al. Edge-based model cleaning and device clustering in federated learning[J]. Chinese Journal of Computers, 2021, 44(12): 2517-2530.

[21] ABADI M, BARHAM P, CHEN J, et al. Tensorflow: A system for large-scale machine learning[C]// 12th USENIX Symposium on Operating Systems Design and Implementation(OSDI 16). 2016: 265-283.

GUO Gui-juan, born in 1994, postgraduate. Her main research interests include edge intelligence and federated learning.

WANG Tian, born in 1982, Ph.D, professor, Ph.D supervisor, is a senior member of China Computer Federation. His main research interests include edge computing and Internet of things.

(责任编辑:喻藜)