联邦学习环境下隐私保护的数据聚合与评估方法

张　黎
闪捷信息科技有限公司　浙江杭州　311100

摘　要：现代深度学习的辉煌成就，得益于大数据的普遍可获取性。然而，随着公众对隐私问题关注度的上升，各国纷纷推出隐私保护相关法律，隐私保护措施日益严格。这一转变加剧了数据的分散，产生了许多“数据孤岛”，成为了制约人工智能技术发展的关键障碍。尽管联邦学习的兴趣日益增加，但在实际应用中仍然面临着不少挑战。虽然各地区的学习计划通过传统模型设定来保障用户培训数据的机密性，但这些模型依旧存在泄露敏感信息的隐患。此外，目前的隐私保护机制也存在一定的安全漏洞。例如，依赖统一单密钥加密的安全聚合系统常常容易受到恶意攻击。值得注意的是，模型训练的有效性不仅依赖于大数据集的规模，数据质量对模型的整体性能同样至关重要。尽管已有大量研究对数据质量进行探讨，但在数据安全评估机制建设上仍显不足。因此，提升数据质量和安全性对于成功实施联邦学习变得尤为关键。

关键词：联邦学习；隐私保护；安全聚合；同态加密；数据评估

前言

现代深度学习的快速进步与大数据的普遍应用紧密相连。然而，随着对数据需求的持续增加，获取免费数据的机会越来越少。一方面，数据供应商开始意识到其数据的商业价值；另一方面，由于行业内部竞争和利益冲突，各方在数据发布方面变得愈发谨慎。此外，随着大数据时代的到来，信息泄露和公民隐私保护的问题也日益突出。许多国家的政府陆续推出相关法律法规，以加强对个人数据收集的监控。例如，自2018年施行的《通用数据保护条例》（GDPR）在数据隐私管理方面带来了显著变革。为了解决用户隐私与数据利用潜力之间的矛盾，联邦学习（FL）近年来逐渐引起了学术界和业界的广泛关注。这种联合学习的方式使多个用户能够在不依赖中央服务器的情况下共同开发公共模型，从而有效地防止用户私人数据的泄露。因此，联邦学习被视为安全敏感任务管理的关键技术，应用于用户偏好预测、自动驾驶和医疗预测等领域。尽管联邦学习在数据隐私保护方面展现了独特的优势，但它仍面临一些挑战和局限性。许多传统聚合系统仍然存在用户隐私的风险。这些系统往往将模型参数直接上传和存储在中央服务器，便于聚合，这样任何合法用户都可以以明文形式获取这些参数。此过程不仅无法完全避免用户数据泄露的风险，还在一定程度上危及了用户数据的保密性。

基金项目：杭州市人工智能重大科技创新项目（项目编号：2022AIZD0058）

一、联邦学习概述

联邦学习是自2016年由谷歌的H. Brendan McMahan及其团队首次提出的一种先进机器学习技术。这项创新允许分散在大量用户（如手机、移动设备或传感器）上的敏感数据有效地被利用，而无需直接收集用户数据。在联邦学习的实施过程中，用户在每个迭代周期下载本地模型参数以形成全局模型。一个集中式服务器负责汇总各用户发送的参数，并将更新后的全局模型反馈给所有用户。由于联邦学习的设计能够有效防止私密数据的泄露，因此它被视为处理安全敏感任务的重要方法之一。开发实用的智能模型，例如疾病预警或交通预测模型，通常需要大量的数据支持。然而，随着全球隐私法规的日益严格，公众对个人隐私的关注也在不断升高，致使获取高质量和大规模数据的难度加大。在这种情况下，联邦学习应运而生，作为一种新兴的训练模式，有望通过有效整合不同来源的数据来解决这一难题，创造更大的价值。联邦学习框架由多个用户设备和云服务组成。原始数据在用户的本地终端上被处理和标准化，以形成本地模型。随后，模型训练的结果（例如部分模型参数或梯度）会被上传至服务器进行参数聚合，最后更新的模型则会传递回每位用户。与传统的集中式机器学习方

法相比，联邦学习能够在不收集用户原始数据的前提下进行训练。这一特点显著降低了数据泄露风险及其相关成本，有效打破了数据孤岛，促进了更为广泛的数据共享与应用。

二、隐私保护和可靠的联邦学习方案

为了应对当前数据保密性和可靠性所面临的挑战，尤其是针对方案的具体难题，我们提出了一种新颖的解决方案——安全可靠的联邦学习系统（PPRFLS）。该系统采用双重服务器架构，由聚合服务器（AS）和平台服务器（PS）构成。在每个学习周期中，参与者将CKKS模型的加密更新发送至聚合服务器。聚合服务器在平台服务器的支持下，通过光学方式汇总参与者的加密模型更新，并利用余弦相似性进行评估。

借助CKKS的同态加法特性，聚合服务器能够高效整合来自不同组的加密模型。当聚合结果传送至平台服务器后，后者负责解密生成的全局子模型，并对其在本地测试数据集上的表现进行验证。如果全局子模型未达到预定的精度标准，将会被直接排除。随后，系统根据质量加权聚合算法对剩余的可靠全局子模型进行进一步整合，从而生成更新后的全球周期模型。

该程序全面满足安全性、可靠性和完整性的标准，有效保障了系统的安全与隐私。在PPRFLS中，参与者仅需向聚合服务器提交加密后的模型更新，得益于CKKS所提供的高安全性，这些模型更新中的明文信息对任何第三方以及外部攻击者来说都是完全不可见的。因此，PPRFLS在数据隐私保护方面展现了显著的优势。

在聚合过程中，PPRFLS仅引入经过验证的可靠子模型，从而确保生成的全局模型更加可信。同时，在更新聚合时，PPRFLS也充分考虑了全局子模型的准确性，以进一步增强隐私保护和安全性。

模拟结果表明，PPRFLS在显著提高全局模型准确率的同时，还降低了潜在攻击者成功实施攻击的概率，从而确保了全局模型的可靠性。此外，其隐私保护能力与联邦学习的基准方案FedAvg不相上下，为数据保护和模型安全开辟了新的可能性。

三、隐私保护、可靠与公平的联邦学习方案

为了应对参与者不愿意无条件地提供自己的计算资源用于联邦学习的问题，我们设定了相应的设计目标，以保障联邦学习中的公平激励机制。具体如下：

在可靠的隐私保护联邦学习方案中，参与者通常被视为愿意投入计算能力和数据。然而，在实际操作中，他们并不总是愿意无条件地贡献自己的资源。针对这一现象，我们在“可靠与公平的联邦学习系统”（PPRFFLS）基础上，开发了一种公平的激励机制。该机制构建了一个基于深度Q网络（DQN）的激励系统，使用隐私评估机制的结果，为参与者集体提供公正的回报，同时不侵犯本地隐私信息。该方案采用了深度卷积神经网络（CNN），以压缩学习状态的表示，并且计算每个回报的Q值。通过激励参与者提供高质量的模型更新，这种激励机制提升了参数服务器（PS）的功能性。对于参与者而言，模型更新的质量和所获得的激励保持了一致，确保了动机上的公平性。

PPRFFLS在保护数据公正性和高度机密性的同时，也重视参与者的隐私和可信度。该提案通过设计基于DQN的激励系统，强调更新是基于模型质量而非个人信息（例如计算资源），从而维护信息的保密性，确保来自其他参与者的本地模型更新不被泄露。此外，通过隐私评估机制实施的分组支付策略，增强了激励的公平性。实验结果表明，PPRFFLS能够有效地激励参与者下载高质量的模型更新，从而提升参数服务器的整体效能。

四、隐私保护、可靠和公平的联邦学习方案

（一）方案概述

在联邦学习及隐私保护的框架下，传统的观点认为参与者应该毫无保留地分享他们的模型更新。然而，在实际操作中，参与者往往只有在获得合理补偿的前提下，才愿意下载质量可靠的模型更新，这就体现了迫切需要一个公正的激励机制。目前，现有的激励措施主要关注与参与者隐私相关的本地信息（如成本和数据量等），却忽视了参与者对隐私保护的需求及其自身的利益。为了解决这个问题，我们提出了一种创新性的方案：一个安全、可靠且公正的联邦学习框架。该框架使用深度Q网络（DQN）来设计公平激励，同时借助深度卷积神经网络（CNN）有效地压缩学习状态空间，并准确预测每次支付对应的Q值。我们确保每位参与者都能获得合理的收益，而不需要依赖本地的隐私信息，这样有效降低了因欺诈性协议造成的利益冲突风险，从根本上保障了激励的公正性。在实际应用中，我们从强化学习的基础要素入手，包括状态、动作和奖励等。接下来，我们将详细介绍基于DQN的激励算法的具体实施步骤。我们的目标是确保参与者能够获得合理的激励，同时有效地保护隐私，推动整个联邦学习系统的有效性和可信赖性。

（二）具体实施

许多联邦学徒程序通常设想一个理想的情境，即参与者能够在全球范围内，无条件地及时更新培训模式。然而，实际情况却面临不少挑战。参与联邦学习任务常常需要占用参与者的资源，比如计算和通信能力，因此他们只有在获得足够回报的情况下才愿意下载可靠的模型更新。此外，一些恶意参与者可能会利用虚假的本地隐私信息（例如成本或数据规模）来骗取丰厚奖励。同时，为了保护隐私，一些参与者可能会选择不下载与个人数据相关的信息。我们所提出的LPPRFLS模型假设参与者会无条件贡献他们的更新，但这一假设在现实中往往难以实现。因此，对于服务提供者而言，设计合理的激励机制显得尤为重要，这不仅能够激励参与者下载可信赖的模型更新，同时也能有效控制成本

（三）结果分析

1.强数据隐私性分析

通过使用CKKS方案，参与者的支付数据和参与水平被加密处理，从而确保即使在受到外部攻击或数据泄露的情况下，敏感信息仍然处于安全状态。PPRFFLS机制的设计还充分考虑到系统的公平性与透明性。由于激励是基于历史观察信息而非个人隐私信息，系统能够更加公正地对待所有参与者，确保他们根据自身的贡献获得相应的激励。这种方式不仅鼓励了参与者积极参与，还提升了整体的参与度和活跃度，从而推动了整个生态系统的健康发展。PPRFFLS不仅在实现隐私保护方面表现突出，还在性能上达到了新的高度。通过高效的加密算法和灵活的激励机制，系统能够处理大量参与者的实时数据，不仅保证了响应速度，也增强了系统的可扩展性。无论是在金融交易、物联网应用还是其他需要隐私保护的场景中，PPRFFLS都展示出了良好的适用性。

2.激励公平性分析

PS为不同的群体量身定制了支付策略，根据其模型的质量对参与者进行分类。为了避免参与者通过获取虚假信息来进行欺诈并从中获得丰厚回报，系统的设计保证了严谨性。不同组别的模型更新质量各有不同，因此各组获得的支付也有所差异，这确保了激励机制的公正性。接下来，我们进行了实验，以验证PPRFFLS是否能够有效激励参与者下载可信的模板更新，同时进一步提升PS的整体效能。

3.模型无损性

为了检验模型的无损性，我们对一组数据进行了对比，分析了PPRFLS、PPRFFLS和FedAvg这三种协议。在本次实验中，非IID强度被设定为0，且未实施任何恶意攻击。PPRFLS协议展示了更快的收敛速度，准确率在五轮内就达到了90%以上，而PPRFFLS虽稍慢，但同样在第六轮结束时达到了接近91%的准确率。相对而言，FedAvg协议在收敛速度和最终准确率上表现平平，大约需要十轮才能超过85%的准确率。PPRFLS和PPRFFLS协议的优势在于有效地利用了参与方之间的梯度信息，降低了模型更新的redundant（冗余）部分。通过引入个性化模型更新策略，这两种协议能够更好地适应不同参与方的局部数据特征，从而提升了整体学习效果。

结论

尽管联邦学习在解决“数据孤岛”问题方面显示出显著的优势，但它仍面临一些挑战。例如，在模型更新的过程中，参与者可能会遇到不确定性，并且可能不愿意无偿提供计算资源。为了解决模型更新的可靠性问题，我们创新性地开发了一套隐私评估机制，以确保联邦学习项目的持续有效性和安全性。通过明确联邦学习的设计目标，专注于保护参与者的隐私，并评估模型更新的质量，以便挑选出最值得信赖的集成模型。此外，针对参与者不愿意无条件分享信息资源的现象，我们设计了一种基于强化学习的PPRFLS改进方案，旨在为学员提供既公正又可靠的学习体验，同时确保他们的隐私得到充分保护。

参考文献

[1]梁琳.文本数据隐私泄露风险评估方法研究[D].中南财经政法大学，2021.

[2]吴恙.跨境数据流通中个人信息保护法律规制研究[D].中南财经政法大学，2021.

[3]张颖，袁海，张继东.智慧家庭隐私泄露风险评估方法及系统[J].现代信息科技，2021，5（01）：143-145.

[4]孙方江.跨境数据流动：数字经济下的全球博弈与中国选择[J].西南金融，2021，（01）：3-13.

[5]刘姿杉，程强，吕博.面向机器学习的隐私保护关键技术研究综述[J].电信科学，2020，36（11）：18-27.