联邦学习在金融数据安全领域的研究与应用*

张海涛
(五矿国际信托有限公司,北京 100027)

摘 要:近年来,金融领域明文数据流通所引起的数据泄露问题日渐突出,传统的跨机构数据融合的机器学习方式面临着新的问题与挑战。因此,立足于金融数据安全领域,从用户隐私和数据安全角度出发,概述联邦学习理论并深入分析其目前在金融行业的应用现状,指出现有的联邦学习还存在通信效率低、数据异构性突出等问题。最后提出健全联邦学习标准体系、时刻关注监管要求等建议,为推动联邦学习在金融数据安全领域中的合法应用提供参考性意见。

关键词:联邦学习;金融数据安全;数据隐私;信用卡欺诈
中图分类号:TP391 文献标识码:A DOI:10.19358/j.issn.2096-5133.2022.01.001
引用格式:张海涛. 联邦学习在金融数据安全领域的研究与应用[J].信息技术与网络安全,2022,41(1):3-9.

Research and application of federated learning in the field of financial data security

Zhang Haitao
(Minmetals International Trust Co., Ltd., Beijing 100027, China)

Abstract: In recent years, the problem of data leakage caused by the circulation of plaintext data in the financial field has become increasingly prominent. The traditional machine learning method of inter agency data fusion faces new problems and challenges. Therefore, based on the field of financial data security, from the perspective of user privacy and data security, this paper summarizes the federated learning theory, deeply analyzes its current application status in the financial industry, and points out that the existing federated learning still has some problems, such as low communication efficiency and prominent data heterogeneity. Finally, it puts forward suggestions on improving the federated learning standard system and paying attention to regulatory requirements at all times, so as to provide reference opinions for promoting the legal application of federated learning in the field of financial data security.

Key words: federated learning; financial data security; data privacy; credit card fraud

0 引言

2020 年 4 月,中共中央、国务院印发了《关于构建更加完善的要素市场化配置体制机制的意见》,明确指出在当今数字经济化时代,数据是至关重要的一种新型生产要素。但是,随着数据赋能研究的不断深入,隐私保护和数据泄露等问题日益突出。如 2018 年 3 月,超 5 000 万 Facebook 用户信息被政治数据公司“剑桥分析”获取并利用,2018 年 11 月,汇丰银行(HSBC Bank)部分客户财务状况和个人信息被泄露。金融作为数据密集型行业,对数据安全、隐私保护以及监管科技等有着更高的要求。实现数据的多方协同和授权共享,得到更优的模型和决策,是当前人工智能赋能金融科技的一个重大挑战[1]。Google 于 2016 年提出联邦学习(Federated Learning)概念为这一困境带来了新的思路与解决办法。目前,联邦学习技术已经在金融科技领域的智能营销、反欺诈、信用卡评分、产品推荐等多个业务场景中得到了具体应用。

1 联邦学习

1.1 联邦学习概述

“联邦学习”,顾名思义,指虚拟地建立一个“联邦世界”,把不同机构的“数据孤岛”像独立的“国

*基金项目:国家高端智库课题“金融科技监管专项方案”(ZXZK202102)

家”一样统一联合起来，彼此间既可以保持一定的独立性(包括机密性、隐私性)，又可以在不共享数据的情况下联合建模，共同受益。联邦学习属于分布式机器学习技术，其明显的特征为“数据不动，模型动”，通过各联邦学习参与方和服务器之间的“参数传导–梯度聚合–模型更新”过程，使得各参与方协同训练，最终聚合出一个拥有更多数据集的可共享的联邦学习模型[2]。

根据联邦学习各参与方数据源分布方式的不同，可将联邦学习分为横向联邦学习(Horizontal Federated Learning，HFL)、纵向联邦学习(Vertical Federated Learning，VFL)和迁移联邦学习(Federated Transfer Learning，FTL)[3]，如图 1 所示。其中，横向联邦学习主要适用于各参与方的数据集在数据特征维度上重叠较多，但在样本 ID 维度上重叠较小，甚至没有重叠的场景。在金融行业领域，不同地域间的多家银行虽然在用户群体上有较小交集，但由于在业务上的相近而拥有了相似的特征空间。在该场景下，多家银行就可以基于横向联邦学习进行联合信贷风控建模[4]。纵向联邦学习主要适用于各参与方的数据集在数据特征维度上仅有较小重叠甚至没有重叠，但在样本 ID 维度上有着较大重叠的现象[5]。仍然以银行业为例，同一地域的某家地方性银行和某家大型零售超市，由于同处一地，导致两家用户群体高度重合，但又由于各自经营业务的不同而导致特征维度交集很小，两家企业即可基于纵向联邦学习分别将不同的特征在加密状态下进行联合建模而实现共同受益。联邦迁移学习适用于多方数据集在数据特征维度和用户 ID 维度重叠度都较低的场景[6]，比如某家国内银行与国外的一家电商公司因业务场景的不同，导致两家机构累积的数据特征交集较小，同时又因地域的不同，两家机构用户群体交集也很小，即可基于联邦迁移技术帮助单一机构进行联合建模。

1.2 联邦学习作用机理

联邦学习由数据源、联邦学习系统、多方客户端三大要素构成。联邦学习的参与方分为客户端与服务端，为了保证训练过程数据的机密性与合规性，客户端负责本地模型的训练，服务端负责对客户端训练的本地模型进行联合构建，从而获得一个共享模型。其训练流程如下：

(1)加密样本对齐。系统使用基于加密的用户 ID 对齐技术，确保各个参与的客户端不需要暴露各自的原始数据即可对齐共同的实体用户，各方可以使用共有的实体数据协同训练机器学习模型。

(2) 服务端初始化联合训练模型。将初始参数下发到每一个客户端，各个客户端开始获得联合数据模型。

(3)客户端利用本地数据集和初始化参数进行模型训练，本地模型训练完成后，计算模型训练梯度，并使用加密或差异隐私等安全技术，将训练好模型参数上传至服务器。

(4)在服务器执行安全聚合操作。通过加权平均的方式对加密的模型参数进行更新计算，得到新的共享数据模型，安全聚合后的结果会再次发送给各个客户端。

(5)各参与方用解密后的梯度信息更新各自的本地模型，重复步骤(2)~步骤(4)，直至收敛，即可完成整个模型的联合训练。

在联邦学习的训练过程中，可以保证各参与方不会将信息泄露给服务器，服务器只负责安全聚合加密后的模型参数，然后发送至所有客户端进行参数的更新。服务器–客户端联合架构如图 2 所示。

2 联邦学习在金融数据安全领域的应用

目前，我国现有的征信体系尚无法满足金融机构在信用评估方面的需求，导致其只能依靠联邦学习等隐私技术来打通外部数据，进而满足信贷业务风控方面的建模需求。关于联邦学习在金融数据安

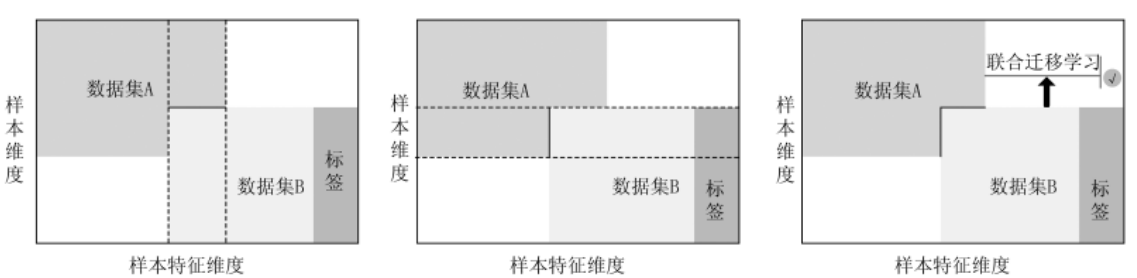

图 1 横向联邦学习(左)、纵向联邦学习(中)、联邦迁移学习(右)[1]

全领域中的应用研究，我国目前主要关注两方联合建模的场景，大多数的应用案例采取时序性联合建模方法。在实际的试点过程中，参与方通常是某金融机构和某科技公司以特定的业务场景来进行。部分国内外金融业联邦学习应用试点如表 1 所示。

目前银行业开展联邦学习的研究与应用主要集中在跨行反洗钱、信贷风控、理财产品营销等不同的业务场景，总体上纵向联邦学习的应用多于横向联邦学习应用。例如，在理财产品营销建模中，基于金融数据信息与互联网用户行为数据信息进行联合建模，可更准确地判断用户偏好，提高推销理财产品的成功率。在信用卡风控建模中，利用信用卡特征数据与外部数据进行联合建模，可在贷前环节预测客户风险，降低人工审核成本。利用联邦学习，在金融行业实现数据的安全融合，以促进金融行业快速高质量发展已成为行业普遍共识。

2.1 欺诈检测问题的定义

欺诈一般被分为内部欺诈和外部欺诈。内部欺诈指由组织或机构内部成员、雇员等产生欺诈意图，通过欺诈手段非法挪用财物；外部欺诈则指由

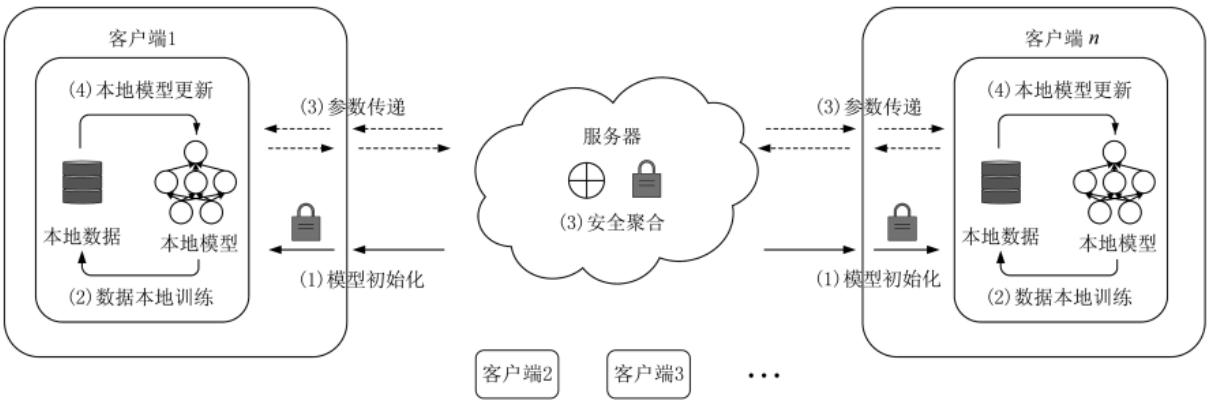

图 2 服务器-客户端联合架构[6]

表 1 国内外金融业联邦学习应用试点[7]

项目内容	平台/试点机构	领域	场景	地区
车险和健康险交叉营销	百度金融安全计算平台(百度)	保险	营销	中国
信贷互联网营销		银行	营销	
线上信贷风控		银行	信贷风控	
保险广告投放 RTA	神盾-联邦计算平台(腾讯安全)	保险	营销	中国
银行卡全生命周期风控		银行	信用卡风控	
网贷短信营销拉新		银行	营销	
江苏银行与腾讯“智能化信用卡管理联合实验室”		银行	信用卡风控	
济宁银行线上信贷业务系统		银行	信贷风控	
保险产品定价	蜂巢联邦智能平台(平安科技)	保险	营销	中国
基于纵向联邦学习的信用卡评分建模	光大银行和某云支付公司	银行	信用卡营销/风控	中国
联邦学习+理财推荐	广州银行	银行	理财产品营销	中国
联邦学习+小微企业贷款风险管理		银行	信贷风控	
联邦模盒	Fedlearn(京东科技)	银行	信贷风控	中国
联邦信贷风控	FATE(微众银行)	银行	信用卡营销/风控	中国
反洗钱联合建模		银行	反洗钱	中国
信用卡反欺诈	PrivPy(华控清交)	银行	信用卡风控	中国
Consilient	Consilient(Consilient 和 Intel)	银行	反洗钱	美国
基于差分隐私+MPC 的联邦学习信用卡反欺诈联合建模	JP Morgan、IBM 和佐治亚理工学院	银行	反洗钱	美国
基于联邦学习的银行间反洗钱分类信息共享	IBM、NICE Actimize、ING、FCA Advanced Analytics	银行	反洗钱	英国
基于联邦图计算的联合金融犯罪侦查	IBM	银行	反洗钱	美国

第三方外部盗窃，或者合作的供应商、客户通过欺诈手段来非法获取财物。

由信用卡产生的欺诈行为多为外部欺诈行为，信用卡欺诈行为的主要表现形式：(1)信用卡遗失。该类型欺诈行为不常见，因为一旦卡丢失，卡会立即被金融机构冻结。(2)信用卡信息被窃取。这是一种比较常见的信用卡欺诈行为的内因，信用卡诈骗犯一般在窃取信用卡信息后，通过电信途径使用在线交易等手段来非法窃取财物。(3)信用卡持有者在无力偿还或无意偿还的情况下，故意透支信用卡进行消费，这种一般也称作破产信用卡欺诈行为[7-8]。

2.2 欺诈行为检测问题所要面对的挑战

传统方案对于信用卡欺骗行为的检测主要依靠数理统计规律结合人工判别来实现，消耗大量的人工和时间成本，导致难以广泛推广，系统面对特殊场景的泛化性也很一般。因此需要建立高效的信用卡欺骗判定检测系统(Fraud Detection System，FDS)，来精确地判定信用卡欺诈行为。国外判定方法有如，IBM 基于 Machine Learning 和 tensor 计算的数据探测(data detectives)，Paypal 和 WePay 利用机器学习来判别复杂的金融欺诈行为。由于国内互联网金融用户规模更加巨大，授权和信用系统更加复杂和特殊，因此需要有更加有效和更为强化的泛化性的系统、模型来应对挑战。

2.2.1 数据孤岛

金融机构之间、个人与机构之间、以及金融机构与第三方外部机构之间，数据流通不畅，是金融行业产生数据孤岛的主要原因，严重影响着各产业融合的发展[9]。金融机构之间通过数据共享来实现机器学习建模的过程，难以避免数据安全或用户隐私安全等问题，导致基于数据的机器学习模型性能严重受限；而各金融机构如仅基于自己的本地数据去设计模型并进行训练，数据维度较少，特征来源单一，泛化性不强，其应用价值有限，根本无法达到金融机构所需要的高精准度判别模型。

2.2.2 数据倾斜

一般来说，金融诈骗行为在交易行为中占比很小。由于该建模问题是二分类的问题，数据却在分类标签上产生了严重的类别倾斜，即欺诈样本数据很少，很难训练出有效的分类模型。在有监督学习中，分类模型的负样本过多，会掩盖掉很多正样本的有效数据特征，从而导致正样本的规律很难被统计分析。因此，训练样本的分类倾斜会严重导致模型适用性有限，对正样本的学习能力较差。

2.2.3 实时性检测需求

一般来说，金融机构对诈骗行为的时间容忍度较短，诈骗行为在发生的较短时间内就能够对金融机构造成巨大的经济损失，金融机构需要更快、更有效地去识别判定是否存在诈骗行为。因此，好的金融欺诈行为检测模型应该兼顾计算资源和时间资源两种维度的消耗。

2.3 金融诈骗检测系统设计

在基于联邦学习的金融诈骗检测系统的整体结构设计中，将其系统设计为三个主要的模块[8]：(1)数据的平衡化模块(解决样本的倾斜问题)；(2)诈骗检测模块；(3)联合金融诈骗检测模块。整个系统的流程示意图如图 3 所示。

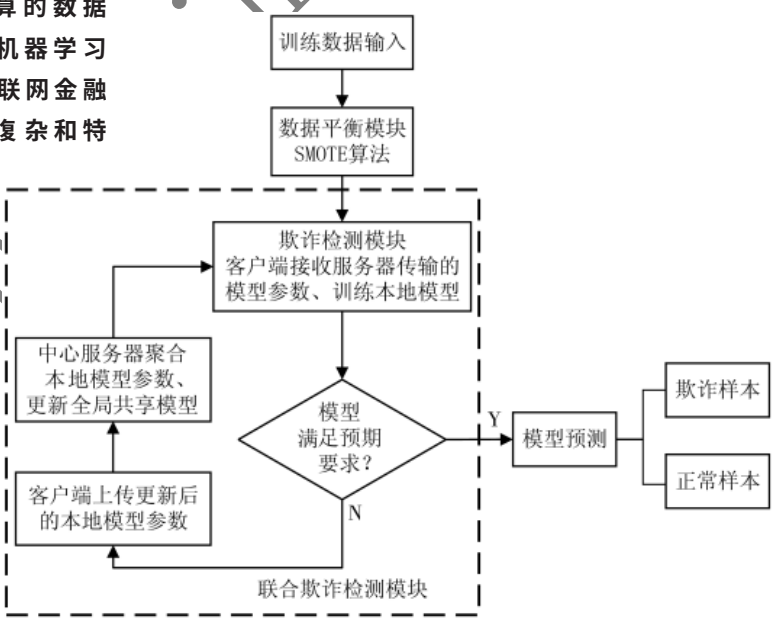

图 3 联合欺诈检测算法流程图[8]

2.3.1 数据化平衡模块

数据化平衡模块主要用于解决在信用卡诈骗问题上正负样本不均衡的问题，通过将正负样本在数据量或者模型训练贡献度平衡在相似或相同的数量级，则模型能够更好地分辨样本类别。正负样本模型训练贡献度在算法层面上又不易量化，因此对正负样本数量进行平衡化则是更易实现的方案。

通常采用的方法是对样本多类的欠采样和对样本少类的过采样，存在问题是：欠采样可能会导致样本集的数据量大规模缩水，造成模型欠拟合；过采样可能会导致负样本的数量过多，造成模型对负样本的过拟合。合成少数类过采样技术(Synthetic Minority Over-sampling Technique，SMOTE)是一种可以更好地避免普通随机过采样方法弊端的方法[9]。

SMOTE 算法的合成策略采用最近邻算法，对每个属于少数类的样本 x_j 从器 K 个近邻随机选取一个样本 $x_{i(k)}$，再生成一个 0 到 1 的随机数 r_i，使得新样本 x_{i1} 为：

$$x_{i1}=x_i+r_i(x_i-x_{i(k)}) \tag{1}$$

使用 SMOTE 算法对数据进行过采样来平衡数据，在计算上更容易实现，不会增加算法的复杂性，更为节省时间，提高效率。

2.3.2 金融诈骗检测模块

金融诈骗检测模块是各金融机构共同维护的诈骗检测模型，是整个系统的基础。模型实现对信用卡的交易记录数据进行判别分类，相当于每个金融机构利用自己的模型对数据进行进行可靠的打标；另一方面，金融机构的本地欺诈检测模型也可以嵌入联邦学习框架，完成整个系统的结构[10]。

在该模块中定义，每个训练样本的参数向量 w 定义损失函数来指导学习过程。损失函数用来量化模型在检测过程中的错误，对样本 (x_i, y_i) 来说，将损失函数定义为 $l((x_i, y_i), w)$，定义学习率为 η。神经网络结构能够更好地契合联邦学习的思想，在本地模型结构和参数的聚合层面能够提供模型同质的优势。因此，一般来讲，对于客户端 c，以固定学习率 η 在当前模型的参数 w_t 下计算本地数据集的梯度为 $\nabla L_c((x_c, y_c); w_t)$，对本地的检测模型进行更新：

$$w_{t+1}^{c} \leftarrow w_t-\eta \nabla L_c(x_c, y_c; w_t) \tag{2}$$

模块网络结构如图 4 所示。

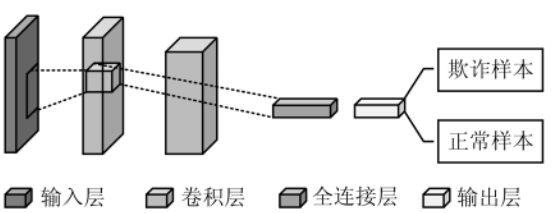

图 4 底层欺诈检测模型网络结构图[8]

2.3.3 联邦诈骗检测模块

联邦诈骗检测模块将金融机构持有的本地数据安全保留在本地，同时能够协同所有金融机构的数据共同建模。通过该模块来实现对数据孤岛的打破，去增加模型所能够学习的数据量，更加有效地去利用多家机构数据的价值[11-14]。

在信用卡诈骗领域，数据隐私安全性保护产生的数据孤岛问题，是金融机构对反诈骗行为模型检测的数据不足的根本原因。因此基于联邦学习的信用卡诈骗检测系统训练过程示意图如图 5 所示。

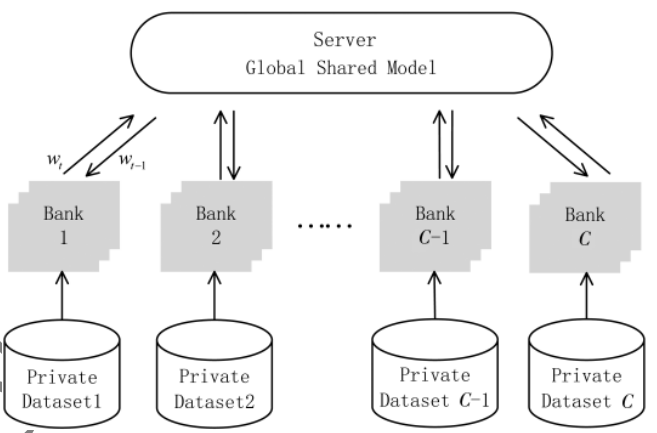

图 5 联邦诈骗检测模块训练过程图[8]

每个银行或金融机构需要利用自己的数据集去训练本地模型，w_t 表示第 t 次同步时向中央服务器传输的本地模型的参数，w_{t+1} 表示由中央服务器对本地模型上传参数进行聚合处理后传输给本地机构的模型参数，只通过中间训练的模型参数的交互来使得各金融机构能够协同训练更有效、更全局化、泛化性更好的诈骗检测模型。在这个过程中，各机构的数据保留在本地不向外传输。在参与整体的联邦诈骗检测模型之前，每个金融机构都在通用模型上进行整合，需求达成一致，因此，非凸模型的学习目标被定义为：

$$\min_{w \in \mathbf{R}^d} l(x, y; w) \text{ where } l(x, y; w) \overset{\text{def}}{=} \frac{1}{n} \sum_{i=1}^{n} l(x_i, y_i; w) \tag{3}$$

其中，任何一个金融机构 c 持有本地数据集 $|D_c|=n_c$，全部涉及的数据量为 $n=\sum_{i=c}^{C} n_c$，因此目标函数的表达式又为：

$$\begin{cases} l(x,y;w)=\sum_{c=1}^{C}\frac{n_c}{n}l_c(x_c,y_c;w) \\ l_c(x_c,y_c;w)=\frac{1}{n}\sum_{i\in D_i}l(x_i,y_i;w) \end{cases} \tag{4}$$

中央服务器初始化模型参数，选择一定比率的机构进行通信并下载模型，每个模型以固定学习率学习，计算平均损失梯度，同步更新后上传给中央服务器，由中央服务器进行参数的聚合和更新：

$$w_{t+1}\leftarrow w_t-\eta\nabla l(x,y;w) \tag{5}$$

$$w_{t+1}\leftarrow w_t-\eta\sum_{c=1}^{C}\frac{n_c}{n}\nabla L_c(x_c,y_c;w) \tag{6}$$

$$w_{t+1}\leftarrow w_t-\eta\sum_{c=1}^{C}\frac{n_c}{n}f_c \tag{7}$$

最后对每个机构c，$w_{t+1}^{c}\leftarrow w_t-\eta f_c$，代入上式则有：

$$w_{t+1}\leftarrow\sum_{c=1}^{C}\frac{n_c}{n}w_{t+1}^{c} \tag{8}$$

进一步，考虑到数据集的不平衡性，结合每个金融机构本地模型性能的准确率作为权重，来聚合更新，则有更新过程的表达式：

$$w_{t+1}\leftarrow\sum_{c=1}^{C}\frac{n_c}{n}\alpha_{t+1}^{c}w_{t+1}^{c} \tag{9}$$

整理后系统的算法流程为联邦平均(FedAvg)算法如下：

算法 1：联邦信用卡诈骗检测系统算法框架(FFD)，B 表示本地模型更新的 batch size，E 表示本地设置的 epoch 数量，t 是当前迭代轮数，η 为学习率。

输入：金融机构或银行的本地数据集

输出：基于联邦学习的信用卡诈骗检测系统模型

中央服务器进行更新：

1. 初始化网络参数 ω_0
2. For 通信轮数 $t=1,2,\cdots$ do
3. 随机选择 max(比率 $*C$, 1)个机构，记作 N_t
4. 对于每一个机构 $c\in N_t(|N_t|=K)$并行执行：
5. $w_{t+1}^{c},\alpha_{t+1}^{c}\leftarrow\text{BankUpdate}(c,w_t)$
6. $w_{t+1}\leftarrow\sum_{c=1}^{K}\frac{n_c}{n}\alpha_{t+1}^{c}w_{t+1}^{c}$

BankUpdate(c, w)：在第 c 个机构模型上执行

数据预处理过程：用 SMOTE 方法平衡原始数据集

7. $w\leftarrow w_t$
8. B←将数据集 D_n 分为大小为 batch size 为 B 的数据
9. 对每个本地 epoch 从 1 到 E 执行：
10. 对于每个 batch 的数据执行：
11. $w\leftarrow w-\eta\nabla l(x,y;w)$
12. 返回模型参数 w 和验证集的准确率 α 到中央服务器

3 联邦学习在金融数据安全领域所面临的挑战

研究表明，联邦学习在金融领域方面可以实现安全融合，促进行业高质量发展，但现有的联邦学习技术在金融行业中大规模的应用仍存在通信效率低下、数据异构及安全性不明晰等主要问题[15]。因此，如果联邦学习想要实现在金融行业大规模商业化的应用，需要主要解决以下三种问题。

3.1 通信效率

与传统的数据持有者提供本地数据，然后集中式地完成训练的方式不同，联邦学习的训练具有参与训练的客户端多、规模大、数据呈分散式分布的特点。因此，各客户端进行本地数据训练时，会受到客户端在线时间、网络环境等因素的影响，很难保证客户端全程参与训练过程。所以在设计联邦学习框架时，通信是需要考虑的一个重要因素。减少通信可从两方面着手，一是减少总通信回合的总轮数，二是减少每轮通信中信息量的传输。现有的解决方法有局部更新(将全局目标函数分解为多个子问题，并行化解决子问题)、模型压缩(稀疏化、二次采样和量化)以及分散训练(分散的拓扑，仅与相邻客户端通信)三种方式。

3.2 数据异构

联邦学习应用尚处于早期，仍在快速发展的阶段。其面临的最大难点在于联邦学习各参与方的源数据集是非独立同分布的，因此，在对数据进行联合建模时以及证明该模型是否收敛时都会存在一定的挑战。另一方面，不同客户端所拥有的本地数据集的大小也不一样，将会给联邦学习带来公平性的问题。比如，在联邦学习训练的过程中，共享模型可能会更偏向于数据量更大的客户端。再者，不同客户端间存在不兼容、差异较大等问题，导致难以实现对接。比如，应用试点较多的是一家金融机构和一家互联网企业相互对接，平台间的异构性导致数据对接繁琐，重复建设，难以满足联合建模的需求。所以，目前来看，想要根本性的解决联邦学习中异构性的这一问题还存在很大的难度。

3.3 隐私安全保护

出于隐私安全的考虑，参与联邦学习的各客户端通过加密技术共享模型的参数信息。虽然无需上传本地数据集至服务端，但是其仍然存在一定的隐私和安全问题：一是复杂模型在训练时对数据有一定的“记忆性”，从而间接导致隐私的泄露；二是安全防御手段不足导致模型的污染，所训练的数据中可能会出现一些虚假数据，从而导致模型不能真实地反映数据的分布特征。

4 结论

综上所述，联邦学习打破了传统机器学习模式，开创了一种新的面向数据隐私保护的机器学习框架，联邦学习各客户端可以通过“联邦学习”这一联合机制实现共赢的局面。目前，虽然联邦学习已在某些实际场景中得到应用，但整体上在国内金融行业还处于探索阶段。未来联邦学习在金融领域中的应用还可以在以下几个方面进行[16]。

(1)技术层面仍需加强创新。在联合建模的过程中，需要使用动态加密、差分隐私等技术去加密模型更新参数。意味着，对于较为复杂的加密系统，解密过程需要花费更多精力，因此需要在数据保护和运行效率之间寻求一个平衡。

(2)规范联邦学习标准体系。目前，有关联邦学习的开源框架层出不穷，有关联邦学习的研究正处于百家争鸣的阶段。由于联邦学习在金融数据安全领域中有着广泛的应用前景，因此相关银行应积极推进联邦学习标准化体系的建设，使联邦学习在金融领域不同业务场景中都可以合规合法地安全应用。

(3)时刻关注监管政策要求。由于金融数据具有高度的敏感性和隐私性，金融业一直都属于强监管性的行业。目前，监管机构对联邦学习在金融行业中的落地还未有明确的表态，因此需要各方密切关注联邦学习安全制度标准的建设，稳步发展联邦学习在相关领域的安全应用。

参考文献

[1] 杨强.AI与数据隐私保护：联邦学习的破解之道[J].信息安全研究，2019，5(11)：961-965.

[2] 王春凯，冯键.联邦学习在保险行业的应用研究[J].保险职业学院学报，2020，34(1)：13-17.

[3] 郑立志.基于联邦学习的数据安全在银行领域的探索[J].中国金融电脑，2020(9)：22-26.

[4] Wang Jianzong，Kong Lingwei，Huang Zhangcheng，et al. Summary of federated learning algorithms[J].Big Data，2020：1-22.

[5] BONAWITZ K，EICHNER H，GRIESKAMP W，et al. Towards federated learning at scale：system design[J]. arXiv preprint arXiv：1902.01046，2019.

[6] 王健宗，孔令炜，黄章成，等.联邦学习隐私保护研究进展[J].大数据，2021，7(3)：130-149.

[7] 陈琨，李艺，王国赛，等.联邦学习在金融行业的应用分析[J].征信，2021，39(10)：29-36.

[8] 阳文斯.基于联邦学习的信用卡欺诈检测系统研究[D].中国科学院大学(中国科学院深圳先进技术研究院)，2020.

[9] YANG Q，LIU Y，CHEN T，et al.Federated machine learning：concept and applications[J].ACM Transactions on Intelligent Systems and Technology，2019，10(2)：1-19.

[10] 李鸣.基于纵向联邦学习的推荐系统技术研究[D].杭州：浙江大学，2021.

[11] 代文，许文彬.基于联邦学习的个人信用风险评估研究[J].商业文化，2021(5)：102-107.

[12] KAIROUZ P，MCMAHAN H B，AVENT B，et al. Advances and open problems in federated learning[J]. arXiv preprint arXiv：1912. 04977，2019.

[13] 郭艳卿，王鑫磊，付海燕，等.面向隐私安全的联邦决策树算法[J].计算机学报，2021，44(10)：2090-2103.

[14] KE G，MEN Q，FINLEY T，et al.LightGBM：a highly efficient gradient boosting decision tree[C]//Proceedings of the Conference and Workshop on Neural Information Processing Systems，2017：3146-3154.

[15] LI T，SAHU A K，TALWALKAR A，et al.Federated learning：challenges，methods，and future directions[J]. IEEE Signal Processing Magazine，2020，37(3)：50-60.

[16] 张艳艳.“联邦学习”及其在金融领域的应用分析[J].农村金融研究，2020(12)：52-58.

(收稿日期：2021-12-20)

作者简介：

张海涛(1983-)，男，硕士，高级工程师，主要研究方向：金融信息安全。

版权声明

经作者授权，本论文版权和信息网络传播权归属于《信息技术与网络安全》杂志，凡未经本刊书面同意任何机构、组织和个人不得擅自复印、汇编、翻译和进行信息网络传播。未经本刊书面同意，禁止一切互联网论文资源平台非法上传、收录本论文。

截至目前，本论文已经授权被中国期刊全文数据库（CNKI）、万方数据知识服务平台、中文科技期刊数据库（维普网）、JST 日本科技技术振兴机构数据库等数据库全文收录。

对于违反上述禁止行为并违法使用本论文的机构、组织和个人，本刊将采取一切必要法律行动来维护正当权益。

特此声明！

《信息技术与网络安全》编辑部

中国电子信息产业集团有限公司第六研究所