

# 在不平衡数据中进行高效通信的联邦学习

舒志鸿<sup>1</sup>, 沈苏彬<sup>2</sup>

(1. 南京邮电大学 计算机学院, 江苏 南京 210046;

2. 南京邮电大学 通信与网络技术国家工程研究中心, 江苏 南京 210046)

**摘 要:**联邦学习(FL)是一种分布式的机器学习方法,它通过中心服务器汇总各个移动终端在本地训练的机器学习模型,使得多个参与方能够协作进行高效率的机器学习。同时,FL不需要将终端的私人数据发送至中心服务器,从而保护了数据隐私。但是与普通的训练数据集不同,终端系统中的数据分布不平衡,这将导致FL的通信效率下降。针对该问题,提出了一种基于数据分布加权聚合的FL算法。通过计算参与方的本地数据集与平衡数据集之间的海林格距离对本地数据集的平衡程度进行了量化,并据此调整了参与方在聚合时的权重,以减少算法收敛或达到目标准确率所需的通信回合。提出的算法利用公开数据集进行了仿真实验。实验结果表明,其与最新的算法联邦平均相比,通信成本降低了14.6%以上,有效提升了数据不平衡时FL的通信效率。

**关键词:**联邦学习;机器学习;不平衡数据;海林格距离;聚合

中图分类号:TP181

文献标识码:A

文章编号:1673-629X(2021)12-0033-06

doi:10.3969/j.issn.1673-629X.2021.12.006

## Communication-efficient Federated Learning from Imbalanced Data

SHU Zhi-hong<sup>1</sup>, SHEN Su-bin<sup>2</sup>

(1. School of Computer Science, Nanjing University of Posts and Telecommunications, Nanjing 210046, China;

2. National Engineering Research Center on Communication and Networking, Nanjing University of Posts and Telecommunications, Nanjing 210046, China)

**Abstract:** Federated learning (FL) is a distributed machine learning method that aggregates machine learning models trained locally by various mobile terminals through a central server, so that multiple participants can collaborate in high-efficiency machine learning. At the same time, FL does not need to send the private data of the terminal to the central server, thereby protecting data privacy. But different from the ordinary training data set, the data distribution in the terminal system is not balanced, which will lead to the decrease of the communication efficiency of FL. To solve this problem, a FL algorithm based on weighted aggregation of data distribution is proposed. The balance of the local data set is quantified by calculating the Hellinger distance between the local data set of the participants and the balanced data set, and the weight of the participants during aggregation is adjusted accordingly to reduce the algorithm convergence or achieve the goal the communication round required for accuracy. The proposed algorithm uses public data sets to conduct simulation experiments. The experimental results show that compared with the latest algorithm Federated Averaging, the communication cost is reduced by more than 14.6%, which effectively improves the communication efficiency of FL when the data is imbalanced.

**Key words:** federated learning; machine learning; imbalanced data; Hellinger distance; aggregation

## 0 引言

随着互联网以及移动终端技术的持续发展,智能手机等移动设备已经成为人们生活中不可或缺的组成部分,并在与用户交互的过程中产生了大量的数据。使用机器学习算法对这些数据进行深度分析可以帮助服务商充分了解用户,为用户提供更好的服务。传统

的机器学习策略要求移动设备将数据上传到云服务器或数据中心进行处理<sup>[1]</sup>,然而,数据隐私和安全性问题越来越受到社会及公众的重视,例如欧盟在2018年推出的《通用数据保护条例》<sup>[2]</sup>中明确规定数据的收集和存储必须在消费者同意的条件下进行。在这种趋势下,传统方法或将不再适用。

收稿日期:2021-01-15

修回日期:2021-05-17

基金项目:中国通信标准化国际标准制订项目(2018外122)

作者简介:舒志鸿(1995-),男,硕士研究生,研究方向为机器学习;沈苏彬,博导,研究员,CCF高级会员(E200005482S),研究方向为物联网及其应用、未来网络及其应用。

为了解决该问题,Google<sup>[3]</sup>提出了一种分布式的机器学习方法,称为联邦学习(FL)。在 FL 中,客户端在本地训练自己的局部模型,然后仅仅将局部模型的参数发送到服务器进行聚合,以更新全局模型。FL 重复上述过程直到全局模型收敛或达到所需的训练准确率为止。区别于典型的分布式机器学习,FL 的一个关键特征是数据的异质性,即:

(1)非独立同分布(Non-IID):移动设备上的本地数据集取决于用户的使用情况,因此任意客户端的本地数据分布都无法代表全局数据的分布。

(2)不平衡:用户使用服务或应用程序的频率不同,造成客户端上的数据量存在差异。

由于模型参数的庞大以及移动设备有限的通信带宽,通信成本成为制约 FL 发展的主要因素之一<sup>[4]</sup>。为了应对这一挑战,McMahan 等人<sup>[5]</sup>提出了目前广为使用的 FL 算法联邦平均(federated averaging, FedAvg),他们以平均的方式聚合各个参与方的局部模型以更新全局模型,并通过增加聚合期间局部模型的训练次数,减少通信开销。这项研究考虑了客户端上数据量的差异,但却假设全局数据平衡分布,即在所有客户端收集到的数据中各个类别的样本数量分布是均衡的。

然而在许多实际应用中都存在全局数据不平衡的情况,例如欺诈检测<sup>[6]</sup>、图像识别<sup>[7]</sup>等。而 Duan 等人<sup>[8]</sup>通过进一步的研究表明,FedAvg 在全局数据不平衡时的表现不佳。

该文主要研究全局数据不平衡时 FL 通信效率的优化,在 FedAvg 的基础上提出了一种基于数据分布加权聚合的 FL 算法(federated learning with data distribution weighted aggregation, FLDWA)。FLDWA 通过客户端的本地数据分布与平衡分布之间的海林格距离衡量本地数据的平衡程度,并将相关信息发送至 FL 服务器。然后,FL 服务器据此为客户端执行加权聚合,从而在全局数据不平衡时更加高效地提取局部模型的信息。

在公开数据集 MNIST<sup>[9]</sup>上使用多种不同的设置进行了仿真实验,以验证所提出方法的正确性和有效性,该数据集已被广泛用于 FL 的相关研究中。实验结果表明,在不平衡的 MNIST 上,FLDWA 可以有效提升 FL 的通信效率。

## 1 相关工作

优化 FL 在全局数据不平衡时的通信效率,一个直接的想法是解决全局数据不平衡问题。在本节中,从不平衡数据上的机器学习以及联邦学习两个角度介绍和分析相关的一些研究。

### (1)不平衡数据上的机器学习。

数据不平衡主要是指数据集中某些类的样本数量远大于另一些类。一般将样本数量非常多的类称为多数类,样本数量较少的则称为少数类<sup>[10]</sup>。该问题可以通过修改训练数据或调整学习策略加以解决<sup>[11]</sup>。前者旨在删除部分多数类(欠采样)或生成部分少数类(过采样)样本使数据分布重新达到平衡状态,文献[12]提出了一种基于聚类的欠采样方法,通过 K 均值聚类算法对多数类进行了聚类,并用聚类中心代替同簇数据。Chawla 等人<sup>[13]</sup>提出了合成少数类过采样技术(SMOTE),通过对少数类进行分析并结合线性插值的方法生成新的少数类样本。调整学习策略的方法则致力于对损失函数进行修改从而削弱算法对少数类的偏见,最受欢迎的方案是代价敏感学习<sup>[14]</sup>。该方法增加了少数类样本的误分类成本从而提高了对少数类的关注。文献[15]中使用了再缩放技术对代价敏感学习进行改善,使其能够适用于多分类任务。

然而,上述方法不适用于 FL。FL 数据仅能够被其所有者所访问,导致采样的方法难以实现。并且由于无从获取整体数据的分布情况,调整学习策略的解决方案仅能在局部使用,这将导致各个用户的局部模型不一致。

### (2)联邦学习。

通信效率的优化一直是 FL 的主要研究方向之一。McMahan 等人<sup>[16]</sup>提出了结构化更新和粗略更新,通过稀疏化和编码技术实现了传输参数的缩小。文献[17]对模型参数进行了有损压缩,并且通过在训练过程中删除固定数量激活单元的方法进一步简化了参数的复杂度,实现了通信开销的优化,从而扩展了文献[16]中的研究。虽然压缩模型参数的方法拥有极强的泛用性,但会导致准确性的牺牲。

Chen 等人<sup>[18]</sup>将神经网络分为深层和浅层,并以不同的频率更新它们的参数,同时根据参数的时效性调整了聚合策略,实现了通信成本的降低。Liu 等人<sup>[19]</sup>设计了一种具有客户端-边缘-云的分层 FL 系统,通过各层之间的协作减少了资源的消耗。Yao 等人<sup>[20]</sup>将最大均值差异(MMD)引入损失函数中,通过最小化局部模型和全局模型的 MMD 损失,减少了算法所需的通信回合。然而,上述工作没有考虑到全局数据不平衡对 FL 的影响。

目前只有少数研究关注全局数据不平衡问题。Duan 等人<sup>[8]</sup>通过数据扩充减轻单个客户端的不平衡程度,并且在服务器与客户端之间设置中介,根据客户端数据分布的 Kullback-Leibler 距离重新安排它们的协作训练。该方法引入了不可忽视的存储和时间开销,这可能会成为应用的限制。

与上述的研究相比,笔者更加关注方法的普适性,致力于在尽量避免额外的成本或牺牲的情况下,提升全局数据不平衡时 FL 的通信效率。

## 2 传统联邦学习方法

通常,FL 包含两个主要的实体,即客户端和服务端。客户端  $k(k=1,2,\dots,K)$  持有一个私有的本地数据集  $D_k$ ,并在  $D_k$  上最小化损失函数  $f_k(w)$  以训练局部模型。令  $D = \cup_{k \in K} D_k$  表示全局数据集,  $|D|$  和  $|D_k|$  分别表示全局数据集以及客户端  $k$  上的本地数据集的数据量。则 FL 的优化目标定义为:

$$\min_{w \in \mathbb{R}^d} F(w) \quad (1)$$

其中,  $F(w)$  为全局模型的损失函数,定义如下:

$$F(w) = \sum_{k=1}^K \frac{|D_k|}{|D|} f_k(w) \quad (2)$$

在 FL 中,服务器收集所有局部模型的模型参数,并通过将它们聚合以更新全局模型。所以,FL 的性能在很大程度上取决于聚合策略。目前广泛使用的聚合策略是由 FedAvg 算法<sup>[5]</sup>中提出的平均聚合,其具体实现由式(3)给出:

$$w_t = \sum_{k=1}^K \frac{|D_k|}{|D|} w_t^k \quad (3)$$

其中,  $w_t^k$  为客户端  $k$  在第  $t$  个通信回合时的局部模型参数,  $w_t$  则表示此时的全局模型参数。

## 3 数据分布加权聚合的联邦学习方法

在对不平衡数据集进行分类操作时,算法为了提高分类精度,会倾向于将多数类分类正确,从而对少数类产生偏见。所以对于同一个机器学习任务,通常使用分布更加平衡的训练数据得到的模型质量也会更高。

由于数据隔离的原因,FL 无法直接使用全局数据进行训练,而是通过聚合局部模型达到学习的目的。当全局数据不平衡时,由于 FL 的本地数据集具有非独立同分布的特点,各个客户端的本地数据在分布平衡程度上出现差异,导致训练出来的局部模型质量也会有所差异。由式(3)可知,FedAvg 算法中的聚合策略仅根据客户端上的数据量确定相应局部模型的权重。这可能并不合理,因为数据分布更为平衡的客户端通常会训练出质量更高的局部模型,应该在聚合时具有更高的权重。

本节提出的数据分布加权聚合的联邦学习 (FLDWA) 方法在 FedAvg 的基础上加入了对数据分布的考虑。量化了每个 FL 本地数据集的平衡程度,据此调整了聚合策略,从而更加有效地提取局部模型的信息。此外,只要本地数据集不发生变化,其平衡程

度也不会改变。所以在算法的执行过程中,客户端只需首次与服务器通信时将相关信息上传即可。

(1) 本地数据集平衡程度的量化。

使用海林格距离对本地数据集的平衡程度进行量化。在概率论和统计理论中,海林格距离被用来度量两个概率分布的相似程度<sup>[21]</sup>。设两个离散概率分布分别为  $U = (u_1, u_2, \dots, u_n)$  和  $V = (v_1, v_2, \dots, v_n)$ , 则它们之间的海林格距离定义为:

$$H(U, V) = \frac{1}{\sqrt{2}} \sqrt{\sum_{i=1}^n (\sqrt{u_i} - \sqrt{v_i})^2} \quad (4)$$

于是,本地数据集的平衡程度可以通过计算其与平衡数据集的海林格距离来刻画。平衡数据集是指各类别样本数量分布均衡的数据集,但目前还没有研究指出各类别样本需要满足何种数量关系才能定义为均衡,该文不解决此问题。由此,定义了一个基准数据集  $D_b$  作为平衡数据集的替代。在基准数据集中,各个类别的样本数量严格遵循  $1:1:\dots:1$  的规律。

本地数据集与基准数据集的海林格距离越小,表示两者的相似度越高,也即该数据集的平衡程度越高。值得注意的是,海林格距离与相似度是负相关的,为了便于后续的计算以及更为直观地表示平衡程度,需要对其进行转换。考虑到海林格距离满足柯西-施瓦兹不等式,所以其具有如下性质:

$$0 \leq H(U, V) \leq 1 \quad (5)$$

最终,使用式(6)所示的  $B_k$  表征本地数据集  $D_k$  的平衡情况,称为平衡度。它是通过将式(7)计算出的本地数据集与基准数据集之间的海林格距离关于其值域进行翻转后得到的:

$$B_k = 1 - H_k \quad (6)$$

$$H_k = H(P_k, P_b) \quad (7)$$

其中,  $P_k$  和  $P_b$  分别为本地数据集  $D_k$  与基准数据集  $D_b$  上的概率分布。

(2) 局部模型的聚合策略。

在 FedAvg 中,聚合策略仅根据客户端的数据量进行加权,其权重为:

$$q_k = \frac{|D_k|}{|D|} \quad (8)$$

在此基础上加入了对客户端上本地数据分布的考虑。为此,将上节中得到的平衡度通过归一化的方法转化为权重的形式。具体计算方式为:

$$s_k = \frac{B_k}{\sum_{i=1}^K B_i} \quad (9)$$

FLDWA 通过结合上述两种权重得到最终的综合权重。考虑到随着实际情况的变化,两种权重会对综合权重产生不同的影响,所以定义了影响因子  $e_q$  和  $e_s$ ,



表示它们对综合权重的影响程度。虽然数据量和平衡度都会对局部模型的质量产生作用,但是权重计算的是客户端在某个因素上的相对差距。例如数据量相同的多个客户端,通过式(8)计算出的权重也是相同的,此时可以仅使用由式(9)所确定的平衡度权重进行加权,因为数据量对局部模型质量的影响不存在差异。

使用与衡量本地数据集平衡程度相类似的方法量化不同客户端在数据量和平衡度两个方面的差异,并据此确定影响因子。由(8)(9)两式可知,两种权重的集合  $Q = \{q_1, q_2, \dots, q_K\}$  和  $S = \{s_1, s_2, \dots, s_K\}$  满足概率分布的特点。令概率分布  $R = \{\frac{1}{K}, \frac{1}{K}, \dots, \frac{1}{K}\}$  为基准分布,表示客户端不存在差异时的权重集合。则  $e_q$  和  $e_s$  的定义为:

$$e_q = \frac{H(Q, R)}{H(Q, R) + H(S, R)} \quad (10)$$

$$e_s = \frac{H(S, R)}{H(Q, R) + H(S, R)} \quad (11)$$

基于此,将式(3)的聚合策略改写为:

$$w_t = \sum_{k=1}^K (e_q \cdot q_k + e_s \cdot s_k) \cdot w_t^k \quad (12)$$

此外,由于  $\sum_{k=1}^K (e_q \cdot q_k + e_s \cdot s_k) = 1$ , 因此 FLDWA 与使用式(3)作为聚合策略的 FedAvg 算法具有相同的收敛性,FLDWA 能否收敛取决于 FedAvg 的收敛情况。FLDWA 的细节在算法 1 中进行了详细的描述。

#### Algorithm 1 FLDWA

Input:  $B$ :小批量梯度下降中的批大小

$E$ :每回合局部模型的训练次数

$\eta$ :学习率

Output: 模型参数

```

//服务器执行:
1: 初始化全局模型参数  $w_0$ 
2: for each round  $t = 1, 2, \dots$  do
3:   for each client  $k = 1, 2, \dots, K$  in parallel do
4:     if  $t = 1$  then // 初次通信时同时接收本地数据集平衡度和局部模型参数
5:        $B_k, w_t^k = \text{CLIENTUPDATE}(t, k, w_t)$ 
6:        $q_k \leftarrow \frac{|D_k|}{|D|}, s_k \leftarrow \frac{B_k}{\sum_{i=1}^K B_i}$  // 计算由数据量和平衡度确定的单项权重
7:        $e_q \leftarrow \frac{H(Q, R)}{H(Q, R) + H(S, R)}, e_s \leftarrow \frac{H(S, R)}{H(Q, R) + H(S, R)}$  // 计算影响因子
8:     else
9:        $w_t^k = \text{CLIENTUPDATE}(t, k, w_t)$ 
10:    end if
11:  end for
12:   $w_t \leftarrow \sum_{k=1}^K (e_q \cdot q_k + e_s \cdot s_k) \cdot w_t^k$  // 聚合
13: end for
14: function CLIENTUPDATE( $t, k, w_t$ ) // 该功能由客户端  $k$  负责执行
15:    $M_d \leftarrow$  (将本地数据集以批大小  $B$  进行切分)
16:   for each local epoch  $i$  from 1 to  $E$  do
17:     for batch  $b \in M_d$  do
18:        $w \leftarrow w - \eta \nabla l(w; b)$ 
19:     end for
20:   end for
21:   if  $t=1$  then // 初次与服务器通信时计算平衡度并上传
22:      $B_k \leftarrow 1 - H(P_k, P_b)$ 
23:     return  $w, B_k$  to server
24:   else
25:     return  $w$  to server
26:   end if
27: end function

```

## 4 仿真实验

文中实验是在 MNIST 数据集上进行的。该数据

集共包含 60 000 张训练图像和 10 000 张测试图像。测试图像中,不同类别对应的图像数量在 892 到 1 135 之间,为了构造出分布平衡的测试集,随机删除了一些图像,最后使所有类别对应的图像数量都为 892 张。

采用多层感知机 (multilayer perceptron, MLP) 和卷积神经网络 (convolutional neural networks, CNN) 两种模型来评估文中的研究,其网络结构与文献[5]中使用的模型保持一致。同时,选择 FedAvg 作为基准算法进行对比,该算法目前已用于众多 FL 实际应用中<sup>[22]</sup>。

为了对 FLDWA 进行测试和评估,设计了两组实验。第一组实验探究了本地数据集的平衡程度对算法的影响,第二组实验则对比了 FLDWA 与基准算法在多种不同设置下的表现。实验中所有结果皆为 10 次独立实验的平均值。

### (1) 本地数据集的平衡程度对算法的影响。

该实验为对照实验,分为正常组和不平衡组。从 MNIST 中抽取了 4 个数据量相同的子集作为本地数据集分发给客户端,正常组中 4 个本地数据集都是平衡数据集,不平衡组中则含有一个极度不平衡的本地数据集,其仅持有一个类别的数据样本。

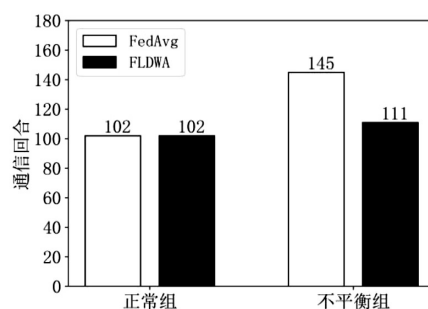


图 1 两种算法达到目标准确率所需通信回合对比

图 1 显示了 FLDWA 和基准算法在正常组和不平衡组使用 CNN 模型达到 98% 准确率所需的通信回合。可以看出,正常组中 FLDWA 与 FedAvg 的性能相同,因为相同的客户端数据分布并未对算法产生影响,此时可以将 FedAvg 视为 FLDWA 的一种特例。在不平衡组中,两种算法所需的通信回合都出现了增长,但 FedAvg 需要额外 30% 的回合才能达到目标准确率。这表明使用平衡度低的本地数据集训练的局部模型质量较低,由于 FedAvg 在聚合时对所有局部模型一视同仁,因此其性能明显低于 FLDWA。

图 2 显示了 FLDWA 中的聚合权重。FLDWA 为数据不平衡的客户端 4 分配了很小的权重,限制了其在聚合时的影响力,其他平衡的客户端被分配以较高的权重。可以发现,FLDWA 能够准确识别本地数据集的平衡度为其分配合适的权重,更加有效地聚合各方的信息。

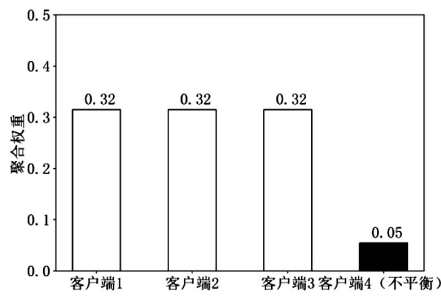
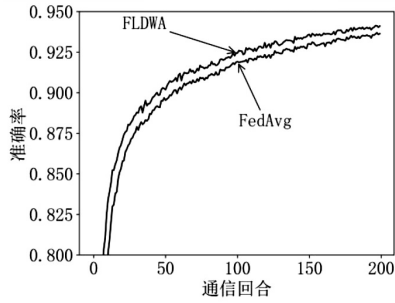
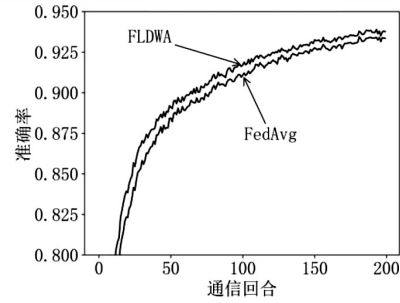


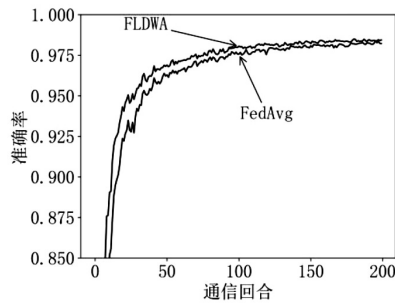
图2 FLDWA 中客户端的聚合权重



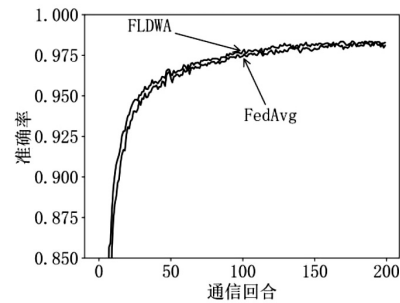
(a)MLP 数据量相同



(b)MLP 数据量不同



(c) CNN 数据量相同



(d) CNN 数据量不同

图3 不同实验设置下的测试集准确率和通信回合的关系

图3比较了两种算法在不同设置下的性能表现。可以观察到,FLDWA在各种设置下都以较少的通信回合达到了同样的准确率,并且在相同的通信回合中,其准确率均优于基线算法。这充分证实了FLDWA的聚合策略更为高效,有助于生成性能更好的全局模型。另一方面,(a)(c)两组实验中,客户端的数据量是相同的,此时FLDWA的聚合权重仅与本地数据集的平衡度有关。这也表明根据平衡度进行全局模型的聚合可以取得良好的效果,是一种有效的方法。

表1 不同设置下FedAvg和FLDWA达到目标准确率所需的通信回合以及以FedAvg为基准的通信回合的减少率

数据量	相同		不相同	
	MLP	CNN	MLP	CNN
FedAvg(回合数)	143.4	145.7	163.2	140.1
FLDWA(回合数)	118.0	122.9	139.4	117.7
减少率/%	17.7	15.6	14.6	15.8

表1中列出了不同的设置下(对应于图3中的(a),(b),(c),(d))使用两种算法达到目标准确率所需的通信回合(MLP模型目标准确率为93%,CNN模

## (2)FLDWA 与 FedAvg 的对比。

在该实验中,分别使用MLP和CNN模型对FLDWA和基准算法进行了对比,并且根据客户端上数据量是否相同,每种模型又分别进行了两组实验。构造客户端数据时,为了反映出FL数据非独立同分布的特点,在保证全局数据不平衡的前提下,随机控制本地数据集中的类别个数以及各个类别样本的数量,并且使得任何本地数据集之间都不存在相同的样本。

型则是98%),以及FLDWA相较基线算法的通信回合减少率。可以观察到在(a)组实验的设置下,FLDWA达到目标准确率平均需要118轮通信回合,而传统的FedAvg则需要大约143轮才能取得同样的效果,从而降低了17.4%的通信成本。尽管实验设置不尽相同,但在(b),(c),(d)三组实验上均可以得出类似的结论,所提出的算法分别将通信成本降低了15.6%、14.6%和15.8%。这证实了在全局数据不平衡时,FLDWA具有更高的通信效率,并且对于不同的数据量情况和不同的机器学习模型都具有很好的鲁棒性。

此外,作为对比和补充,也对FLDWA在全局数据平衡时的表现进行了评估。实验结果在表2中进行了展示。

表2 全局数据平衡时FedAvg和FLDWA达到目标准确率所需的通信回合

模型	MLP	CNN
FedAvg(回合数)	91.5	91.2
FLDWA(回合数)	90.3	90.2

从数据中可以看出在全局数据平衡时,两种算法

达到目标准确率所需的通信回合差异较小,FLDWA 仅带来了微弱的提升。这可能归因于全局数据平衡时,虽然也可能会出现分布不平衡的本地数据集,但是该数据集中缺失或冗余的信息恰好与其他本地数据集对应的部分互补,于是局部模型通过平均聚合能够获得很好的调整。同时这也表明,FLDWA 同样适用于全局数据平衡任务,并且在多数情况下表现出优于 FedAvg 的通信效率。

## 5 结束语

该文提出了一种基于数据分布加权聚合的 FL 算法 FLDWA,旨在提升 FL 在全局数据不平衡时的通信效率。该算法基于海林格距离对客户端的本地数据分布平衡程度进行了量化,并据此重新确定了其在 FL 聚合时的权重,使算法在更少的通信回合内收敛或达到目标准确率。实验结果表明,与流行的 FL 算法 FedAvg 相比,该方法有效降低了通信成本,并且在采用不同的机器学习模型和本地数据集大小时都有着很好的表现,具有较强的泛用性。

在接下来的工作中,将对该算法进行扩展,结合同态加密、安全多方计算等技术进一步为 FL 提供更强大的安全保证。此外,可能存在恶意节点向 FL 服务器发送错误的模型更新信息,从而降低 FL 的性能。如何检测和避免恶意攻击也是接下来重点关注的研究方向。

## 参考文献:

- [1] LI P, LI J, HUANG Z, et al. Multi-key privacy-preserving deep learning in cloud computing [J]. *Future Generation Computer Systems*, 2017, 74: 76–85.
- [2] TIKKINEN-PIRI C, ROHUNEN A, MARKKULA J. EU general data protection regulation: changes and implications for personal data collecting companies [J]. *Computer Law & Security Review*, 2018, 34(1): 134–153.
- [3] KONECNY J, MCMAHAN H B, YU F X, et al. Federated learning: strategies for improving communication efficiency [J]. *arXiv*:1610.05492, 2016.
- [4] 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 and Tutorials*, 2020, 22(3): 2031–2063.
- [5] MCMAHAN H B, MOORE E, RAMAGE D, et al. Communication-efficient learning of deep networks from decentralized data [J]. *arXiv*:1602.05629, 2016.
- [6] WEI W, LI J, CAO L, et al. Effective detection of sophisticated online banking fraud on extremely imbalanced data [J]. *World Wide Web*, 2013, 16(4): 449–475.
- [7] KUBAT M, HOLTE R C, MATWIN S. Machine learning for the detection of oil spills in satellite radar images [J]. *Machine Learning*, 1998, 30(2–3): 195–215.
- [8] DUAN M, LIU D, CHEN X, et al. Astraea: self-balancing federated learning for improving classification accuracy of mobile deep learning applications [C]//2019 IEEE 37th international conference on computer design (ICCD). New York: IEEE, 2019: 246–254.
- [9] DENG L. The mnist database of handwritten digit images for machine learning research [best of the web] [J]. *IEEE Signal Processing Magazine*, 2012, 29(6): 141–142.
- [10] 徐玲玲, 迟冬祥. 面向不平衡数据集的机器学习分类策略 [J]. *计算机工程与应用*, 2020, 56(24): 12–27.
- [11] KRAWCZYK B. Learning from imbalanced data: open challenges and future directions [J]. *Progress in Artificial Intelligence*, 2016, 5(4): 221–232.
- [12] LIN W C, TSAI C F, HU Y H, et al. Clustering-based under-sampling in class-imbalanced data [J]. *Information Sciences*, 2017, 409: 17–26.
- [13] CHAWLA N V, BOWYER K W, HALL L O, et al. SMOTE: synthetic minority over-sampling technique [J]. *Journal of Artificial Intelligence Research*, 2002, 16: 321–357.
- [14] ZADROZNY B, LANGFORD J, ABE N. Cost-sensitive learning by cost-proportionate example weighting [C]//Third IEEE international conference on data mining. Los Alamitos: IEEE, 2003: 435–442.
- [15] ZHOU Z H, LIU X Y. On multi-class cost-sensitive learning [J]. *Computational Intelligence*, 2010, 26(3): 232–257.
- [16] KONECNY J, MCMAHAN H B, YU F X, et al. Federated learning: strategies for improving communication efficiency [J]. *arXiv*:1610.05492, 2016.
- [17] CALDAS S, KONECNY J, MCMAHAN H B, et al. Expanding the reach of federated learning by reducing client resource requirements [J]. *arXiv*:1812.07210, 2018.
- [18] CHEN Y, SUN X, JIN Y. Communication-efficient federated deep learning with asynchronous model update and temporally weighted aggregation [J]. *arXiv*:1903.07424, 2019.
- [19] LIU L, ZHANG J, SONG S H, et al. Client-edge-cloud hierarchical federated learning [J]. *arXiv*:1905.06641, 2019.
- [20] YAO X, HUANG T, WU C, et al. Federated learning with additional mechanisms on clients to reduce communication costs [J]. *arXiv*:1908.05891, 2019.
- [21] 赵亮, 刘建辉, 王星. 基于 Hellinger 距离的混合数据集中分类变量相似度分析 [J]. *计算机科学*, 2016, 43(6): 280–282.
- [22] YANG T, ANDREW G, EICHNER H, et al. Applied federated learning: improving google keyboard query suggestions [J]. *arXiv*:1812.02903, 2018.