计算机科学
COMPUTER SCIENCE

基于区块链的联邦蒸馏数据共享模型研究

刘炜, 刘宇昭, 唐琮轲, 王媛媛, 佘维, 田钊

引用本文

刘炜, 刘宇昭, 唐琮轲, 王媛媛, 佘维, 田钊. 基于区块链的联邦蒸馏数据共享模型研究[J]. 计算机科学, 2024, 51(3): 39-47.

LIU Wei, LIU Yuzhao, TANG Congke, WANG Yuanyuan, SHE Wei, TIAN Zhao. Study on Blockchain Based Federated Distillation Data Sharing Model [J]. Computer Science, 2024, 51(3): 39-47.

计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.230700186

基于区块链的联邦蒸馏数据共享模型研究

刘　炜[1,2,3]　刘宇昭[1,3]　唐琮轲[1,3]　王媛媛[5]　佘　维[1,3,4]　田　钊[1,3]
1 郑州大学网络空间安全学院　郑州 450002
2 河南省网络密码技术重点实验室(信息工程大学)　郑州 450000
3 郑州市区块链与数据智能重点实验室(郑州大学)　郑州 450000
4 嵩山实验室　郑州 450000
5 国网许昌供电公司　河南 许昌 461000
(wliu@zzu.edu.cn)

摘　要　零散、孤立的海量数据形成“数据孤岛”使得数据无法交互和连接,如何在保护原始数据隐私的前提下安全有效地共享数据中的知识信息已成为热点研究问题。基于以上内容,提出了一种基于区块链的联邦蒸馏数据共享模型(BFDS)。区别于中心化架构,采用区块链联合多参与方组建教师网络,实现分布式协同工作;通过交换蒸馏输出的方式,传递数据中的知识信息,联合训练轻量化模型;提出了一种多权重节点可信评估算法,调用智能合约分配权重并生成可溯源全局软标签,降低因参与方质量差异而产生的负向影响。实验结果表明,BFDS模型能联合多参与方安全可信共享数据知识,协同蒸馏训练模型,降低了模型的部署成本;所提出的多权重节点评估算法能有效减小低质量节点的负向影响,提高了全局软标签的质量与安全性。

关键词:区块链;知识蒸馏;数据共享;智能合约
中图分类号　TP391

Study on Blockchain Based Federated Distillation Data Sharing Model

LIU Wei[1,2,3],LIU Yuzhao[1,3],TANG Congke[1,3],WANG Yuanyuan[5],SHE Wei[1,3,4] and TIAN Zhao[1,3]
1 School of Cyber Science and Engineering,Zhengzhou University,Zhengzhou 450002,China
2 Henan Key Laboratory of Network Cryptography Technology(Information Engineering University),Zhengzhou 450000,China
3 Zhengzhou Key Laboratory of Blockchain and Data Intelligence(Zhengzhou University),Zhengzhou 450000,China
4 Songshan Laboratory,Zhengzhou 450000,China
5 State Grid Henan Electric Power Company,Xuchang,Henan 461000,China

Abstract　The privacy of raw data makes it difficult to be directly shared among multiple participants. The issue of data security sharing and privacy-preserving has become a hot research topic. To solve this problem,this paper proposes a blockchain-based federated distillation data sharing model(BFDS). It utilizes blockchain to form a collaborative teacher network with multiple participants. Through distilled output exchange,the knowledge from complex teacher networks is transferred and used to train lightweight models. A novel multi-weight node trust evaluation algorithm is proposed that uses smart contracts to generate traceable global soft labels. It can reduce the negative impact caused by quality differences among participants. Experimental results show that BFDS can collaborate with multiple parties to share data knowledge reliably,distill training models collaboratively,and reduce model deployment costs. The proposed algorithm can effectively reduce the negative impact of low-quality nodes and improve the quality and security of global soft labels.

Keywords　Blockchain,Knowledge distillation,Data sharing,Smart contracts

1　引言

随着人类社会进入大数据时代,蕴含知识信息的海量数据成为了重要的生产要素和资源。机器学习[1]可以帮助人们从海量数据中挖掘规律、做出决策进而理解世界。机器学习模型的性能和精度取决于训练数据的质量、数量和多样性,

到稿日期:2023-07-24　返修日期:2023-11-30
基金项目:河南省高校科技创新人才支持计划(21HASTIT031);河南省网络密码技术重点实验室研究课题(LNCT2022-A04);河南省高等学校重点科研项目(24A520045);嵩山实验室预研项目(YYYY022022003)
This work was supported by the Henan Province University Science and Technology Innovation Talent Support Plan(21HASTIT031),Research Project of Henan Provincial Key Laboratory of Network Cryptography Technology(LNCT2022-A04), Key Scientific Research Project of Colleges and Universities in Henan Province(24A520045) and Songshan Laboratory Pre-research Project(YYYY022022003).
通信作者:田钊(tianzhao@zzu.edu.cn)

不充分、质量差或是单一的训练数据集会限制模型适应新数据的能力。然而，现实生活中数据通常分散在不同的数据中心或机构中，数据的隐私属性也意味着无法直接分享原始数据信息，数据无法交互和连接，形成了“数据孤岛”。如何在保护原始数据隐私的前提下，有效地共享数据中的知识信息已成为热点研究问题。

联邦学习(Federated Learning，FL)[2]区别于上传数据的集中训练方式，采用“数据不动模型动”的思想。在全局模型迭代过程中，各数据中心使用本地数据训练初始模型，仅上传本地训练的梯度更新，再由中心服务器聚合成全局模型，进而实现多数据中心进行联合训练，训练出更具泛化能力的模型以适应不同数据分布和用户行为。但联邦学习也存在一些问题：如中央服务器易发生单点故障[3]；上传梯度参数的交互方式易受到推理攻击，存在隐私泄露风险[4]；多个数据中心参与训练缺乏足够的信任，低质量的参与方影响训练的结果；联合训练得到的模型参数量较大，通信成本与模型的部署成本较高等。针对以上问题，需要提出新的多参与方数据共享范式，以适应复杂的分布式隐私数据共享环境。

区块链作为一种利用点对点网络构建的分布式账本，因其具有去中心化、不可篡改、可溯源等特点而被应用于许多领域[5]。基于区块链技术构建多中心参与的内生信任训练环境，解决单点故障问题，利用智能合约自动执行预定义的规则操作，使任务过程可信可溯源。联邦蒸馏(Federated Distillation，FD)算法通过交换蒸馏输出的方式传递知识信息。相比传递参数信息，通信开销大幅降低并能够解决 FL 中的推理攻击问题。区块链结合联邦蒸馏可以构建分布式安全可信的数据共享模型，实现数据不出本地的知识信息安全共享。

综上所述，本文提出了一种基于区块链的分布式联邦蒸馏数据共享模型，采用区块链代替中心服务器，防止出现单点故障问题；设计智能合约评估节点并分配权重，减小低质量节点负向影响；采用软标签蒸馏训练轻量化模型，在准确率略有损失的情况下压缩模型，降低通信开销和部署成本[6]。本文的主要贡献如下：

1)提出了一种基于区块链的联邦蒸馏数据共享训练模型，利用区块链取代中央服务器，解决多参与方联合训练过程中的单点故障问题，构建去中心化的可信多方联合训练环境，保证各方原始数据不出本地的情况下提取学习数据的知识并用于模型训练。

2)提出了多权重节点可信评估算法，利用智能合约进行参与方的筛选与权重分配，降低了低质量节点影响全局训练结果的影响。将全局标签生成信息上链，使训练过程可溯源，提高了训练的可信度与安全性。

3)采用软标签知识蒸馏的轻量化模型训练，减少模型参数量，降低通信开销与模型部署成本，在保证各参与方数据不出本地和容许模型异构的前提下，进行安全有效的联合训练。

2 相关工作

自知识蒸馏技术于 2015 年被机器学习之父 Geoffrey Hinton 提出以来[7]，专家学者对其进行了大量的研究与应用。Wu 等[8]提出了利用知识蒸馏进行高效通信联邦学习的方法，有效降低了联邦学习的通信成本，但整体训练过程需要中心服务器参与，易出现单点故障问题。Xu 等[9]提出了利用生成对抗网络进行知识蒸馏训练的方案，可以将知识从教师网络转移到学生网络，但仅在相对较小的学生网络有效。Chen 等[10]提出了一种基于注意力联邦蒸馏的推荐算法，减少教师网络和学生网络的差异性影响，但中心化的架构易出现单点故障问题。Sun 等[11]提出了一种基于合作博弈和知识蒸馏的个性化联邦学习算法，以衡量多客户端间多重协作的影响，但也存在单点故障的隐患。Cheng 等[12]利用无标签数据避免隐私泄露问题，在数据共享上仍有改进空间。Mo 等[13]提出了一种联邦蒸馏算法 FedDQ，利用联邦蒸馏减小联邦学习中的通信开销，但仍需中心服务器进行聚合，易遭受单点故障问题。Yao 等[14]提出了 FEDGKD 方法，使用全局模型蒸馏进行本地训练，解决局部训练偏差问题，但也存在单发生点故障的风险。Li 等[15]提出了一种基于区块链和联邦蒸馏的远程医疗数据共享系统，采用平均算法未考虑低质量标签的影响。Li 等[16]提出了 HBMD-FL，使用区块链取代联邦学习中心服务器，利用模型蒸馏解决模型异构性，但在聚合过程中未考虑低质量模型标签的影响。

已有的研究工作中，基于中央服务器架构的方案易出现单点故障问题；使用公共模型联邦蒸馏的方案，在联合多方训练新模型方面仍有不足；在多参与方蒸馏输出结果的质量差异对全局软标签的影响方面也有所忽视。对于构建安全的区块链联邦蒸馏数据共享模型，仍需进一步深入研究。

3 系统模型

在多参与方的联合学习场景中，任务发布方和接受方通过基于区块链的任务发布平台进行撮合与匹配，依靠区块链的去中心化、可溯源、不可篡改等特性在多个互不信任的节点间建立信任。利用知识蒸馏完成各节点间知识的提取与传递。本章构建了基于区块链的联邦蒸馏数据共享模型(Blockchain Federated Distillation Data Sharing Model，BFDS)，引入多权重节点可信评估算法评估节点可信度与训练权重，从联合教师网络中提取“知识”用于训练学生模型，并将训练过程进行上链，对节点的训练表现进行评估和打分，使训练过程公开透明、可溯源。

3.1 模型概述

BFDS 利用区块链取代传统联合训练的中心服务器，将多个数据量较大和准确率较高的本地优质模型的节点联合起来，组成大型分布式联合教师网络，避免中心化服务器出现单点故障而影响模型训练过程。通过知识蒸馏的方式，利用智能合约评估参与方节点，自动分配相应的权重，用于投票产生训练学生模型所需的软标签(Soft Labels)，记录训练过程，标签生成可追溯、透明化，减小学生模型的预测误差。该模型成员主要由 3 部分组成：任务发布方、教师网络参与方节点和共识节点。系统的符号和说明如表 1 所列，整体框架如图 1 所示。

1)任务发布方 $P0$ 是一次多方联合学习任务的发起者，其首先将训练集 $D_{tr}(P0)$ 和测试集 $D_{te}(P0)$ 上传到区块链，再根据训练所需要的训练数据类型 C_{P0}^{*}、目标分类数 N_{P0}^{*}、全局

蒸馏温度 T、选取参与组建教师网络的节点的阈值、期望加入教师网络节点的最低测试集准确率 $Acc^*(D_{te})$ 和历史信誉值评分 SH 发起联邦学习任务。满足要求的节点可以通过部署在区块链上的智能合约申请参与该联合学习任务。随着节点的加入和训练，任务发布方最终将得到来自不同节点的教师网络软标签，用于蒸馏训练本地模型，最终得到参数量规模较小但性能接近于教师网络的轻量化模型。

2）教师网络参与方节点 R_j 是若干个对任务发布方 $P0$ 发起的联合训练任务感兴趣并通过准入控制审核的节点，是海量数据节点 $R_i(i=1,2,\cdots,n)$ 的子集。如医院、银行、工厂、移动运营商等本身具有大量蕴含知识的数据，受限于法律或是隐私需要，数据无法进行直接分享参与联合训练。参与方根据自身拥有的数据类型，经由不同的任务通告加入不同领域的联合训练任务。这些节点作为数据的拥有者，本身已经从自身拥有的大量数据中提取了一定的知识，具有准确率较高的模型，但是由于地理位置、地域分布、环境差异等因素，本身数据具有一定的局限性。存在着“信息茧房”效应，参与组成教师网络的节点既是训练过程软标签生成的参与方，也是受益方。通过联合学习生成的软标签可以汲取其他不同地区和环境因素影响的节点所具有的数据中的知识，参与方可以对该标签进行利用，蒸馏自身本地模型，进一步提高模型的泛化能力和准确率。

3）共识节点 B_i 负责维护区块链的安全性和稳定性。系统的联盟链对加入链的节点的身份进行审核和验证，通过准入控制实现对节点的入链管理。能够通过预先审核的节点是具有相对稳定性的参与方，由众多参与方共同维护区块链账本的一致性、安全性、稳定性。训练任务信息和节点的参与信息等以交易的形式记录，并经由共识节点验证后打包成区块存储在区块链上。在主链上，共识节点由其他所有通过了联盟链准入控制的节点组成，并且通过 PBFT 共识算法进行共识，得到超过 2/3 的节点验证的信息将被写入区块。在子链上，共识节点由所有参与组建教师网络的节点组成，节点间通过委员会共识算法进行共识，无须消耗算力进行数学难题计算，根据选出的共识委员会，按照预定好的顺序轮流作为记账人验证和打包区块。

表 1　符号和描述

Table 1　Symbols and descriptions

符号	描述
$P0$	任务发布方
$R_i(i=1,2,\cdots,n)$	海量数据节点
$R_j(j=1,2,\cdots,k),k\leqslant n$	组建教师网络节点
N_{P0}^*	目标分类数
C_{P0}^*	训练数据类别
T	全局蒸馏温度
$D_{tr}(P0)$	发布方训练集
$D_{te}(P0)$	发布方测试集
$Acc^*(Dte)$	在发布方测试集的准确率
SH	历史信誉值评分
SC	本次评估得分
softmax T	温度 T 的蒸馏输出
L	学生模型损失函数
L_{soft}	softlabel 损失函数
L_{hard}	hardlabel 损失函数
W_j	合约生成的 R_j 权重
θ	L_{soft} 权重因子
ξ	得分系数因子

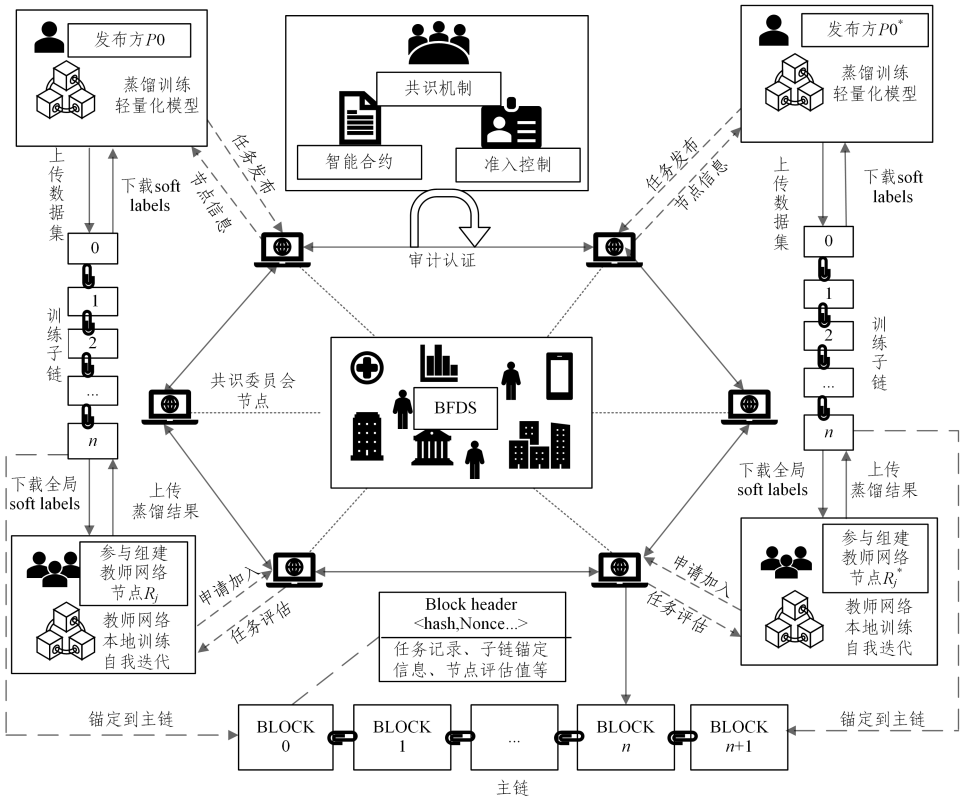


图 1　系统整体框架

Fig. 1　Overall system framework

3.2 系统流程

在多方联合训练的场景下，不同的节点往往具有不同的数据集和异构的本地模型，联合不同节点执行多分类模型训练任务需要考虑节点间的异构性。

现有的多种知识蒸馏方法中，KD 知识蒸馏算法的效果最好，算法思路简单且适用于任意网络结构，在多分类任务中表现出了优秀的性能[17]。因此该模型采用 KD 蒸馏算法，对教师网络模型输出的 logits 进行蒸馏，通过生成全局软标签的方式，将大型教师网络输出结果中蕴含的知识传递给学生模型进行学习。整体训练流程如图 2 所示。

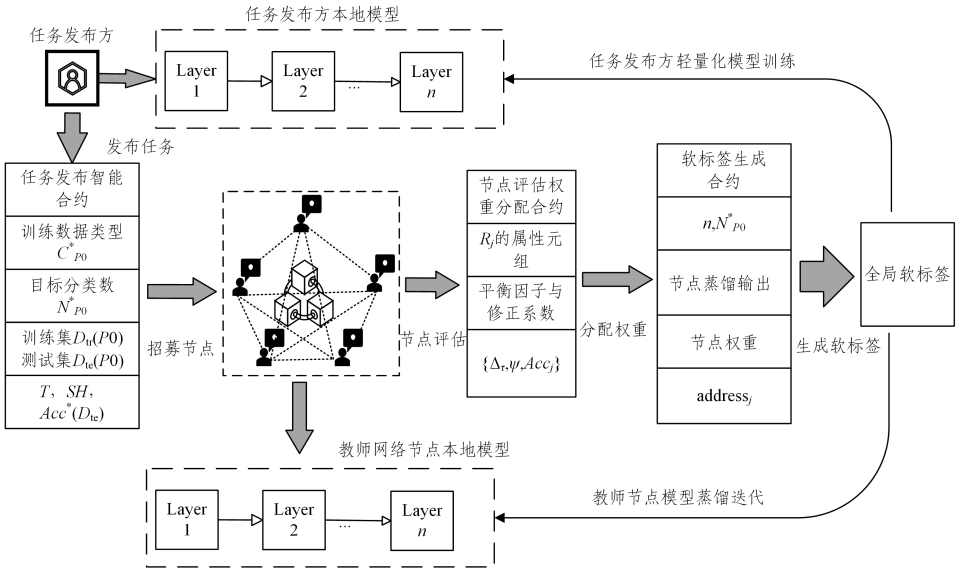


图 2 系统整体训练流程

Fig. 2 Overall training process of system

整体多参与方训练的流程如下：

1)任务发布方 $P0$ 通过任务发布合约，输入期望训练任务的属性集，发布训练任务。智能合约将对应的任务发布在区块链主链，同时 $P0$ 上传自己的训练集图像和测试集作为模型训练子链的创世纪区块，等待其他节点参与进行多方联合训练任务。子链的所有权归任务发起者拥有，参与组件教师网络的节点享有访问权和使用权。子链以交易的形式记录智能合约生成软标签的过程，以及各节点在软标签生成过程中的权重、输入信息等。

2)对多方联合训练任务产生兴趣并具有合适数据集类型和数据大小的节点 R_j，通过节点验证合约审核加入多方联合训练任务中。智能合约将对应的节点信息打包上链并返回给任务发布方 $P0$。

3)R_j 下载 $P0$ 上传的数据集图像，使用自己本地数据集训练出的本地模型对训练数据集在蒸馏温度 T 下生成软标签，将包含节点签名信息的软标签上传到训练子链的区块链网络中，经由共识节点的审计与验证打包成区块，记录在子链上，其中式(1)为温度 T 下的蒸馏输出。

$$\text{softmax } T=\frac{e^{\frac{x_i}{T}}}{\sum_{i=0}^{n} e^{\frac{x_i}{T}}} \tag{1}$$

4)节点评估权重分配智能合约根据打包的节点信息为每个参与方节点 R_j 生成对应的权重 W_j，并将权重信息记录在链上。软标签生成智能合约根据节点权重 W_j 和链上对应的各个节点生成的软标签，进行数据标准化处理后生成最终的全局软标签，将对应的生成结果打包在区块链上以供任务发布方和参与方访问和使用。

5)任务发布方 $P0$ 通过训练子链获取最终的全局软标签，用于自身本地训练轻量化学生模型。式(2)为学生模型的损失函数：

$$L=\theta * L_{\text{soft}}+(1-\theta)L_{\text{hard}} \tag{2}$$

其中，θ 为软标签的学习权重因子，L_{soft} 和 L_{hard} 分别如式(3)和式(4)所示：

$$L_{\text{soft}}=CE(p^T, q^T) \tag{3}$$

其中，L_{soft} 表示教师网络与学生模型在蒸馏温度 T 下的学习损失函数，p 表示教师网络通过温度 T 的 softmax 分布，q 表示相同温度下学生模型的 softmax 分布，求取两者的交叉熵。

$$L_{\text{hard}}=CE(y, q^{T=1}) \tag{4}$$

其中，L_{hard} 为学生模型自己训练的损失函数，y 是数据集的 $label$，q 表示学生模型在 $T=1$ 时的输出结果。任务发布方 $P0$ 通过上述方式从区块链网络的训练子链得到全局软标签进行本地的轻量化模型蒸馏训练。教师网络的参与方节点也可以通过全局软标签进一步迭代更新本地的模型。

6)任务发布方 $P0$ 对参与方节点进行打分，并将打分结果上链，根据当前打分结果与各节点的历史评分，更新各节点的分数，其中 ξ 为得分系数因子。式(5)为节点分数的更新方式：

$$SC_{\text{new}}=\frac{SC}{SH}\{SH+\xi(SH-SC)\} \tag{5}$$

3.3 多权重节点可信评估算法

参与组建教师网络的节点，因其数据集与模型具有差异性，不同节点在训练任务中的表现情况不同，这会对全局软标签的生成产生不同的影响。为使产生的软标签有利于教师网络的知识向学生模型转移，需要为优质的节点和模型分配更高的权重，减小低质量节点对模型的影响。对此提出了一种

多权重的节点可信评估算法，对节点的可信度和质量进行评估。该算法将数据集质量与模型质量作为评估依据，对参与组建教师网络的节点进行综合考量。其中参与方数据集质量主要从3方面进行评估：数据集新鲜度、数据集大小以及本地数据集与上传数据集比值。参与方节点的模型质量也从模型参数量规模、模型泛化能力和测试集准确率3方面进行评估。

数据的时效性对持续产生新数据记录的参与方具有重要意义，新鲜度高的数据对模型训练准确性和可靠性的提升起到了至关重要的作用。故引入平均数据年龄对参与方整体数据集的新鲜度进行评估。式(6)为数据集的平均数据年龄 $\Delta\tau$：

$$\begin{cases} U(k)=t_k \\ \Delta(k)=t_{\mathrm{cur}}-U(k) \\ \Delta_\tau=\dfrac{1}{\tau}\sum\limits_{k=1}^{\tau}\Delta(k) \end{cases} \tag{6}$$

其中，$U(k)$为时间戳函数，代表一条数据集记录的产生时间戳；t_{cur}表示当前时间；$\Delta\tau$ 为数据集的平均数据年龄，平均数据年龄越小代表数据的新鲜度越高。节点 i 的数据集质量评估得分 Ω_i 如式(7)所示：

$$\Omega_i=\frac{\gamma\cdot L_i^2}{S\cdot\Delta_{\tau_i}} \tag{7}$$

其中，L_i 为参与方数据集大小；S 为任务发布方上传的训练集大小；γ 为修正系数，避免数据集大小的绝对值过大从而影响其他评估因素生效。任务发布方可以根据参与方的训练情况和个人偏好进行个性化调整，调整权重信息的比重。

定义1(模型泛化误差)　泛化能力指模型对未知数据的预测能力，是机器学习的重要性质，一般采用测试误差评价模型的泛化能力，但是该评价体系对测试数据集的要求较高，有限的测试数据集可能得出的评价结果是不可靠的。为此，引入泛化误差上界[18]对模型的泛化能力进行评估。对于泛化误差：设模型为 $\hat{f}$，则该模型对未知数据预测的误差即为泛化误差，其表达式如式(8)所示：

$$\begin{aligned} R_{\exp}(\hat{f})&=E_P[L(Y,\hat{f}(X))] \\ &=\int_{X*Y}L(Y,\hat{f}(X))P(x,y)\mathrm{d}x\mathrm{d}y \end{aligned} \tag{8}$$

则泛化误差等价于期望风险 $E_P[L(Y,\hat{f}(x))]$，期望风险 $R(f)$ 与训练误差 $\hat{R}(f)$ 满足不等式(9)：

$$\begin{aligned} &R(f)\leqslant\hat{R}(f)+\varepsilon(d,N,\delta) \\ &\varepsilon(d,N,\delta)=\sqrt{\frac{1}{2N}\left(\log d+\log\frac{1}{\delta}\right)} \end{aligned} \tag{9}$$

其中，d 为函数的个数，表示为 $\Gamma\{f_1,f_2,\cdots,f_d\}$，$N$ 为样本容量，δ 为置信度。通过上述分析可以得出模型泛化误差上界是 N 的单调递减函数，是 d 的阶函数。泛化误差上界越小的模型，其期望风险越小，泛化能力越强，泛化误差上界是期望值，故其相对大小的意义要大于绝对大小。因此定义泛化能力因子 ζ 如式(10)所示：

$$\begin{cases} \zeta_j=\dfrac{1}{p_j} \\ p_j=\dfrac{R(f_j)}{\sum\limits_{i=0}^{n}R(f_j)} \end{cases} \tag{10}$$

其中，节点 i 的模型质量评估得分 ψ_i 如式(11)所示。φ_i 表示节点 i 的模型参数量，$\varphi_{\max}$ 和 $\varphi_{\min}$ 分别表示最大节点参数量与最小节点参数量，σ 为平衡系数，μ 为非零参数，Acc_i 表示节点 i 的公共测试集准确率，由该节点在 $P0$ 给出的测试集上验证得出。

$$\psi_i=\frac{\zeta_i\cdot(\varphi_i-\varphi_{\min}+\mu)\cdot Acc_i}{\sigma\cdot(\varphi_{\max}-\varphi_{\min})} \tag{11}$$

综上所述，节点的最终评估得分如式(12)所示，λ 与 ε 为平衡因子，且 $\lambda+\varepsilon=1$。

$$Score_i=\lambda\frac{\gamma\cdot L_i^2}{S\cdot\Delta_{\tau_i}}+\varepsilon\psi_i \tag{12}$$

3.4　智能合约设计

3.4.1　节点评估权重分配合约

在多节点联合训练的过程中，节点间存在着质量的差异，节点本地的数据集大小、模型泛化能力与准确率都不相同。选取出质量较高的节点并为其分配更高的权重，可以在产生全局软标签的过程中获得更优的输出分布结果。基于此，提出了节点评估权重分配算法，并由智能合约自动执行，记录上链。节点评估权重分配算法的伪代码如算法1所示。

算法1　节点评估权重分配算法

输入：R_j 的属性元组$\{L,\Delta\tau,\varphi,R(f),Acc\}_j$；$\lambda$；$\varepsilon$；$\gamma$；$\sigma$；$S$；n1

输出：各节点权重$\{w_1,w_2,w_3,\cdots,w_j\}$

```
1. / * 参数初始化 * /
2. int Ω[n1],ζ[n1],ψ[n1],
3. int Sc[n1],tp_R=0,m=0,tp=0
4. int w[n1],i=0
5. / * 计算各节点数据集质量评估得分 * /
6. for(int j=0;j<n1;j++)
7.     Ω[j]←Ω_j=γ·L_j^2/(S·Δτ_j)
8. / * 累加计算总泛化误差上界 * /
9.     tp_R +=R(f)[j]
10. / * 遍历 φj 两两对比求取 φMax 与 φMin * /
11. / * 计算模型得分与最终得分 * /
12. for(int j=0;j<n1;j++) #
13. / * 计算泛化能力因子存入数组 * /
14.   ζ[j] ←tp_R /R(f)[j]
15.   while m<=n1
16.   {
17.   ψ[m] ←ψ_i=ζ_i·(φ_i-φ_min+μ)·Acc_i/(σ·(φ_max-φ_min))
18.   Sc[m] ←Score_m=λγ·L_m^2/(S·Δτ_m)+εψ_m
19.   tp +=Sc[m],m++
20.   }
21.   / * 输出节点权重列表 * /
22.   do{
23.     w[i]=Sc[i]/tp
24.     print w[i];i++;
25.     }
26.     while(i<=n 1)
27.   return w[]
28.   End
```

3.4.2　软标签生成合约

通过节点评估权重分配算法为各个节点分配相应的权重

后，就可以进一步调用软标签生成合约。依据分配的权重与各个参与方节点的蒸馏输出结果，聚合生成全局软标签，将对应的软标签生成过程记录在链上，使整个过程可溯源。软标签生成合约的伪代码如算法 2 所示。

算法 2 软标签加权生成算法

输入：P0 训练集图像数量 n 预期分类数 N_{P0}^{*}；$label\_output_{nl}\{key1:value1, key2:value2, \cdots, key\ N*P0:valueN*P0\}_n$；节点权重 $\{w_1, w_2, w_3, \cdots, w_j\}$；节点地址 $address_j$

输出：S[n][n_class]

```
1. /* 参数初始化，绑定地址 */
2. int i=1
3. struct teacher{uint weight,bool succ,}
4. mapping(address=> teacher) public teachers;
5. /* 循环遍历蒸馏输出，加权聚合 */
6. for(i=1;i<=n;i++)
7.    for(k=1;k<=m;k++)
8.       S[i][k]=Σ^nl wj * outputj{k:value}
9. /* 输出投票信息打包上链 */
10. return S[n][n_class]
11. End
```

4 实验与分析

为验证方案的效果与有效性，实验分别在包含 60 000 个训练样本和 10 000 个测试样本的 MNIST 数据集和包含 100 个类别的 CIFAR100 数据集上进行测试。实验环境为 1 台 12th Gen Intel(R) Core(TM) i5-12490F CPU，3.00 GHz，32 GB 机带 RAM，NVIDIA GeForce RTX3060 GPU，操作系统为 Windows10。实验采用 cuda_11.6，Python3.9 和 Pytorch 实现教师网络模型和学生模型的编写与搭建，使用 FISCO BCOS 区块链系统模拟区块链环境，利用 FISCO BCOS 所提供的 Python SDK 实现系统的逻辑，智能合约使用 Solidity 编写。

4.1 安全性分析

本节从隐私性、可信性和可用性 3 方面对所提出的数据共享模型进行安全性分析。

1)隐私性：在联邦学习中，各参与方在不交换本地数据的情况下，通过上传梯度更新信息协作训练全局模型，敌手可以通过 DLG(Deep Leakage from Gradients)攻击手段[19]推理出参与方本地数据及标签信息，导致数据隐私泄露。图 3 给出了通过 DLG 攻击手段逐步恢复参与方原始数据的结果。在 BFDS 模型中，各参与方通过上传对任务发布方图像数据的蒸馏输出结果构建全局软标签，敌手仅通过软标签输出结果难以发起反演攻击[20]或成员推理攻击[21]，无法得到参与方本地原始数据信息。任务发布方上传的无标签图像数据信息不包含图像的真实标签与其他关联性隐私信息，并可以通过上传指定的公开数据集或是通过上传 GAN 生成对抗网络产生的虚拟图片，进一步保护本地原始数据降低数据隐私泄露风险。相比原始数据信息的直接泄露，该模型的隐私泄露风险较低，能为各参与方提供相对可靠的数据隐私保护。

2)可信性：在传统联邦学习中，中央服务器占据主导地位易遭受单点故障问题，对参与方上传的数据缺少相应的审查机制。在 BFDS 模型中，通过引入区块链构建去中心化训练框架可以有效解决单点故障问题，通过联盟链的准入机制对参与方身份进行审核，避免恶意节点的加入，利用共识委员会和智能合约对参与方上传的数据进行评估、打包上链，整体流程可溯源、可审计，保障了联合学习的可信性。

3)可用性：在传统联邦学习中，每轮全局模型迭代均需要各参与方上传本地梯度更新信息，模型参数量大且通信轮次多，低质量局部模型也会影响全局模型的可用性。在 BFDS 模型中，通信行为仅发生在任务发布方上传训练图像信息和下载全局软标签以及参与方下载训练图像信息和上传本地蒸馏输出的过程中。由于通信过程中仅交互软标签信息，因此能有效减少通信轮次，降低通信开销[22]。并且 BFDS 模型通过多权重节点评估算法为低质量节点分配更小的聚合权重，可以进一步保障全局软标签的可用性。

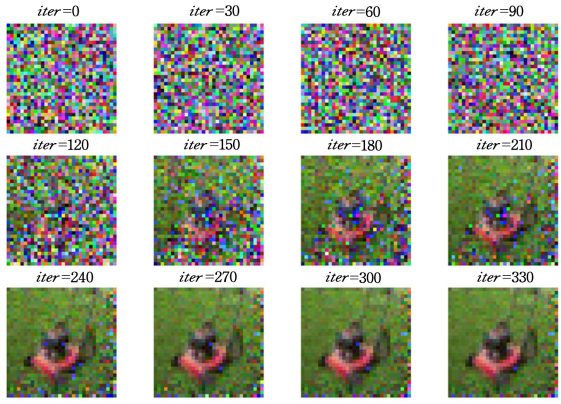


图 3 通过 DLG 还原参与方数据

Fig. 3 Restore data through DLG

4.2 知识蒸馏实验

4.2.1 知识蒸馏温度实验

首先对蒸馏温度进行测试，选取出合适的蒸馏温度以便于后续实验的开展。利用 pytorch 搭建一个包含 fc1，fc2，fc3，dropout，参数分别为(784，2 400)，(2 400，2 400)，(2 400，10)的网络模型，前向预测和反向传播进行 50 轮模型训练，准确率为 98.77%。随机抽取一个样本的全连接层输出作为蒸馏产生软标签的输入 logits。分别在蒸馏温度 $T=1,3,5,10,100$ 的情况下测试软标签的输出结果。实验结果如图 4 所示。

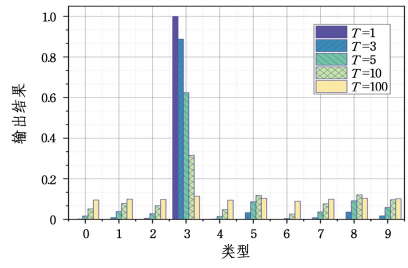


图 4 蒸馏温度对比

Fig. 4 Distillation temperature comparison

当 $T=1$ 时，softmax 的输出仅有单个极大值，其他类别的输出结果接近 0，此时的输出结果为 hard label，不包含各类别之间的联系信息，学生模型无法学习到更多的知识，因而泛化能力降低。当 $T=3$，$T=5$，$T=10$ 时，softmax 输出结果给出了该样本真实类别与其他分类间联系的紧密程度，学生

模型可以从输出结果学习到各类别间的联系，进而提高泛化能力。当 $T=100$ 时，各类别输出结果的分布近似均匀，此时蒸馏温度过高，正确类别和其他类别的输出分布结果差异性被过度拉低，无法有效学习到知识。

4.2.2　模型蒸馏对比实验

考虑到联合学习场景下，教师模型与学生模型的结构往往存在差异，设计了 3 组不同教师模型与学生模型组合，探究在 CIFAR100 数据集上的训练效果。Net-T 和 Net-S 分别表示教师与学生模型的类别，Para-T 和 Para-S 分别表示教师与学生模型的参数量，Acc-T 和 Acc-S 分别表示教师与学生模型独自训练的准确率情况，Acc-KDS 表示学生模型蒸馏训练的结果。训练结果如表 2 所列。

表 2　3 组模型的蒸馏实验

Table 2　Three sets of model distillation experiments

Net-T	WRN-40-2	WRN-40-2	RESNET50
Net-S	WRN-16-2	WRN-40-1	MobileNetV2
Para-T	2255156	2255156	25636712
Para-S	703284	569780	3504872
Acc-T	75.88%	75.88%	79.55%
Acc-S	72.98%	71.77%	65.21%
Acc-KDS	74.96%	73.52%	67.36%

其中，WRN-40-2，WRN-16-2 和 WRN-40-1 是同构模型，RESNET50 和 MobileNetV2 是异构模型，通过知识蒸馏均能有效提高学生模型的准确率，使其逼近教师模型的训练效果。学生模型相比教师模型在参数量上分别减少了 68.81%，74.73%和 86.32%，蒸馏训练对比单独训练学生模型准确率分别提升了 1.98%，1.75%和 2.15%，证明了 KD 蒸馏算法在同构与异构模型间能够降低模型的部署成本，提高学生模型的准确率。

4.2.3　基于区块链的联邦蒸馏有效性实验

在知识蒸馏温度实验的基础上，选取 $T=5$ 作为全局蒸馏温度，利用 pytorch 构建 10 个相互独立的教师网络节点客户端，模型结构如表 3 所列。

表 3　教师模型结构及参数量

Table 3　Structure and parameter quantity of teacher model

Layer(type:depth-idx)	Param #
TeacherModel:	—
ReLU:1-1	—
Linear:1-2	1884000
Linear:1-3	5762400
Linear:1-4	24010
Dropout:1-5	—
Total params:	7670410
Trainable params:	7670410
Non-Trainable params	0

各教师网络节点首先本地训练 10 轮，其平均准确率达到 98.43%，再调用软标签生成智能合约，根据权重分配生成软标签蒸馏训练学生模型。利用软标签与学生模型在 $T=5$ 时产生的 Soft Predictions 求取 loss_soft，并与 loss_hard 加权求和产生 loss_total，实验的 loss_hard 权重为 0.8。学生模型结构及参数量如表 4 所列。

表 4　学生模型的结构及参数量

Table 4　Structure and parameter quantity of student model

Layer(type:depth-idx)	Param #
StudentModel:	—
ReLU:1-1	—
Linear:1-2	7850
Linear:1-3	110
Linear:1-4	110
Dropout:1-5	—
Total params:	8070
Trainable params:	8070
Non-Trainable params	0

直接进行模型训练的学生模型记作 SD1，使用 BFDS 生成软标签联合训练的学生模型记作 SD2。通过实验对比在相同模型结构、训练集和测试集的情况下，两种方式各自训练 50 轮的准确率与收敛速度。结果取 10 次重复实验的平均值，如图 5 所示。

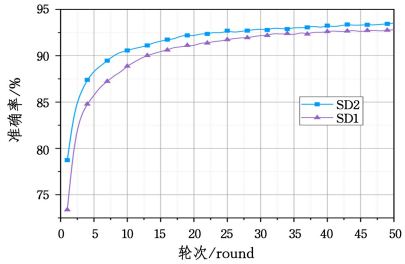


图 5　50 轮 SD1 与 SD2 的对比

Fig. 5　Comparison of 50 rounds of SD1 and SD2

SD1 准确率为 92.81%，SD2 准确率达到 93.52%，SD2 相比 SD1 准确率提升了 0.71%。在训练过程中 SD2 的整体效果和收敛速度均优于 SD1。设置 6 个具用相同初始模型的节点分别记作 TD1，TD2，TD3，TD4，TD5，TD6，将训练集随机平均分为 6 份，每个节点拥有 10000 个不同的样本。其中 TD1-TD5 利用本地样本训练 50 轮，并通过 BDFS 对节点 TD6 的训练集生成软标签，节点 TD6 使用软标签和本地训练集进行 50 轮模型训练，实验结果如图 6 所示。

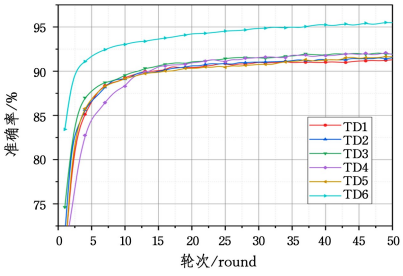


图 6　参与方节点实验

Fig. 6　Participant node experiment

TD1-TD5 的准确率分别为 91.29%，91.54%，91.96%，91.88%，91.62%，而 TD6 的准确率达到了 95.49%，相比 TD1-TD5 准确率平均提高了 3.83%。实验结果表明，在多个模型参数量和本地数据量接近的节点中，使用 BDFS 生成的软标签仍对模型准确率的提高具有明显的作用，即意味着参与组建教师网络的节点也能经 BDFS 进一步蒸馏训练本地模型，提高模型的泛化能力和准确率。由于在知识传递过程中使用的是生成的软标签，最终只传递分类信息与对应的分布

概率，而不受各节点模型异构的影响，因此 BDFS 同样适用于异构模型间进行联合训练，其具有一定的通用性。

4.3 节点评估实验

实验设置两种不同规模的教师网络节点：节点 GN(模型结构如表 3 所列)和节点 BN(模型结构如表 5 所列)。表 6 列出了不同模型结构训练的平均准确率，GN 为 98.06%，BN 为 90.02%，学生模型为 88.11%。进行 3 组实验，对比不同的节点组合对训练结果的影响，训练结果如图 7 所示。

表 5 BN 的模型结构及参数量

Table 5 Model structure and parameter quantity of BN

Layer(type:depth-idx)	Param #
BN_Model:	—
ReLU: 1-1	—
Linear: 1-2	15700
Linear: 1-3	420
Linear: 1-4	210
Dropout: 1-5	—
Total params:	16330
Trainable params:	8070
Non-Trainable params	0

表 6 不同模型结构训练平均准确率

Table 6 Average accuracy of training for different model structures

模型结构	GN	BN	学生模型
平均准确率	98.06%	90.02	88.11%

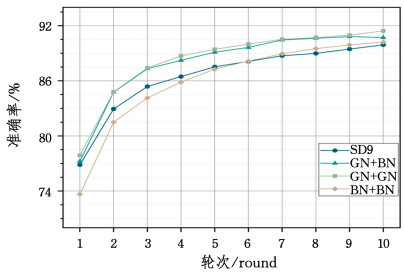
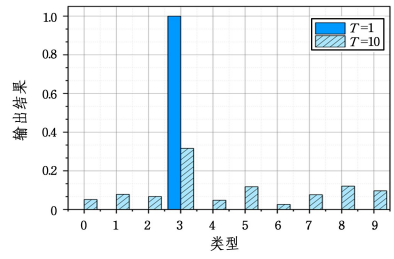


图 7 3 组教师网络对比

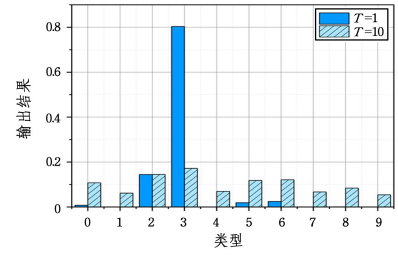
Fig. 7 Comparison of three groups of teacher networks

组成教师网络的节点的平均质量越高，对蒸馏训练学生模型产生的效果就越好。为优质节点分配更高的聚合权重进行联合训练任务，可以提高蒸馏训练学生模型的质量。

取两组不同的模型对随机抽取的同一张图片在 $T=1$ 与 $T=10$ 时的 logits 输出进行对照实验。图 8(a)给出了准确率为 98.03%、参数量与数据集规模较大的模型输出结果；图 8(b)给出了准确率为 75%、参数量与数据集规模较小的模型输出结果。通过对比可得，图 8(a)输出结果间的关联性更强。图 9 给出了使用多权重节点评估算法生成软标签与使用 HBMD-FL[16] 的平均算法 AVG 生成软标签的输出结果。在优劣模型输出分布接近时，两种算法并无明显差异；在优劣模型输出差异较大时，AVG 在错误类别“6”的输出结果为 7.35%，多权重节点评估算法输出为 5.24%，对比减少了 2.11%。采用 BFDS 的评估算法分配权重生成的软标签保留了优质模型输出结果所具有的类别间的关联性，对于不同的类别具有更好的辨识度，受到较差模型输出结果的影响小于使用 AVG 生成的软标签，证明了算法的有效性。



(a) Image 1



(b) Image 2

图 8 模型输出对比

Fig. 8 Model output comparison

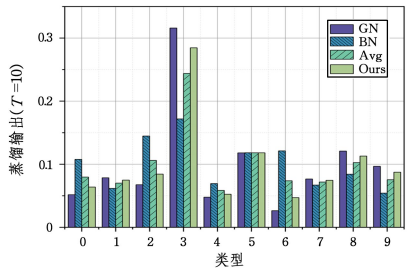


图 9 输出结果对比

Fig. 9 Comparison of output results

结束语 本文提出了一种基于区块链的联邦蒸馏数据共享模型，解决了多方联合训练场景下的数据孤岛问题。利用区块链技术构建分布式教师网络，在保证原始数据不出本地的前提下，将本地输出结果用于蒸馏训练轻量化学生模型，避免了单点故障，减小了参数规模，降低了部署成本，提高了训练过程的可信度。本文提出了一种节点评估与权重分配算法，保证更好地分配聚合权重，减小低质量节点带来的负向影响。实验结果表明，在该模型中各参与方可以实现数据知识的安全可信共享，发布方与参与方均能经由 BFDS 蒸馏训练本地模型，同时验证了本文算法在促进全局软标签聚合过程中的可信性与安全性。在下一步的工作中，将进一步探索对任务发布方数据集的隐私保护，为评估模型中各方的贡献设计平衡的激励机制，以提高参与方的积极性。

参 考 文 献

[1] SAMUEL A L. Some studies in machine learning using the game of checkers [J]. IBM Journal of Research and Development, 2000, 44(1/2): 206-226.

[2] MCMAHAN H B, MOORE E, RAMAGE D, et al. Federated learning of deep networks using model averaging [J]. arXiv: 1602.05629, 2016, 2.

[3] ZHOU X, XU M, WU Y, et al. Deep model poisoning attack on federated learning[J]. Future Internet, 2021, 13(3): 73.

[4] LIU W，TANG C K，MA J，et al. Application Research and Progress of Blockchain in Privacy Computing [J]. Journal of Zhengzhou University (Natural Science Edition)，2021，38(3)：675-679.

[5] YUAN Y，WANG F Y. Development status and prospect of blockchain technology[J]. Acta Automatica Sinica，2016，42(4)：481-494.

[6] MORA A，TENISON I，BELLAVISTA P，et al. Knowledge Distillation for Federated Learning：a Practical Guide[J]. arXiv：2211.04742，2022.

[7] GEOFFREY H，ORIOL V，JEFF D. Distilling the knowledge in a neural network[J]. arXiv：1503.02531，2015.

[8] WU C，WU F，LYU L，et al. Communication-efficient federated learning via knowledge distillation[J]. Nature Communications，2022，13(1)：2032.

[9] XU Z，HSU Y C，HUANG J. Training shallow and thin networks for acceleration via knowledge distillation with conditional adversarial networks[J]. arXiv：1709.00513，2017.

[10] CHEN M，ZHANG L，MA T Y. Recommendation Approach Based on Attentive Federated Distillation [J]. Journal of Software，2021，32(12)：3852-3868.

[11] SUN Y H，SHI Y H，LI M，et al. Personalized Federated Learning Method Based on Collation Game and Knowledge Distillation[J]. Journal of Electronics & Information Technology，2023，45(10)：3702-3709.

[12] CHENG X M，DENG C H. Compression Algorithm of Face Recognition Model Based on Unlabeled Knowledge Distillation[J]. Computer Science，2022，49(6)：245-253.

[13] MO Z，GAO Z，ZHAO C，et al. FedDQ：A communication-efficient federated learning approach for Internet of Vehicles[J]. Journal of Systems Architecture，2022，131：102690.

[14] YAO D，PAN W，DAI Y，et al. Local-global knowledge distillation in heterogeneous federated learning with non-iid data[J]. arXiv：2107.00051，2021.

[15] LI N，ZHANG R，ZHU C，et al. A data sharing method for remote medical system based on federated distillation learning and consortium blockchain[J]. Connection Science，2023，35(1)：2186315.

[16] LI Y，ZHANG J，ZHU J，et al. HBMD-FL：Heterogeneous Federated Learning Algorithm Based on Blockchain and Model Distillation[C]// Emerging Information Security and Applications：Third International Conference (EISA 2022). Cham：Springer Nature Switzerland，2023：145-159.

[17] MENG X F，LIU F，LI G，et al. Review of Knowledge Distillation in Convolutional Neural Network Compression[J]. Journal of Frontiers of Computer Science and Technology，2021，15(10)：1812-1829.

[18] LI H. Statistical learning methods[M]. Beijing：Tsinghua University Press，2019.

[19] ZHU L G，LIU Z J，HAN S. Deep Leakage from Gradients [C]// Neural Information Processing Systems. 2019.

[20] ZHOU C Y，CHEN D W，WANG S，et al. Research and Challenge of Distributed Deep Learning Privacy and Security Attack [J]. Journal of Computer Research and Development，2021，58(5)：927-943.

[21] HE Y Z，HU X B，HE J W，et al. Privacy and Security Issues in Machine Learning System：A Survey[J]. Journal of Computer Research and Development，2019，56(10)：2049-2070.

[22] QIU X Y，YE Z C，CUI X L，et al. Survey of communition overhead of federated learning[J]. Journal of Computer Applications，2022，42(2)：333-342.

LIU Wei，born in 1981，Ph.D，associate professor，Ph.D supervisor，is a member of CCF(No. 49811M). His main research interests include blockchain technology，privacy protection and smart healthcare.

TIAN Zhao，born in 1985，Ph.D，associate professor，is a member of CCF (No. K7436M). His main research interests include blockchain technology，information security，and intelligent transport.

（责任编辑：喻藜）