面向隐私保护的联邦学习模型训练效率优化与跨设备协同学习实践

刘有缘

上海学昶信息技术有限公司，上海闵行区，201100；

摘要：在隐私保护背景下，联邦学习模型训练过程中，数据的非唯一性、不可预测性、分布不均匀等特点使得数据和模型的安全性受到挑战。如何在保证数据隐私和安全性的同时，实现模型的高效训练，成为联邦学习技术研究的关键。本文首先探讨了面向隐私保护的联邦学习模型训练过程中存在的问题，然后介绍了基于数据传输与同步技术和差分隐私技术的联邦学习模型训练效率优化方法，并基于上述优化方法进行了模型训练效率优化实践。最后以跨设备协同学习为例，探讨了基于隐私保护和性能提升相结合的联邦学习模型训练方法，并在数据安全和隐私保护下实现了跨设备协同学习。

关键词：面向隐私保护；联邦学习；模型训练效率优化；跨设备协同学习

DOI：10.64216/3080-1508.25.09.065

引言

随着大数据时代的到来，数据的安全与隐私问题愈发突出，联邦学习技术在保护用户数据安全的同时，能有效整合多方数据，提升算法训练效率，为用户提供更精准的数据服务。但在联邦学习模型训练过程中，数据传输过程中存在的安全风险以及通信时延等问题制约了模型训练效率。因此，如何在保证用户数据安全的前提下，实现模型的高效训练成为联邦学习技术研究的关键。本文将重点针对上述问题开展研究工作，并基于联邦学习技术在跨设备协同学习领域进行实践，以期为联邦学习技术应用提供参考。

1 联邦学习概述

联邦学习技术是一种在云端利用分布式计算框架，将训练模型的本地数据集分布到各参与方，由参与方基于联邦学习框架进行本地训练，在完成各自本地模型训练后，将模型结果返回到云端进行汇总、建模和联合学习的一种机器学习技术。在联邦学习技术框架下，参与方在保护用户隐私的前提下，将数据集合到一个公共服务器上进行统一训练，各参与方在完成各自数据训练后将结果返回给该公共服务器[1]。

2 面向隐私保护的联邦学习模型

联邦学习模型训练过程中，数据隐私和安全问题是其面临的首要问题，随着联邦学习技术的不断发展，联邦学习模型训练过程中的数据安全与隐私保护问题日益突出。如何在保证数据安全和隐私保护的前提下，实现联邦学习模型训练效率优化，成为联邦学习技术研究的关键。目前，面向隐私保护的联邦学习模型训练算法主要有两类，一是基于模型并行化的联邦学习算法；二是基于模型分组化的联邦学习算法。此外，针对隐私保护技术在联邦学习中的应用问题，主要有差分隐私技术、同态加密技术、多方安全计算技术和安全多方计算技术等。

2.1 面向隐私保护的联邦学习算法

联邦学习的核心思想是基于数据拥有者的私有信息，在多方参与的情况下进行联合训练，在不共享数据的前提下，实现模型共享，从而保护数据拥有者的隐私。联邦学习算法主要分为两种类型：基于模型并行化的联邦学习算法和基于模型分组化的联邦学习算法。基于模型并行化的联邦学习算法将联邦学习过程中所涉及的各个参与方联合起来[2]，共同参与训练；而基于模型分组化的联邦学习算法在确保数据不出本地的前提下，通过模型分组以及对参数进行归一化等处理，对不同参与方之间数据进行加密与转换，从而保证其在本地进行隐私保护训练。

2.2 隐私保护技术在联邦学习中的应用

针对上述面向隐私保护的联邦学习算法，本文在联邦学习模型训练过程中主要采用了差分隐私技术、同态加密技术、多方安全计算技术和安全多方计算技术等。差分隐私技术是指当数据加密密文出现微小差异时，可以通过差分的方法来保护数据的隐私，从而在保证数据安全的同时，实现模型训练任务。同态加密技术是指对密文进行加密处理后，可以对加密后的密文进行计算，但是密文中的某些信息无法被解密。多方安全计算是指在联邦学习中，不同参与方之间不需要共享隐私信息，

但是需要合作完成模型训练任务。

3 模型训练效率优化

3.1 联邦学习模型训练过程分析

在联邦学习中，数据参与方需要建立自己的本地训练模型，并在本地训练模型的过程中，参与方不能获取对方的数据和参数。模型的训练过程需要三个阶段：预处理（Pre-expression）、特征提取（Feature Expression）和模型训练（Training）。预处理阶段，参与方对数据进行预处理，生成原始数据和本地模型；特征提取阶段，参与方对原始数据进行特征提取，将原始数据转换为目标模型的输入数据；模型训练阶段，参与方对目标模型进行训练并生成目标模型的输出结果。该过程与传统机器学习方式下的训练过程相似。

3.2 模型训练效率问题分析

为了研究在隐私保护情况下联邦学习模型训练效率问题，我们对其进行了实验分析。首先，在数据不存在隐私保护的情况下，在训练模型的过程中，会发生以下几个问题：一是模型训练需要遍历整个数据集，数据处理和特征提取都需要时间；二是在训练过程中，会出现不同的网络模型之间进行模型训练时，需要做网络聚合，存在大量的计算量[3]；三是在模型训练过程中，会出现梯度消失和梯度爆炸问题。为了解决上述问题，我们采用了基于循环神经网络的联邦学习框架。该框架将联邦学习任务分解为多个子任务，每个子任务由多个任务进行并行处理。

3.3 模型训练效率优化方法探讨

针对上述问题，结合业界已有研究成果，我们提出了一种基于联邦学习框架的模型训练优化方案，具体工作流程如下。方案一：采用双机分布式训练，在保证模型训练效率的前提下，尽量减少单机训练的时间开销。方案二：采用联邦学习框架对训练数据进行预处理，通过对数据进行分布式并行化处理，提升联邦模型的训练效率。方案三：通过结合联邦学习框架和模型迁移学习，在保证模型训练效率的前提下，将模型迁移至其他终端设备上进行并行化处理。方案四：采用分布式深度学习框架 TensorFlow 进行联邦学习框架的开发。

4 跨设备协同学习实践

4.1 跨设备协同学习框架设计

根据前面的分析，基于云平台的分布式训练，可以在不泄露用户隐私的情况下完成模型训练。但是在实际应用中，模型数据需要在不同设备之间进行传输，那么就需要考虑如何保证数据在不同设备之间的一致性，即如何在保证数据安全的情况下实现模型协同。

本文设计了跨设备协同学习框架，并提出了基于云平台的跨设备协同学习算法，能够保证模型训练中的数据一致性和模型在不同设备之间的一致性。在具体的框架设计上，本文的框架主要由模型协同训练两个部分组成。在模型协同训练阶段，通过搭建统一的云平台，并对不同设备进行配置，在云平台中完成模型参数的收集、本地设备的训练以及云端服务器的训练；在模型协同训练阶段，通过构建一个统一的分布式架构，将多个设备上收集到的数据进行融合处理，并将数据发送至云端服务器进行模型训练。在该框架中，模型协同训练部分主要包含两个子模块：一是构建统一的云平台；二是对不同设备上收集到的数据进行融合处理，并将计算结果发送至云端服务器进行模型训练。

4.2 跨设备数据传输与同步技术

在实际应用中，设备间的数据同步和传输是一个比较耗时的环节。通常情况下，我们会采取如下几种方法进行数据同步：（1）根据数据量的大小，选择最合适的传输方式；（2）如果数据量不大，可以采用简单的共享表方式，由客户端直接进行同步；（3）如果数据量较大，可以采用缓存和分布式消息队列进行同步；（4）如果有多个设备，可以采取异步数据同步方法，即将不同设备的数据更新到不同的设备上。具体方法是将更新到该设备上的最新数据进行缓存，然后再发送到其他设备[4]。在实践中，我们选择使用第二种方式进行跨设备的协同训练。具体来说，在训练阶段，我们可以采用分布式消息队列和共享表两种方式来同步数据。分布式消息队列可以实现多台设备的数据同步，但是效率较低，不适合大规模的协同训练；共享表方式简单、高效，但是需要将所有设备的数据更新到一张共享表上，效率很低。在实践中，我们采用了一种介于分布式消息队列和共享表之间的技术来解决这个问题。

4.3 实验设计与结果分析

为验证模型训练效率的提升效果，本文采用跨设备数据同步方案进行跨设备协同训练。实验使用同一数据集，在同一实验平台上对不同的模型进行训练。其中，模型的训练精度表示为 RMSE，目标是使训练结果在目标精度附近，最小化模型的方差；模型的训练效率表示为运行时长，目标是使训练时间最小化。实验中选择了多个模型，并进行了多次实验。在本实验环境下，基于自适应精度分配策略的跨设备数据同步方案能够显著降低模型训练耗时和方差。经过多次实验验证，相比于其他方案，本方案在保证精度的前提下大幅降低了模型

训练耗时和方差。在训练精度上，通过优化精度分配策略，本方案取得了比传统的训练效率优化技术更高的 RMSE，是当前实现方案中的最佳选择。在训练时间上，本方案使得模型训练速度比其他方案有显著提升。实验中选择了多个模型，验证了跨设备数据同步技术在不同场景下的适用性。为了进一步验证本方案对于隐私保护的支持，本文对不同隐私保护强度下的模型训练结果进行了对比分析。对于低强度隐私保护场景，本方案与其他算法相比在 RMSE 方面不具有显著优势；对于高强度隐私保护场景，本方案与其他算法相比在 RMSE 方面具有显著优势。

5 跨设备协同学习实践案例分析

5.1 跨设备协同学习流程设计

在此，我们以数据智能行业中的某银行客户为例，结合具体的业务场景，进行跨设备协同学习的实践。我们将跨设备协同学习分为三个阶段：数据收集、模型训练与模型验证。数据收集阶段，首先针对该银行客户进行模型训练，并通过数据智能行业中的智能算法模型进行验证。模型训练阶段，数据智能行业中的智能算法模型在该银行客户设备端上运行，通过对该客户的业务特征进行分析，并基于该银行客户的实际业务数据，进行训练。模型验证阶段，该银行客户业务特征输入到上述智能算法模型中，并基于上述模型进行验证。最终，完成跨设备协同学习。

在本案例中，数据智能行业的智能算法模型在某银行客户设备端运行，但由于该银行客户与数据智能行业之间并不存在直接的交互，所以，无法进行直接的跨设备协同学习。而该银行客户与数据智能行业之间存在交互，通过特定的交互协议，可以完成数据、算法模型和训练结果的共享。在此基础上，我们需要完成跨设备协同学习，将该银行客户与数据智能行业之间的交互信息进行有效地处理，通过特定的协议将模型训练结果和训练模型运行时产生的隐私保护参数提供给数据智能行业。最终完成跨设备协同学习任务。

5.2 实验设计与数据分析

本文主要针对基于联邦学习的模型训练效率优化问题开展实验设计，并选取不同的参数进行分析，对比不同参数组合对模型训练效率的影响。同时，针对多设备协同学习场景，实验设计了三种不同的协同学习策略，即基于安全多方计算的策略、基于联邦学习的策略和基于隐私保护联邦学习的策略，并在真实数据上进行了模型训练效率优化效果评估。对比不同协同学习策略下模型训练效率的变化情况[5]。可以发现在本文所提协作学习策略下，模型训练效率整体得到了较大提升，同时在部分场景中验证了隐私保护联邦学习策略对模型训练效率提升的有效性。

5.3 结果讨论与展望

本文研究成果在智慧医疗领域得到了应用，取得了很好的效果。本研究实现了跨设备协同学习在移动端上的应用，并取得了理想的效果。实验结果表明，移动端与 PC 端的模型训练效果较好，可以满足智慧医疗领域的实际需求。但目前该方法还存在一些不足，主要表现在以下三个方面：第一，移动端与 PC 端之间数据传输存在一定的限制，在实际应用中，移动端和 PC 端需要同时访问同一个数据才能取得最优的效果。第二，在移动端上训练模型时，只考虑了数据量对模型效果的影响。未来可以考虑进一步考虑将数据量与模型效果结合起来进行分析。第三，由于移动设备与 PC 设备存在差异，不能通用。

6 结语

在大数据时代，数据智能在各行各业中得到了广泛应用，但在数据使用过程中，不可避免会涉及隐私保护问题。联邦学习作为一种新兴的机器学习方法，可以有效解决数据不互通、数据孤岛等问题，因此受到了广泛关注。本文以联邦学习模型训练效率优化为研究对象，分析了当前联邦学习模型训练效率优化方面存在的问题，提出了基于联邦学习框架的模型训练效率优化方法，并通过实验进行了验证。此外，本文还结合实际应用场景，探索了跨设备数据同步技术在模型训练中的应用。最后，本文基于跨设备数据同步技术设计并实现了跨设备协同学习框架。

参考文献

[1]刘洵.联邦学习的安全梯度聚合及训练效率优化方法[D].中国科学技术大学,2025.

[2]王建树,张南方,徐方,等.基于卷积神经网络的异型卷烟分拣优化算法研究[J].湖北工程学院学报,2024,44(06):82-87.

[3]沈良铎.无线联邦学习的能量效率优化方法[D].北京邮电大学,2024.

[4]曹婧.面向异构分布式系统的深度神经网络训练效率优化方法研究[D].中国科学技术大学,2024.

[5]姚飞翔.异构 GPU 集群下分布式深度学习训练效率优化研究[D].西北农林科技大学,2023.