中 文 摘 要

近年来，相关国家和机构对于隐私保护问题越来越重视，期望在敏感数据不对外公布的前提下获得性能良好的机器学习模型。因此，联邦学习旨在数据保存在本地进行机器学习模型的非集中式训练，即客户端持有本地隐私数据，在训练过程中，数据不对外公开发布，客户端彼此交换的是机器学习模型。然而，在隐私保护背景下的模型训练存在许多问题和挑战，数据非独立同分布是影响全局模型性能的主要原因之一。本文针对数据非独立同分布下各客户端数据标签分布偏移情况进行了深入的研究。具体工作如下：

（1）提出基于正负梯度样本加权的训练方法和聚合策略。从模型训练入手，基于反向传播的角度分析，首先阐明了数据标签分布偏移对梯度带来的影响，然后引入正负样本以及正负梯度的概念，并更改传统训练的 softmax 分类函数和损失函数，使得在模型训练过程中，对于客户端存在的类别，只受到存在标签样本正负梯度的影响，对于客户端中缺失的样本类别，将完全舍弃其他存在类别样本带来的负梯度影响。此外，模型发送至服务器时，根据余弦相似度衡量客户端模型之间的差异并聚合，以提高联邦学习算法的收敛性和收敛速度。

（2）基于正负梯度样本加权的联邦学习算法中，全局模型关注的多分类优化问题在客户端上变为贴合本地数据分布的子优化问题，通过延缓本地模型优化速度进而促进全局模型的收敛。但在全局优化问题下，拆解后的子优化问题丢失了来自负类样本的监督信息，因此提出了分类器类别嵌入向量正则化策略，优化各分类向量在特征空间中的角度以获得更好的分类边界，以期望提升全局模型的泛化性能。

关键词：联邦学习；数据非独立同分布；正负梯度；类别嵌入向量正则化

ABSTRACT

In recent years, relevant countries and institutions have paid more and more attention to privacy protection issues, and hope to obtain high-performance machine learning models without disclosing sensitive data to the public. Therefore, federated learning aims at non-centralized training of machine learning models where data is kept locally. In other words, during the training process, the client holds local private datas which are not publicly released, and the client exchanges the machine learning model with the server. However, there are many problems and challenges in model training under the background of privacy protection, data with non-independently and identically distribution is one of the most important reasons affecting the performance of the global model. Therefore, This paper conducts in-depth research on the label distribution skew on various clients under non-independently and identically distribution .The specific work is as follows:

（1） Propose a training method and aggregation strategy based on weighted positive and negative gradient samples. From the perspective of model training and back propagation, this paper firstly analyzes the impact of data label distribution skew on the gradient, then introduces the concept of the positive and negative samples and the positive and negative gradients. Meanwhile, the method changes the softmax classification function and the loss function of traditional training, so that the classes which are existing on the client are only affected by the positive and negative gradients of the label samples in the process of model training. For the missing sample classes in the client, the negative gradient effect brought by different labels samples will be completely discarded. In addition, when the model are uploaded to the server, the differences between client models are measured and aggregated by cosine similarity , so as to improve the convergence and the speed of the federated learning algorithm.

（2） In the federated learning base on the positive and negative gradients samples weighted, The multi-classification optimization problem concerned by the global model becomes a sub-optimization problem that fits the local data distribution on the client side, and the convergence of the global model is promoted by slowing down the local model optimization speed. However, under the global optimization problem, the disintegrated sub-optimization problem loses the supervision information from the negative class samples.. In order to improve the generalization

performance of the global model, therefore, a regularization strategy is proposed to optimize the angle of each classification vector in the feature space to obtain a better classification boundary.

Keywords： Federated Learning; Non-independently and identically distribution data; Positive and negative gradients; Class embedding vector regularization

目 录

第一章 绪论

1.1 研究背景与意义

近些年，得益于深度学习技术的进步，人工智能领域飞速发展，并在多个领域如人脸识别、推荐系统、目标检测、智能翻译等获得了广泛应用。但影响深度学习模型性能的关键在于拥有高质量且充足的训练数据以及强大的计算能力。目前实现深度学习算法的主要步骤为，构造数据集、确定使用的模型结构、确定损失函数，然后使用梯度下降法对模型参数进行优化求解，当获得最小损失函数值时，算法收敛。然而在当今数据爆炸的时代，单个计算机硬件的计算能力制约着算法的实现，为此，出现了分布式机器学习，其利用多个计算节点协同完成全局模型的训练。然而，随着隐私保护等法律法规的提出，很多数据源之间无法直接交换数据，导致“数据孤岛”现象的出现，制约了人工智能的发展。2018 年欧洲联盟出台《通用数据保护条例》，旨在保护用户的个人隐私和数据安全。用户可以删除或撤回其个人数据，没有用户的允许，公司的不可以有其它用途。违反该法案的公司将面临高额罚款。同时我国在 2021 年 11 月 1 日出台《中华人民共和国个人信息保护法》，禁止公司不合理使用用户数据。一个 AI 项目可能涉及多个领域、多个机构，可能需要融合多个公司、多个部门的数据，（比如居民线上消费问题，可能需要多个银行的数据），因为涉及到用户隐私，数据的拥有者通常拒绝将数据对外发布，因此将这些分散的数据整合到一起几乎是不可能的。分布式机器学习虽然可以解决算力不足的问题，但却无法解决隐私保护[1]的问题，因此联邦学习应运而生。2015 年，Google 公司提出了联邦学习[2-4]（Federated Learning）的概念，旨在解决数据敏感场景下机器学习问题。该方法将数据的独立拥有方（公司或机构）作为为客户端，引入受信任的中心服务器协调训练，各个客户端在不对外公布涉及到用户隐私和本地用户偏好的私有数据的情况下，完成基于本地数据的模型训练，并将其训练后的模型上传至服务器。中心服务器作为受信任方，基于接收到的客户端模型完成聚合获得全局模型后广播给客户端。通过迭代更新客户端模型和服务端全局模型，最终得到表现良好的全局模型。

根据在联邦学习[5]训练过程中各客户端特征属性和样本数据之间的差异，可分为横向联邦学习（Horizontal Federated Learning, HFL）、纵向联邦学习（Vertical Federated Learning, VFL）以及联邦迁移学习（Federated Transfer Learning, FTL）。在横向联邦学习中，不同的客户端在数据特征方面是重叠的，但是数据样本，即给定特征下数据样

本重叠程度不高，例如，联邦学习的参与方是属于两个地区不同的商业银行，他们具有相同的商业模式和经营业务，因此具有相同的特征属性，但他们的用户可能是距离各自最近地区的居民，因此客户群体的差异度比较大。在纵向联邦学习中，联邦学习客户端在特征属性方面重叠不多甚至没有重叠，但是给定特征下数据样本重叠程度比较高。例如两家公司（电子商务公司和银行），它们经营的业务不同，对相同的用户提供不同的服务，拥有客户不同特征的数据，但是客户群体高度重叠。在联邦迁移学习中，不同的联邦学习客户端在数据样本和特征属性方面都高度不同，因此，一般使用迁移学习将源域的知识迁移到目标域或者互相迁移，从而获得表现良好的机器学习模型。本文仅考虑横向联邦学习的情况。

联邦学习在各客户端数据独立同分布(Independently Identically Distribution，IID)的情况下能够获得较好的全局模型。然而，不同的客户端数据源之间其数据分布并不一定完全一致，即客户端之间数据是非独立同分布(Non-IID)的。各客户端模型均是基于其本地数据训练获得，因此，当各客户端之间数据非独立同分布时，其训练的模型之间差异较大，导致服务端全局模型的泛化性能变差，收敛速度降低，从而增加了模型训练的通信成本[6, 7]，造成资源的浪费和损耗。

目前常见的数据非独立同分布[8-10]问题主要包括以下几种:

(1)类别标签分布偏移(Label Distribution Skew)。这类 Non-IID 数据集主要是指数据类别标签分布发生了偏移。例如，Mnist 数据集中有 10 个标签，假设有 3 个客户端，每个客户端并不完全包含这 10 个标签的数据，也就是说只有 10 个标签中若干标签的数据样本。

(2)特征分布偏移(Feature Distribution Skew)。这类 Non-IID 数据集主要是说特征值的分布发生了变化，即每个客户端上样本特征值的差异比较大。

(3)标签相同但特征不同(Same Label But Different Features)。此类情况下不同客户端之间存在相同的标签，但是其数据样本差异较大。这种情况在实际问题中很常见，例如对于同样标签为猫的样本，在不同客户端上是不同品种的猫，它们的颜色、大小、样子可能都不相同。

(4)标签分布不同但是却特征相同(Same Features But Different Labels)。此类问题主要是不同客户端之间存在相同的特征样本，但其具有不同的标签。例如在不同客户端上颜色、大小、样子都相同的样本，其对应的标签一个可能为美短猫，另外一个可能为英短猫。

(5)数量不同(Quantity Skew)。此类情况下主要是在每个客户端下，其不同标签的

数据量不同。例如，针对 Mnist 数据集，在某个客户端中，标签为 1 的样本数量有 100 条，标签为 9 的样本数量有 2000 条。

类别标签分布偏移在实际的应用问题中广泛存在，如眼科医院和肿瘤医院中相同病人的诊断标签是不同的。而由于客户端模型仅考虑本地数据样本，导致服务端模型聚合后泛化性能变差，且很难获得提升。为了在类别标签分布偏移的非独立同分布数据情况下获得较好的全局模型，本文引入正负样本和正负梯度的概念，提出了客户端正负样本加权训练策略和服务端聚合策略。

1.2 国内外相关工作研究进展

联邦学习旨在协同建立一个基于分布式数据集的机器学习模型，一般为深度神经网络，在联邦学习算法训练过程中，与模型权重参数相关的信息可以在服务器和客户端之间交换（或者以加密的方式交换），但是每个客户端节点持有的数据集不能交换，这么做旨在最大限度的保护隐私，已经训练好的联邦机器学习模型可置于联邦学习的各参与方，也可以在多方之间共享。

1.2.1 联邦学习的主要研究方向

目前针对联邦学习中存在的问题，主要有以下几个方面的研究：

1.在数据非独立同分布情况下，主要解决全局训练过程中的收敛问题，在收敛速度和全局模型泛化精度间达到平衡，提高全局模型的泛化能力。

2.降低联邦学习算法的通信频次，用少量的通信达到收敛。网络中的数据传输是比较昂贵的问题，如果能用少量的通信次数快速提升联邦学习算法的泛化能力，将大大减少联邦学习的资源[11]消耗。该研究在数据独立同分布情况下，有三值压缩、二值化等方式压缩通信时传送的数据量，在保证全局模型泛化效果的情况下减少通信数据量。

3.联邦学习中的隐私保护问题，联邦学习本身只保证本地数据集不会直接对外发布，但是模型的梯度、参数等信息是向服务器上传的，在通信过程中，很容易被泄露并且从中推断出用户的数据。

4.联邦学习模型的鲁棒性，容易有恶意节点发送错误的梯度或模型给服务器，因此需要设计防范恶意攻击的算法。

1.2.2 数据非独立同分布下的联邦学习

由于实际应用中很多问题的数据集是非独立同分布的，因此很难获得泛化性能较

好的全局模型。为此，学者们针对数据集非独立同分布的情况展开了研究，其大致可以分为四类，下面对其分别介绍。

1.2.2.1 正则化训练和聚合方法

当数据异构的时候，随着优化过程的不断进行，不同客户端的模型会偏离中心服务器的模型，不同的客户端存在权重差异，这种差异和数据的异构程度呈正相关，数据差异越大，模型差异越大，如果只对它们进行平均，得到的全局模型泛化效果很差。为此，TL[12]等人提出了 FedProx 算法，该算法允许各个局部模型（客户端模型）不必每次都优化到本地最优解，在客户端本地模型训练时在传统损失函数的基础上加入了近端算子（proximal term），将损失函数修改为：

$$F(w) = l(w,x) + \frac{\mu}{2}\left|\left|w - w^*\right|\right|^2 \tag{1.1}$$

在这个式子中主要是多了惩罚项，w^*代表的是全局模型，意为在边缘节点优化时，容忍客户端模型的异构，但是不能太过于偏离全局最优解。虽然引入惩罚项后，在一定程度上可以减缓 Non-IID 对于全局模型的影响，但是却又额外引入了新的超参数。因此其存在的缺点是边缘节点可能需要仔细调整 μ，用以保证良好的收敛性和泛化能力。如果μ太小，那么正则化项几乎没有影响。如果μ太大，那么本地更新就非常小，收敛速度也很慢。

在FedNova[13]中，客户端节点拥有不同的算力，在相同的训练时间约束下，当不同客户端拥有不同的数据量大小或数据集是非独立同分布时，具有较多本地步骤的客户端将有更大的本地更新，如果简单平均，这将对全局更新有更显著的影响。因此，为了确保全局更新不存在偏差，FedNova在更新全局模型之前，根据他们的本地步骤数对每个边缘节点的本地更新进行归一化和缩放。具体的权重更新公式为:

$$w^{t+1} = w^t - \eta \frac{\sum_{i \in S_t} \left|D^i\right|_\tau}{n} \sum i \in S_t \frac{\left|D^i\right| \Delta {w_i}^t}{n\tau_i} \tag{1.2}$$

其中S_t代表的是边缘节点的数量，D代表的是每个边缘节点的数据集，τ代表每轮通信本地的优化次数。FedNova 对 FedAvg 引入了轻量级的修改，并且在更新全局模型时的计算开销可以忽略不计。

文献[14]提出的 Scaffold 算法引入了方差，并应用方差减少技术，为服务器(即 C)和各客户端(即 Ci)引入了控制变量，用于估计服务器模型的更新方向和每个客户端的更新方向。然后用这两个更新方向的差值来近似局部训练的漂移。因此，Scaffold 算法通过在局部训练中添加漂移项来修正局部更新，通过计算全局模型中局部数据的梯度

或重复使用先前计算的梯度两种方法来更新局部控制变量。两者各有优劣，而第一种更新方法更稳定，第二种方法的计算成本更低。与 FedAvg 相比，由于引入额外的控制变量，Scaffold 每轮的通信大小增加了一倍。虽然在一定程度上克服了数据非独立同分布问题，但通信量太过巨大，尤其网络模型较大时更为明显。

1.2.2.2 个性化联邦学习方法

为了应对联邦学习中数据非独立同分布带来的挑战，一种比较有效的方法是根据不同的客户端的本地数据分布不同，对客户端训练的模型做个性化[15-18]处理，为每个客户端定制高质量个性化模型。该方法不再关注全局模型的泛化能力，服务器的作用也不是聚合好的全局模型而是协助个性化模型的建立。个性化联邦学习分为两种，一种是全局模型个性化，意在提升在异质数据上的联邦学习共享模型的性能；另一种是直接训练个性化模型，提供个性化解决方案，这类方法在客户端上训练单个的个性化模型，通过修改传统联邦学习的聚合过程来建立个性化模型。大多数个性化模型技术分为两步：先在服务器上获得全局模型广播给客户端，然后客户端再使用自己的私有数据进行全局模型的个性化处理。然而，两步个性化学习可能存在问题，即在服务器上协作训练好的全局模型并不利于本地的个性化训练，在服务器上表现好的模型可能在客户端上进行若干次梯度下降之后收敛到局部最优点难以跳出，因此Jiang[19]等人提出了让联邦个性化学习更有意义的方法，该方法要同时解决以下三个问题：1.要训练一个有利于大多数客户端的全局模型，2.训练一个精确的模型，使得在本地数据较少的客户端也能够得到最佳的全局模型，3.能够在很少的训练次数下实现收敛（对于全局模型来说），减少训练次数主要是为了防止过拟合。Dinh[20]等人利用Moreau Envelope作为正则化损失函数，提出了一个新的面向个性化联邦学习的双层优化算法pFedMe。该方法将个性化模型的优化过程与全局模型的学习过程解耦。因此，pFedMe以类似于FedAvg等标准联邦学习算法的方式更新全局模型，同时以较低的复杂度并行优化个性化模型，该方法取得了不错的效果。

2021 年 Collins[21]等人提出了一种基于个性化层的方法，该方法通过找到全局共享特征提取层、个性化分类层的方法来提高联邦学习模型的泛化能力。先将高维的数据输入映射为低维的特征表示，并将这个映射(模型)与其他客户端的特征提取器聚合获得全局特征提取层参数，随后再通过每个客户端上个性化分类器层将经过特征提取层的低维特征表示映射为网络输出结果。该方法中的特征提取层由不同的客户端共享，同时个性化了分类器层，模型的泛化性能主要由客户端私有的分类器层决定。

1.2.2.3 针对分类器层做特殊操作的方法

此类方法采用分层训练的方式，将模型分成两个部分，特征提取器和分类器，先训练特征提取器再单独矫正分类器。

在人脸识别领域，Yu[22]等人在 2020 年提出了联邦学习只在正样本下训练的方法，人脸识别场景是典型的数据非独立同分布场景下数据标签分布偏移问题，其中人脸数据又具有强隐私性，因此设计一个算法提高人脸识别模型的准确率是棘手的问题。该文将人脸识别模型分为两部分，针对分类器层做特殊训练，使其保持一定的分类间隔和距离，提高了模型的准确率。Rawat[23] 等人仅在正样本下训练的方法也是对分类器进行训练，将神经网络模型等价为特征提取模型和网络分类器两部分，使用矩阵表示和对比学习[24]的方法，将每个人脸的特征向量在服务器端使用 Spreadout Regularize 重新训练，使人脸特征向量保持一定的距离并聚合，提高了全局模型泛化能力。

1.2.2.4 使用知识蒸馏的方法

知识蒸馏(Knowledge Distillation)[25-27]或者迁移学习[28-32](Transfer Learning)方法是在保护本地数据集的情况下使用自编码器或其他生成模型产生边缘模拟数据，促进全局模型的聚合。

OneShotFL[33]是 Zhou 等人提出的全新联邦学习[34]训练方法，将知识蒸馏引入到联邦学习中，更改了传统的训练方式和聚合方式。其具体做法为将多轮次的与服务器通信改为一次，同时改变传统的聚合策略，将传统联邦学习的参数平均修改为知识蒸馏的方式。客户端首先在本地将服务器训练的足够好，其次将客户机模型发送到服务器，服务器使用公共数据集将客户端模型作为 teacher 模型，将模型的知识从客户端模型转移到服务器模型中。该方法降低了通信频次，提高了联邦学习算法的效率。文献[35]提出了一种数据集蒸馏的新方法，该方法改变传统蒸馏模型的方式，从蒸馏模型转移到了蒸馏数据，使用一个小批量的数据训练出的模型和使用原始数据训练出的模型泛化效果几乎一致。其通过构建一个双层优化问题来寻找一个更好的蒸馏数据集，在数据集蒸馏时，得到的蒸馏图片比较模糊，这在一定程度上保护了客户端的隐私。Zhou[33]等人在数据集蒸馏的基础上进一步扩展了这种训练方式，其具体做法是每个客户端蒸馏本地数据集，将蒸馏过后的本地数据集传到服务器，服务器使用蒸馏过的数据训练全局模型。该工作的核心假设为蒸馏过的数据即使对外发布，也不会对隐私造成泄露。Lin[36]等人在联邦平均基础上引入知识蒸馏，在每轮联邦学习模型聚合完成之后，使用知识蒸馏微调模型，使得模型参数对于数据异构的敏感性没那么强，提高了

模型的鲁棒性。

文献[37]使用生成对抗网络解决联邦学习中的域偏移问题，通过对抗性技术解决联邦学习系统中的域偏移问题。通过为每个源域节点训练一个模型并使用源域梯度的聚合更新目标模型来保护数据隐私，设计了对抗域适应方法。该方法在自然语言处理和图像识别多个任务上测试，取得了不错的效果，但该方法存在通信代价过大的问题。

FedGen[38]通过在服务器端学习一个仅在用户模型的预测规则上派生的生成模型，通过给定一个目标标签就可以生成和客户端数据分布一致的特征表示，同时广播该生成器给每一轮参与联邦训练的客户端，使用生成的增广特征表示和本地的数据来训练客户端模型。这体现了来自于其他客户端的知识，维护一个远小于真实数据集特征输入的潜在空间。FedGen[38]学习到的生成器是轻量级的，在克服 Non-IID 问题的同时为整个训练过程引入了很小的通信开销。该算法的性能取决于生成器的选择以及生成器学习的效果好坏，同时容易受到数据集的影响，在复杂的数据集中，该算法难以发挥作用。

1.3 主要研究工作和内容

本文主要针对横向联邦学习中数据标签分布偏移情况下对影响联邦学习性能的因素进行分析和研究，引入正负样本和正负梯度的概念，提出了客户端正负样本加权训练策略和服务端聚合策略。随后提出自适应调节分类器权重的优化算法，并在最后对全文进行了总结和展望，本文的具体内容组织如下：

第一章介绍了论文的研究背景及意义，分析了目前联邦学习的主要研究方向，回顾了数据非独立同分布下现有的联邦学习算法。

第二章简要介绍了神经网络以及基本的联邦学习算法。

第三章详细介绍了基于正负梯度加权的客户端训练算法和服务端聚合算法。分析了在标签分布偏移场景下，客户端训练存在的主要问题。引入正负梯度[39]概念，模型训练中对正负梯度进行加权，以保证算法的可收敛性。考虑到正负梯度加权带来的不同类别向量在不同客户端训练收敛速度不一致的问题，提出了基于余弦相似度的聚合策略，以期有效加快标签分布偏移这种数据非独立同分布情况下全局模型的收敛速度。

第四章详细介绍了基于 softmax 类别嵌入向量正则化的改进联邦学习算法。基于正负梯度加权的联邦学习算法，使用延缓本地收敛的方式正则化客户端的训练过程，以保证联邦聚合算法的收敛，同时也将一个多分类的问题转变成了若干个子问题来解决，但是该方法没有考虑到分类任务类别特征向量类间的相似度。因此，使用 softmax 作为

分类器，自适应调节类别嵌入向量在空间中的夹角，使得它们在空间中的分布更均匀，以此提高最后模型的准确率。

第五章对本文进行了总结，并对今后工作进行了展望。