基于同态加密的金融数据安全联邦学习

莎莉尼·Dhiman
计算机科学系;工程国家理工学院
**，印度
shalini.phd22@nitap.ac.i
n 0000-0003-1384-7238

Sumitra Nayak
计算机科学系;工程国家理工学院
**，印度
sumitra.mtech.cse.21@nitap.ac.i
n 0000-0002-1910-6337

Ganesh Kumar Mahato
计算机科学系;工程国家理工学院
**，印度
ganesh.phd20@nitap.ac.i
n 0000-0002-2925-7438

Anil Ram
计算机科学系;工程国家理工学院
**，印度
anil.phd20@nitap.ac.in
0000-0002-3581-0519

Swarnendu Kumar Chakraborty
计算机科学系;工程国家理工学院
**，印度
swarnendu@nitap.ac.in
0000-0002-6223-9362

2023 第四届计算与通信系统国际会议(I3CS) | 979-8-3503-2377-1/23/$31.00 ©2023 IEEE | DOI: 10.1109 / I3CS58314.2023.10127502

联邦学习是一种分布式机器学习技术，它可以实现设备上的训练，而无需在集中式服务器上交换敏感数据。在本文中，联邦学习被用于在金融或业务系统中的物联网应用的帮助下训练设备上的金融数据模型。这创造了更先进、更安全的金融商业模式。我们应用了同态加密和加密原语的机制，包括屏蔽和本地模型保护，以防止私人金融或商业数据中的任何类型的推断，而多个攻击者通常会找到模型反转或重建攻击的方法。我们使用各个金融部门的数据集作为主要度量，而不是采用深度学习中通常使用的数据集的大小，以获得正确度量本地模型训练的每一节对全局模型训练的贡献率。如果计算出的在线客户端数量超过预定的阈值，那么联邦学习过程将继续使用容错计划。安全研究表明，建议的解决方案满足数据隐私要求。计算成本和通信成本也从理论上进行了分析。根据研究观察，与现有方案相比，该方法在保证隐私保护的同时取得了良好的效果。

索引术语:同态加密、联邦学习、隐私保护

I. 介绍

物联网(IoT)是一个有益的概念，有许多不同的应用。通过物联网金融业务，投资者或市场分析师可以获取金融信息从任何位置实时获取用户的业务数据，实现智能交易和持续监控。此外，物联网还提供计算机辅助财务管理、财务分析和金融危机预测服务。这可以使用户获得更好的结果并使金融运营商更有效地工作。使用基于物联网的业务模块和财务分析技术有几个好处，可以提高效率[1]。共享电子金融数据可以帮助市场分析人员预测投资计划的波动特别是在金融部门。共享财务报告有助于整个行业的发展然而，由于投资报告的敏感性，现在很难共享资源。大多数市场分析师分布在不同的地区，在不同的行政管理下工作[2]。由于担心违反隐私法或损失财务利益的风险，例如暴露业务或市场策略或敏感的金融投资，市场分析师对披露投资者或行业敏感的财务数据犹豫不决。这些隐私问题同样是其他行业需要解决的重大挑战。

[3]和[4]中的研究人员开发了一个联邦学习框架，该框架使来自本地存储库的数据能够为机器学习做出贡献，同时保护隐私。其目的是在不损害用户隐私的情况下实现数据共享的好处。通过人工智能的应用，联邦学习具有改善多个领域的潜力

979-8-3503-2377-1/23/$31.00 ©2023 ieee

各公共部门的情报。然而，它也可能对用户隐私构成威胁在现实世界中实现联邦学习，比如基于物联网的金融业务是一个重大挑战。虽然本地数据不直接共享，但本地模型被聚合到一个集线器中，使它们容易受到攻击。

由于这些担忧，人们越来越关注实现保护用户隐私的联邦学习过程。然而，联邦学习的应用程序可能会带来各种技术挑战，例如确保准确的全局模型、管理客户退出以及最小化计算和通信成本。

为了解决这些问题，本研究提出了一种隐私保护系统，用于物联网工业应用的几个金融部门的联邦学习，重点是金融数据安全。联邦学习体系结构使用多方计算实现，这已经是一种安全的格式，并使用同态加密来维护数据隐私该系统还能够处理多个参与者之间的串通。

此外，我们已透过改善现有计划，解决成本问题。此外，我们考虑了不同市场领域数据的不同特征，优先考虑了模型的准确性，同时提高了安全流程的效率。

II. 背景

A. 联合学习

联邦学习是一种分布式协作 AI 解决方案，它有助于训练数据集，数据的训练是通过多个设备与中央服务器的协调来完成的，实际数据集的使用并不多，例如，在智能物联网网络中，多个物联网设备作为工作人员与服务器连接服务器也可以被聚合器知道[3]。服务器首先从一个全局模型开始，其中包含一些更精确的学习参数。每个工作人员从聚合器下载当前模型，然后使用本地数据集更新模型并将本地计算的更新卸载到聚合器。最后，服务器通过使用所有的本地模型来创建一个新的全局模型。服务器可以利用分布式工作者的处理能力来提高培训质量，从而减少分布式工作者的隐私泄露[4]。最后，来自聚合器的本地工作者下载所有全局更新，并为下一次本地更新做好准备直到培训完成。

联邦学习主要有三种类型，下面将进行讨论。

- 水平联邦学习(HFL): 当两个数据集之间存在有限的用户重叠但显著的用户特征重叠时，HFL 是一种适合的方法。在 HFL 中，数据集按用户维度水平划分，删除具有相同用户属性但不同用户的数据进行训练。换句话说，相同的数据特征出现在不同的行中

对传统的联邦学习逻辑回归算法。作者还实现了更快的训练时间和更好的数据安全性。

在[6]，[7]，[17]中，作者指出，信任问题出现在

联邦机器学习是指多方参与分布式数据集的联合数据分析。为了解决这个问题

为了解决这个问题，他们提出了一个叫做梯度增强的方案机器(GBM)是一种基于半同态加密的多方计算模型。在这种情况下，每一方都可以获得自己的共享梯度提升模型

机器，这使他们能够保护他们的数据

剥削。作者大多在“两党制”下工作。

双方都有独特的视图和特性的场景

标签共享是不允许的。为了达到这个目的，他们[18]，[19] 介绍了 LightGBM。用于安全通信和

计算时，他们引入了 SecureGBM。 结果，

按用户特征对齐的数据。因此，联邦水平学习可以扩大用户样本。

- 垂直联合学习(VFL): VFL 是两个给定数据集的用户特征之间几乎没有重叠的情况，但在用户中有明显的重叠。在 VFL 中，数据集垂直划分(按用户特征维度)，去除用户相同但用户属性不同的数据进行训练。换句话说，相同的用户出现在跨列的数据中(按用户对齐)。因此，使用 VFL 可以增加训练数据的特征维数。
- 联邦迁移学习(FTL): FTL 是一种学习方案，它涉及将知识从具有丰富特征空间的数据集转移到具有有限特征或标签的数据集，以训练高性能模型。这在处理来自不同方面具有不同级别特征的多个数据集时特别有用。

B. 同态加密

同态加密是一种特殊类型的数学运算，它对加密后的数据进行加密，而不保证加密。直接对明文进行计算得到的结果与对密文进行计算得到的结果基本相当[7][20]。在这里，执行计算的服务提供者不知道纯文本，因为他们在加密状态下执行计算。因此，明文的安全性得到了保证。

III. 相关工作

[4]中的作者遇到了一个网络安全挑战，其中数据聚合变得困难。为了应对这一挑战，作者使用了智能电网的概念，它为增强操作系统和改善电力的分配、发电和传输提供了解决方案。然而，智能电网也面临安全和隐私问题，并产生高昂的计算和通信成本。为了克服这些问题，作者提出了一种保护隐私的基于同态加密的深度学习和数据聚合模型。该模型通过使用不同的身份验证技术执行智能计量，减少了对闪现工作负载预测模型准确性的负面影响。为了增强安全性，作者设计了一个基于云计算和区块链的数据聚合框架。结果表明，该模型能达到 80

根据[5]，传统的执行联邦学习任务的方法会降低训练速度、准确性和数据安全性。在使用垂直联邦学习进行任务分类时，传统方法会产生较大的无效参数。为了克服这一问题，作者提出了使用梯度压缩和参数共享来实现双方通信，并在参数传输上改进同构加密。通过比较，得出了较好的分类模型评价指标

算法 1 讨论了具有安全性的模型训练，算法 2 提出了具有数据安全性的多层感知器联邦学习。

具有安全性的训练算法 1: 输入: ΣWt1Σ，
ΣWt2Σ，。。。， ΣWtNc Σ

2：CSP 生成 Ni 个随机向量 R1 R2。。。，RNj
然后用公钥加密。3:对于 j≤Ni doΣ Σ

4：CSP (Sk) $W_j^{k_j^l} \leftarrow \cdot(Rj)$

5：CSP 部分解密[Si]i 使用 SK (j)
6: end for

7：CSP 将向量以解密形式发送给 CP

作者实现了 3 到 64 倍的性能下降，但由于 SecureGBM 的使用，效率得到了提高，导致了更低的下降率和更大的训练集。表 1 将我们提出的模型与以前完成的类似工作进行了比较。

IV. 系统模型

A. 模型设计

本节讨论金融环境中的 FL 系统，该系统由分布式客户端和模型聚合服务器组成，如图 1 所示。该系统采用安全的多方计算和同态加密，即使在真诚但好奇的参与者之间存在串通的情况下，也能确保数据隐私。该研究提出了一种改进的系统，解决了影响以前基于加密的方案的成本问题，同时提高了不同金融行业数据属性的模型准确性和性能结果。

为了减少通信开销，通过调整算法，将高级 Paillier 加密技术的加性同态改为乘性同态。此外，与传统加密不同联邦学习中的同态加密不加密每个模型参数。高维数据的同态加密，如在深度学习模型中所见，将导致巨大的计算成本。本文提出的方案仅对每个成就计划中每个客户端的数据质量变量进行加密。这种方法防止了计算开销的急剧增加。

所提出的模型基于服务器和客户端的相互协作。 本系统的客户端基本上是金融机构，它由大量的金融原始数据组成。客户端在本地金融数据集上训练本地模型，并将其提交给服务器。客户端提交所有被屏蔽的金融本地模型，一旦被屏蔽模型被提交，服务器将收集所有被屏蔽的金融本地模型，并执行一系列操作来执行安全聚合。在执行此过程时，需要执行一些隐私保护步骤，这些步骤可以在给定的算法中看到

算法 2 数据安全的多层感知器联合学习

输入: 金融数据集 Dx

输出: 最终 *模型* $_{\phi}$

1：初始化模型 ϕ
2: for I 迭代做
3： 向前传播:输出 j = *hq (xj, j);*
4： 损失计算: *Cj* = loss (h * $^{(xj),}$ out $_{tj);}$
5:如果 *cj* <ϵ *然*后
6： 打破
7:其他
8： 传播: Grad j = *bp* (xj， $_{\theta j,\ cj);}$
9： 向执行计算的服务器发送梯度;10：
升级: *;;;;;;;;;* −
11:结束 if
12: end for
13:返回参数为 ϕfinal 的模型

B. 实验环境

该算法是用 Python 开发的，并在 3.6 GHz 四核处理器和 16gb RAM 的机器上进行了评估。利用联邦学习仿真平台 FLUTE 进行实验。在该模型中训练和测试了近 76000 个数据集。

C. 数据集

在这种情况下，来自金融部门的数据用于训练模型并对加密状态执行同态计算。使用了 76000 多个不同参数的财务数据。 数据收集自[22]，其中包括股票市场、银行交易、贷款部门等的数据。

D. 安全分析

数据的计算是以加密形式进行的，云服务提供商和计算服务器不知道纯文本。云服务提供商只有在对敏感数据进行加密后才会将数据发送到计算服务器，这样就没有机会让数据落入入侵者手中。

表我
类似方案的比较

Ref。	计划使用		数据聚合	服务器	加密的数据访问		体系结构	同态加密方法
[4]	区块链		是的	云	是的部分		集中	基于数据聚合的研究
[5]	区块链		是的	零	是的完全		集中	阈值秘密共享
[6]	零		是的	中央服务器	是的零		集中	数据聚合
[7]	机器学习		是 fl 服务器是无分散可信执行环境					
[8]	机器学习							
[9]	云计算		是 fl 服务器是无去中心化去中心化阈值可加性 Homo- 否无是是集中式 fl					形态学加密
[10]	机器学习		没有	零	是的	没有	没有一个	算法的同态加密 大致的数字
[11]	机	学习----	是的	中央服务器	是的	是的	集中	加性同态加密
[12]	零		没有	云服务器	是的	没有	分散的	边缘层节点的秘密共享
[13]	机	学习----	是的	中央服务器	是的	是的	分散的	加性同态加密
[14]	人工	易达利-	没有	主机/客户服务器	是的	没有	没有一个	数据聚合
[15]	机	学习----	是的	云服务器	是的	是的	集中	Fashion-MNIST
[16]	机	学习----	没有	零	是的	没有	集中	梯度增强机
我们的	联邦	学习----	是的	云	是的	没有	集中	相加和相乘

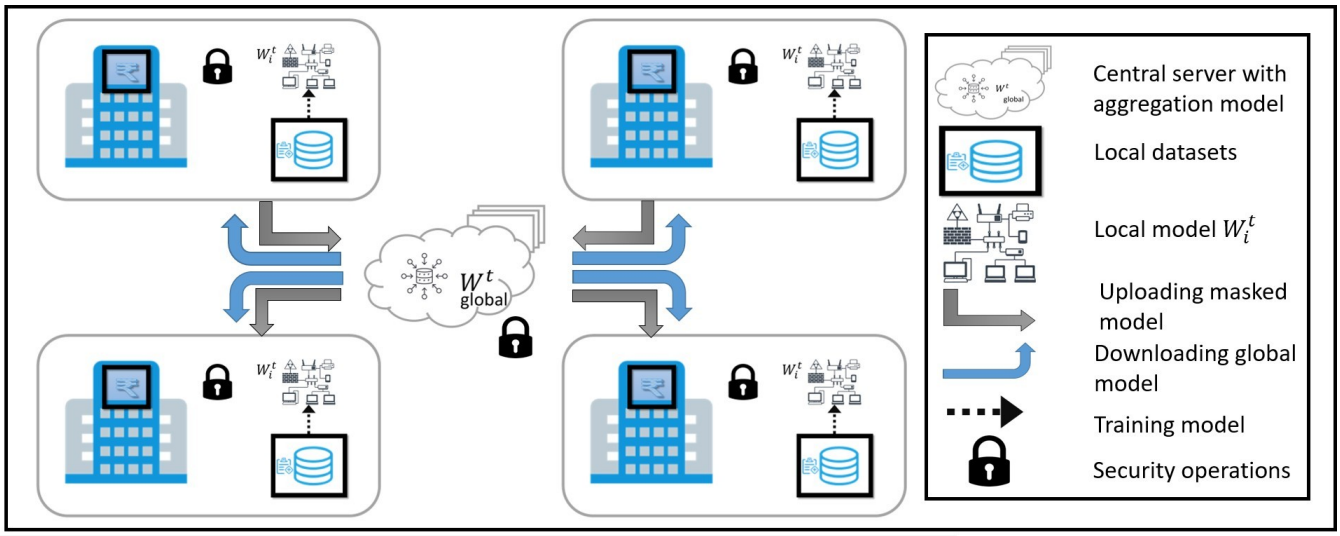

图 1 所示。基于同态加密的安全金融数据联邦学习。

E. 绩效评估

所建议模型的性能列于表二和表三。前者描述了不同键大小的各种迭代所消耗的时间，而后者显示了在不同层训练模型所花费的时间。

加密后的数据使用 128 ~ 2048 位的密钥进行加密。随着密钥大小的增加，时间消耗也随之增加。同时，随着隐藏层的增加，训练时间也会增加。如图 2 所示。显示所提出的模型针对不同的密钥大小进行密钥生成、加密、计算和解密所花费的时间。图 3 所示。显示使用不同密钥长度的加密数据在不同迭代中训练模型所花费的时间。

结论

在我们提出的论文中，我们提出了一个独特的联邦学习框架，该框架专为基于物联网的金融营销系统设计，优先考虑保护隐私。我们共同

表二世
不同密钥大小的各种迭代的时间消耗(秒)

算法	密钥大小(位)	时间(秒)
先进 Paillier	128	547.65
	256	642.7
	512	1112.25
	1024	2025.27
	2048	3768.56

表 3
用于训练的各个隐藏层的时间消耗(秒)

使用数据集	隐藏层(单位)	时间(秒)
财务数据	3.25	2432.54
	3.65	3642.7
	4.23	4112.25
	4.54	5025.27
	5.25	8768.56

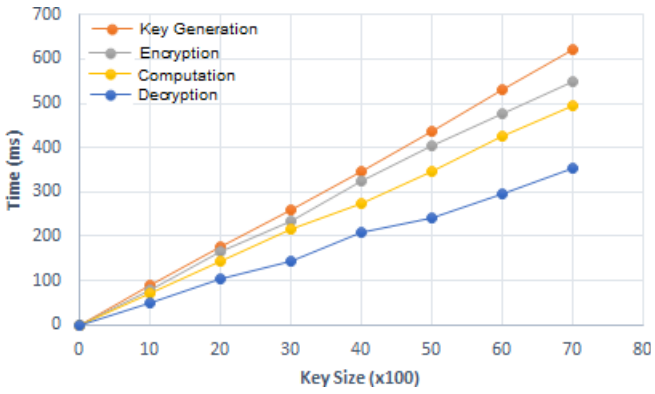

图 2 所示。花在同态密钥生成、加密、计算和解密上的时间

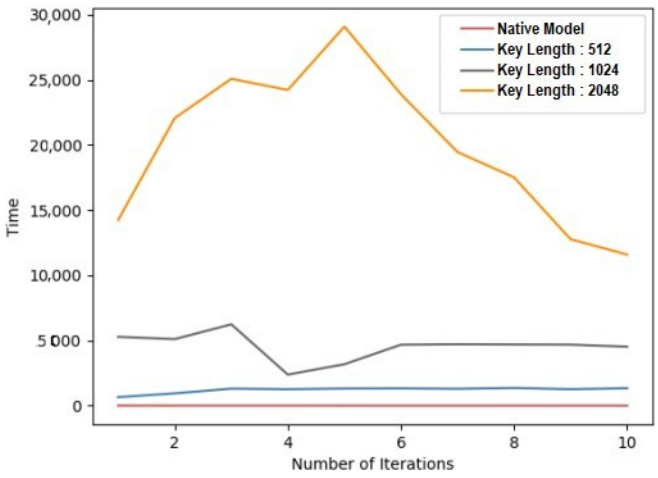

图 3 所示。在不同键大小的财务数据集上花费的时间

结论是，这个框架可能会在保持数据隐私的同时实现更有效的数据利用，从而彻底改变当前的系统。为了实现这个目标，我们进行了修改

传统系统技术[21]中使用的计算以前是基于数据量的。相反，我们使用了一种基于数据质量的加权平均算法，我们发现这种算法更有效。在联邦学习的上下文中，我们还介绍了一种全新的屏蔽方法，该方法同时利用了安全的多方计算和同态加密。这种方法确保来自多个设备的数据被安全有效地利用，而不会损害用户隐私。然而，我们意识到对这种通常与深度学习模型相关的高维数据进行同态加密将导致显著的计算开销。因此，我们设计的同态加密模型不单独加密每个模型参数，以减少计算成本，同时保持隐私。

我们提出的技术旨在通过以固定形式为每个客户端仅加密数据质量变量来最小化与同态加密相关的计算开销。这种方法防止了计算开销的显著增加，使其在实践中更加可行。为此，我们对 Paillier 加密技术进行了改进，将其同态乘法的性质转化为可加同态，使其适用于我们所提出的系统。

总的来说，我们认为我们提出的技术提供了一个有前途的解决方案，可以解决与同态加密相关的计算开销问题，同时保持合理的成本。通过仅加密必要的变量并利用有效的加密技术，我们能够为基于物联网的金融营销系统中的联邦学习提供实用且安全的解决方案。

参考文献

[1] Zhang, Li, J. Xu, P. Vijayakumar, P. K. Sharma, U. Ghosh，“基于同态加密的医疗保健系统中的隐私保护联邦学习”，IEEE Transl，网络科学与工程，2022。

[2] Kim, Andrey, A. Papadimitriou 和 Y. Polyakov，“近似同形加密与减少近似误差”，Springer, Cham, RSA 会议，密码学家轨道，pp. 120- 144,2022。

[3] Kara, Mostefa, A. Laouid, M. A. Yagoub, R. Euler, S. Medileh, M. Ham- moudeh, A. Eleyan 和 A. Bounceur，“基于魔术数分片和 El-Gamal 加密的完全同态加密：智能医疗用例，“专家系统 39”，no。5, p. 12767, 2022。
[4] P.辛格，M.马苏德，M.S.王晓明，“基于区块链和同态加密的智能电网数据聚合模型”，《计算机工程与电子技术》，第 3 期，第 1 - 4 页。
[5] R. Kumar, J. Kumar, A.A. 汗，H.阿里，c.m.伯纳德，R.U. 张晓明，“基于区块链和同态加密的医学图像隐私保护模型聚合”，《医学信息学报》，第 2 期，p. 1023 - 1023 页，2012。
[6] 蒋志峰，王伟，刘毅，“闪光: 一种跨孤岛联合学习的加性对称半纯态加密算法，[j] .中文信息学报，2004,21(9):675 - 675,2021。
[7] Mondal, Arup, Y. More, R. H. Rooparaghunath 和 D. Gupta。“Flatee: 跨可信执行环境的联邦学习”，arXiv 预印本 arXiv, 2111.06867(2021)。
[8] 张锋，张海波，张海波，李斌，“基于分散阈值加性同态加密的安全线性聚合联合学习”，中国科学 d 卷第 1 期，第 1 期，第 3 期。
[9] 李俊毅，黄辉，“基于分布式同态加密的快速安全数据挖掘”[26]。知识发现&;数据挖掘，pp. 2706-2714, 2020]。
[10] 李登浩，王俊，孔丽丽，司树豪，黄志强，黄春春，等肖杰，“金融领域垂直联邦学习的最近邻欠采样策略”，《信息隐藏与多媒体安全研讨会文集》，第 123-128 页，2022 年。
[11] Madi, Abbass, O. Stan, A. Mayoue, A. G. Se´bert, C. G. Pailler, and R. Sirdey，“使用同态加密和可验证计算的安全联邦学习框架”，IEEE。协调数据分析、自动化、隐私和安全: 大数据挑战，2021 (RDAAPS)， pp. 1-8, 2021。
[12] 李志强，李志强，李志强，“基于多模态加密的 IoMT 隐私保护”，应用科学 11，第 1 期。18(2021)， 8757, 2021。
[13] 张成良，李生，夏军，王伟，闫峰，刘毅，“批加密”; “跨竖井联合学习的高效同态加密”，USENIX 年度技术会议。(USENIX ATC 20)， pp. 493-506, 2020。
[14] 杨奎和，宋正杰，张勇，周勇，孙晓霞，王俊，“基于垂直联邦学习的模型优化方法”，IEEE 国际电路与系统研讨会，2021,pp 1-5。IEEE 2021。
[15] 周春义，付安，于生，杨伟，王慧，张勇，“雾计算中保护隐私的联邦学习”，《IEEE 物联网学报》第 7 期。11 (2020): 10782 - 10793。
[16] 冯志，熊宏辉，宋超，杨树林，赵波，王丽丽，陈振中，杨树林，刘林，欢杰，“安全模型”; 安全多方梯度增强，”2019 年 IEEE 国际会议。《大数据》，第 1312-1321 页，2019。
[17] Mammen, P. Mary，“联邦学习:机遇与挑战”，《中国管理杂志》，2010 年第 4 期，2101.05428 2021。
[18] 李，田，A. K. Sahu, A. Talwalkar 和 V. Smith，“联邦学习: 挑战、方法和未来方向，《IEEE 信号处理杂志》第 37 期。3, pp. 50-60, 2020。
[19] 张勇，谢勇，李伟，于斌，“联合学习研究”，《知识管理系统》第 16 卷第 1 期，2013。
[20] Mahato, g.k.， Chakraborty, s.k.: 云安全中同构加密的比较研究。IETE 研究杂志。 泰勒和弗朗西斯，第 1-10 页，2021 年。
[21] Mahato, g.k.， Chakraborty, s.k.: 基于同态加密的边缘计算隐私保护。模式识别和数据分析与应用(第 395-407 页)。施普林格、新加坡。(2022)
[22] https://www.kaggle.com/datasets/theworldbank/global-financial-的发展。