计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.191000126

一种可用于数据和模型分享的模型链

闫凯伦[1] 张继连[1,2]
1 广西师范大学计算机科学与信息工程学院 广西多源信息挖掘与安全重点实验室 广西 桂林 541004
2 暨南大学网络空间安全学院 广州 510632
(victory_yan@foxmail.com)

摘 要 机器学习开始在越来越多的行业中得到应用,但使用机器学习执行任务的软件一直受限于第三方软件商更新模型。文中基于区块链,将训练神经网络消耗的算力和区块链的工作量证明机制相结合,提出并实现了模型链。模型链作为一种可用于分享数据和机器学习模型的区块链,基于骨架网络训练神经网络模型,以全网节点匿名分享的数据作为训练模型的数据集,实现了不依赖第三方更新神经网络模型。模型链使用环签名来保护用户数据隐私,节点训练的模型使用统一的测试集评估,通过评估的模型将作为节点的工作量证明用于投票达成一致共识。文中提出了两种可行的激励机制,即物质奖励和模型奖励。对于潜在的威胁,如账本分析、脏数据攻击和欺骗投票,给出了相应的解决方案.实现了一个用于数字识别的模型链。实验结果表明,模型链中的模型可以适应实际场景下发生的用户变迁和数据变化。

关键词:区块链;投票共识机制;数据分享;工作量证明;神经网络
中图法分类号 TP311

Model Chain for Data and Model Sharing

YAN Kai-lun[1] and ZHANG Ji-lian[1,2]
1 Guangxi Key Lab of MIMS,College of Computer Science and Information Engineering,Guangxi Normal University,Guilin,Guangxi 541004,China
2 College of Cyber Security,Jinan University,Guangzhou 510632,China

Abstract Machine learning has been applied in more and more scenarios,but software that employs machine learning to perform tasks depends on third-party to update the models. This paper proposes and implements a model chain by utilizing computation power of training neural network consumption with proof-of-work. As a blockchain that can be used to share data and machine learning models,the data shared anonymously by the whole network node are used in the model chain,and the neural network model is explored based on the primary network,thus realizing neural network model update without relying on the third-party. The shared data are signed with a ring signature to protect local data privacy. The whole network uses the same test set to evaluate the model,and the adopted model can be regarded as proof-of-work. This paper proposes two reward mechanisms,i.e.,material reward and model reward. To deal with potential threats,e.g.,blockchain ledger analysis,dirty data attacks and fraudulent voting,this paper proposes ideal ring signature scheme and several solutions. Finally,extensive experiments on real data are conducted,and the results show that the model in the model chain can adapt to the user changes and data changes.

Keywords Blockchain,Voting consensus mechanism,Data Sharing,Proof-of-work,Neural network

1 引言

近年来,神经网络的快速发展和优异表现让一些日常软件,如输入法[1]、购物软件[2],开始使用机器学习来改善用户体验。软件公司训练机器学习模型的数据一般来自用户收集或公开数据集。收集用户的数据存在用户隐私泄露的风险,而且训练过程中一般未考虑数据时间序列属性,训练的模型也不能适应用户的动态变化[3]。同时,机器学习模型的更新完全依赖软件公司,存在使用软件的用户群体已经变更,但软件中的模型未及时更新的情况。在保护用户隐私的前提下,

到稿日期:2019-10-21 返修日期:2020-04-23 本文已加入开放科学计划(OSID),请扫描上方二维码获取补充信息。
基金项目:国家自然科学基金(61932011,61972177,61877029,62020106013);广西多源信息挖掘与安全重点实验室开放课题(MIMS18-09);广东省计算机网络重点实验室开放课题(CCNL201903)
This work was supported by the National Natural Science Foundation of China(61932011,61972177,61877029,62020106013),Guangxi Key Lab of MIMS (MIMS18-09) and Communication and Computer Network Lab of Guangdong(CCNL201903).
通信作者:张继连(zhangjilian@jnu.edu.cn)

不需要依靠第三方更新机器学习模型，是一个亟待解决的问题。

以区块链为基础的比特币自诞生便得到了工业界和学术界的广泛关注和研究[4]。区块链的去中心化、可编程等特点让用户可以不依赖第三方共同维护机器学习模型。比特币使用工作量证明作为共识机制，保证了区块链的安全性和一致性[5]，但在比特币中大量算力被用于计算无意义的哈希值。由于训练神经网络模型也需要大量算力[1]，因此可以将机器学习的算力消耗作为工作量证明，从而把区块链和机器学习进行有机结合。

基于以上所述，本文提出了一种可用于数据和模型分享的区块链——模型链(Model chain)，在保护节点隐私的前提下分享数据，将节点训练的模型作为工作量证明，全网节点投票达成共识，以实现在对等网络下不依赖第三方更新神经网络模型。

2 研究背景和现有工作

目前，区块链和机器学习结合的关注点是如何通过智能合约的方式交易机器学习模型，如 Algorithmia 推出 DanKu 协议用于在以太坊上对机器学习模型进行评估和交易[6]，但需要用户提供数据集，且并未解决不依靠第三方更新模型的问题。机器学习和区块链的交叉应用主要面临模型选择、数据集来源、共识机制等问题。本节将针对这些问题，介绍相关的研究背景和现有工作。

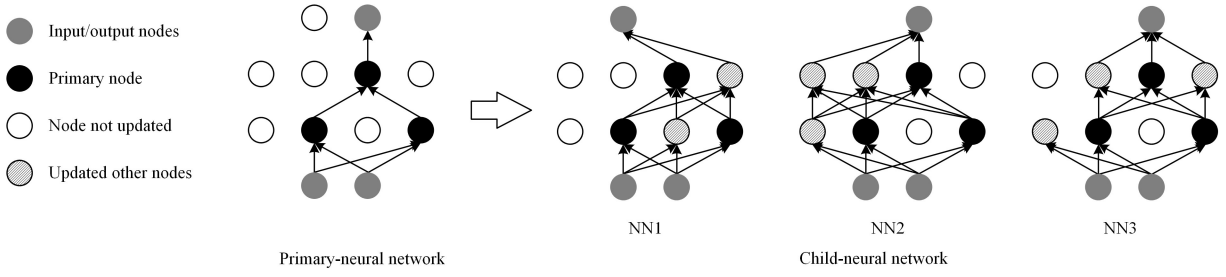
2.1 具有隐私保护的数据分享

一般而言，某个应用场景中收集的数据具有关联度高、时序化等特点[3]。在实际应用中，大部分软件、网站都会收集用户数据用于数据分析或改进软件，例如，谷歌输入法收集用户数据，使用用户设备进行分布式机器学习以改善输入法体验[1]；亚马逊根据用户购物习惯推荐商品[2]。同样地，模型链中用户分享的数据可以作为数据集，用于模型的训练和评价。

虽然区块链中的公钥具有一定的匿名性，但仍存在通过对公开账本进行分析来侵犯用户隐私的风险[7-8]，而把用户分享的数据直接暴露在区块链中更增加了上述可能性。环签名(Ring signature)作为一种数字签名方案[9]，具有匿名性、不可伪造、签名者可以自由决定匿名范围、无需可信第三方等特点。模型链作为分享链通过环签名对分享的数据进行签名，有效地破坏了数据和用户间的关联。用户在分享数据时可以通过改变数据时间戳、打乱有序数据、保留敏感数据来保护隐私。通过环签名和可控的分享策略，其他用户可以对数据的正确性进行签名验证，但无法知道该数据属于哪个公钥拥有者以及确切的产生时间，实现了具有隐私保护的数据分享。

2.2 基于骨架神经网络的探索训练

神经网络存在一定的冗余性，可以寻找一个规模比它小得多的子网络来替代当前模型，以获得相同的模型表达能力[10]。文献[11]提出一种针对神经网络模型的随机划分方法，通过模型压缩和删减[12]选取骨架网络作为训练的基础，同时每个节点选取一些其他神经节点进行探索训练，如图 1 所示。



注：3 个子网络基于同一个骨架神经网络进行探索

图 1 神经网络随机划分

Fig. 1 Random partitioning of a neural network

模型链将骨架神经网络与区块链相结合，即区块链中分布式节点基于骨架神经网络进行探索训练。模型链中的节点需要在限定时间内训练出神经网络模型。区块数量的增多保证了模型的更替，从而让链中的模型更适应实际场景下发生的概念迁移和概念演化。

2.3 投票共识

共识机制是区块链的核心，常见的共识机制有 PoW，PoS，DPoS，PBFT 等[13]。其中比特币采用的 PoW 是得到最广泛认可的共识机制。比特币的工作量证明机制本质上是节点通过计算一个特定哈希值作为工作量证明来获得记账权，使得全网达成一致，从而保证了区块链的安全。

模型链采用单节点单票制对区块进行投票，得票最多的区块作为节点一致的新区块，节点自行打包新区块并添加到区块链中，从而解决了节点作恶和偷懒问题。无代价的投票容易产生“女巫攻击”[14]，因此在模型链中投票是有代价的。节点训练的结果在当前测试集上的评价要优于上一个区块最优模型在当前测试集上的评价才可以被其他节点认可。以上一个区块中最优模型在当前区块测试集上的评价为基准评价，即节点需要训练出满足基准评价的模型才可以作为工作量证明进行投票，节点在分享模型的同时进行投票，不产生额外的通信开销。

2.4 星际文件系统

星际文件系统(InterPlanetary File System，IPFS)[15]是一种对等分布式文件系统，使用哈希表来解决数据传输和查找问题。模型链把数据文件对应的哈希值存储到区块链中作为索引，使用 IPFS 来存储数据文件实体，既实现了区块链的轻量化也保证了数据安全。目前已有很多相关工作把区块链和 IPFS 相结合[16]，限于篇幅，本文不详细阐述 IPFS 在区块链上的存储方案。

3 模型链系统

本节主要介绍模型链中节点如何实现数据分享和投票达成全网共识。

3.1 系统架构

模型链面向使用神经网络模型执行任务的设备终端(下述为节点)。图 2 给出了模型链的架构，节点分享的数据和模型通过区块链保证安全。节点训练的模型只对网络中的节点负责，使用统一的测试集对训练结果进行评价，将通过评价的模型作为节点的工作量证明。节点间通过带工作量证明的选票选出一致区块。一个区块的生成主要包括数据收集和模型竞争两个阶段，数据收集阶段包括数据分享和数据划分，模型竞争阶段包括模型分享和区块同步。

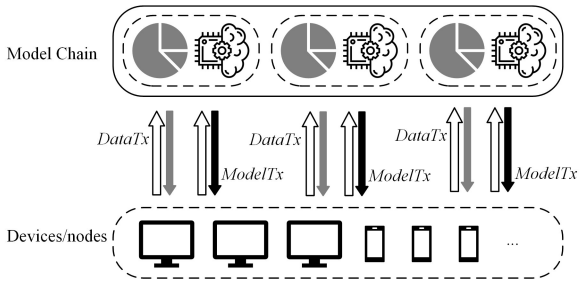


图 2 设备/节点通过模型链分享数据和模型，使用模型链中的模型执行本地预测或回归任务(图中省略了 IPFS 分布式存储部分)

Fig. 2 Devices/nodes share data and models through model chain, and the models in model chain can be used for local prediction or regression tasks (IPFS-based entity storage is omitted for brevity)

3.2 数据收集

3.2.1 数据分享

节点通过数据事务(Data transaction)进行数据分享，一个数据事务由事务哈希值、数据哈希值、时间戳、环签名、公钥集合组成，如图 3 所示。

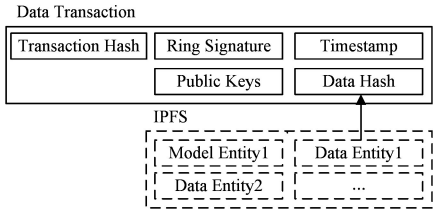


图 3 一个数据事务的组成

Fig. 3 Composition of data transaction

假设某节点有公钥 pk 和私钥 sk，已使用公钥的集合为 UPK，未使用的公钥集合为 PK，节点通过数据事务 $DataTx$ 分享数据 $data$，数据的哈希值为 H_d，时间戳为 ts。节点从 PK 中随机选取一组公钥 PKG，使用 PKG 和私钥 sk 对数据事务 $DataTx$ 进行环签名，签名为 sig。

算法 1 数据事务分享

1. PKG=getPKG(PK,UPK,n) / * 获取签名需要的公钥集合 UPK * /
2. add(pk,PKG) / * 添加本节点公钥 * /
3. sig=signature(H_d,ts,PKG,sk) / * 签名 * /
4. DataTx=pack(H_d,ts,PKG,sig) / * 打包 * /
5. push(DataTx) / * 广播 * /

数据事务的接收遵循以下两个准则：

准则 1 如果收到数据事务的时间戳超过当前时刻 T 的一定范围 L，则不接受数据事务；

准则 2 节点开始划分初集后，不再接收划分时刻 T_d 之前的数据事务。

其中，时间 L 和时刻 T_d 依据数据划分情况而定。准则 1 是为了防止节点把数据划分到下一个区块，准则 2 是为了防止不诚实节点在划分阶段声称未分享的数据是已分享的，从而破坏数据集划分。节点对接收到的数据事务 $DataTx$ 验证签名 sig 的合法性，通过验证的数据事务会保存到当前区块的初集 G 中。节点还会收集事务环签名使用的公钥集合 PKG 中的新公钥，添加到本地未使用的公钥集合 PK 中。算法 2 给出了数据事务的接收过程。

算法 2 数据事务的接收

1. verifyDataTx(DataTx) / * 验证事务 * /
   1.1. if(ts−T>L or ts < T_d) / * 是否符合准则一、准则二 * /
   1.2. exit / * 结束 * /
   1.3. if(!verifySig(H_d,ts,PKG,sig)) / * 验证签名 * /
   1.4. exit
2. add(DataTx,G) / * 保存到初集 * /
3. push(DataTx) / * 转发 * /
4. updatePK(PKG,UPK,PK) / * 添加新公钥到未使用公钥集合中 * /
   4.1. for every pk^* in PKG do
   4.2. if (notIn(pk^*,UPK) and notIn (pk^*,PK)) / * 如果 pk^* 不在使用过的也不在未使用的公钥集合中 * /
   4.3. add(pk^*,PK)
   4.4. end for

3.2.2 数据划分

在节点数据划分之前，先讨论分布式环境中随机种子的选取。选取方式有链内和链外两种：

(1)链内选取可以使用基于公开可验证密码的共享方案(PVSS)[17]，使用 PVSS 实现的分布式可验证随机数生成器有 RandHound，RandHerd[18]，Ouroboros[19]。该方法的缺点是系统需要依赖复杂的交互过程。

(2)链外选取可以避免复杂的算法，但依赖可信的随机源。有区块链使用超级大乐透开奖结果作为随机种子，这种方法对于模型链而言频率过低。比特币大约每 10 min 产生一个区块，区块哈希值和具体出块时间都是未知的，因此非常适合作为模型链中数据划分的随机种子[20]。

模型链中区块的数据集包括初集、训练集和测试集，数据事务中的数据在未划分前属于初集。每个区块中测试集大小是确定的，由网络中节点数量和上一个区块数据集的大小决定。数据划分阶段会进行多次初集划分，初集每次随机划分出训练集的一个子集和测试集的一个子集。

使用不可预测、公开可验证的随机种子进行数据划分可防止单节点的自私行为，保证了数据集的随机性和安全性，同时也避免了节点间为达成共识进行大量通信。从上一个测试集收集结束到当前测试集收集完成称为一个数据收集阶段。一个区块的测试集收集完成后，马上进入本区块的模型竞争阶段。

由于网络延迟等原因，存在数据事务没有及时广播到全网节点造成数据集划分不一致的问题。节点接收到随机种子后会往前推一段时间作为“缓冲时间”L_d，准则 2 中的划分时刻 T_d 称为“决断时刻”。节点只对数据事务时间戳小于决断时刻的数据进行划分，时间戳大于决断时刻的数据事务在下一个初集划分。缓冲时间可以有效解决节点数据划分不一致的问题。本文用比特币区块的哈希值作为随机种子，将比特币区块的时间戳提前 1 min 作为决断时刻 T_d，缓冲时间 L_d 为 1 min，准则 1 中的时间 L 设置为 10 min。

3.3 模型竞争

3.3.1 模型分享

节点使用模型事务(Model transaction)进行模型分享，具体组成如图 4 所示。训练的模型和训练集都是由节点自行决定，不同的节点可以选择不同的骨架网络探索训练。模型事务由 3 个部分组成，事务的基本部分包括事务哈希值、时间戳、签名和公钥；模型部分包括模型的哈希值、模型评价和模型训练集的 Merkle root，这是模型事务的工作量证明；投票部分包括上一个区块哈希值和当前区块数据集 Merkle root。

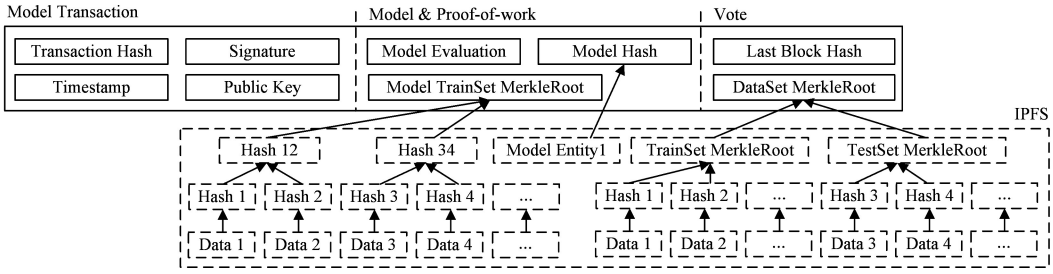


图 4 模型事务的组成

Fig. 4 Composition of model transaction

3.3.2 区块同步

模型链采用单节点单票制对区块进行投票，得票最多的区块作为一致新区块，节点自行打包新区块到模型链中，因此单节点无法作恶和偷懒。一个区块包括区块哈希值、上一个区块哈希值、区块最优模型哈希值(即对应的模型事务哈希值)、时间戳和投票结果的 Merkle root。时间戳规定为最后一个初集划分的决断时间。相同提名的选票组成一个 Merkle tree，此 Merkle tree 的根即代表此种提名所有投票的选票。然后，这些 Merkle tree 的根依据包括选票的数量递减排序，再作为叶子节点，组成一个新的 Merkle tree，作为最终的投票结果，保存在区块中。

节点训练出模型后会使用本节点数据集中划分出的测试集进行评价，这可视为节点对当前区块数据集投票的保证。节点的骨架模型和训练数据来自当前区块和之前区块，模型在测试集上的评价作为节点的工作量证明，这可视为节点对区块投票的保证。节点接收到模型事务并通过合法性验证后，会进行两次计票。

第 1 次投票：对模型事务中的数据集进行计票，选出获得最多选票的数据集 D^*；

第 2 次投票：将第 1 次投票中投给数据集 D^* 的模型事务视作有效票，对有效票中上一个区块哈希值计票，选出得票最多的上一个区块。

在第 2 次投票中，每种提名依据选票数量递减排列，作为叶子节点保存在 Merkle tree 中。每种提名中，选票依据模型的评价结果递减排序，并保存在 Merkle tree 中，因此第一种提名中第一个选票的模型自然为区块最优模型，其数据集为区块统一的数据集。投票 2 保证了全网节点的模型链一致。

在当前区块 B_i 模型竞争阶段，有节点 node1 分享的模型事务 $ModelTx$。$ModelTx$ 包括上一个区块哈希 HB_{i-1}、数据集 D、训练集 TR、测试集 TE、模型 m 等。某网络良好的诚实节点 node2 及时接收了所有合法事务。节点 node2 有上一个区块中的最优模型 m^*、数据集 D^*(有测试集 $TE^* \subseteq D^*$)，基准评价 $Es=eval(m^*, TE^*)$，即上一个区块最优模型在当前区块测试集上的评价)，投票 1 集合 $V_D=\{v_1, v_2, \cdots, v_m\}$，投票 2 集合 $V_B=\{v_1, v_2, \cdots, v_n\}$，节点 node2 对于收到的 $ModelTx$ 使用如下投票算法。

算法 3 模型事务投票算法

```
1. verifyModelTx(ModelTx) / * 验证事务 * /
   1.1. if(!verifySig(ModelTx) or eval(m,TE*)<Es) / * 验证签名
        正确性、模型是否满足基准评价 * /
   1.2. exit / * 结束 * /
2. push(ModelTx) / * 转发 * /
3. vote(D,V_D) / * 投票 1 * /
4. order(V_D) / * 对投票 1 中提名降序排序 * /
5. vote(HB_{i-1},V_B) / * 投票 2 * /
6. order(V_B) / * 对投票 2 中提名降序排序 * /
```

实际应用中节点并不知道自己是否及时接收到所有的数据事务，因此节点在收到一些模型事务并确定可能性最大的数据集后，再开始验证每个模型事务中的模型评价，然后进行归票。

下一个区块数据收集阶段结束时，马上结束当前区块的模型竞争阶段，开始下一个区块新的模型竞争。各节点会根据投票结果同步上一个区块，然后自行打包当前区块并添加至区块链中。

3.4 分享和激励

为得到对本节点有利的模型，节点趋向于分享数据，只有分享的数据才会被其他节点用于训练或作为测试数据。节点希望其拥有的数据尽可能多地划分到测试集中，这样可以筛选出对其有利的模型。在公有链环境下，可以引入激励机制来防止不进行训练的寄生节点，因此模型链的激励主要表现在节点物质奖励和模型奖励两方面。

代币：代币是一种简单有效的物质激励手段，在比特币、以太坊等区块链中广泛使用。通过代币奖励训练出优秀模型

的节点可以有效鼓励节点参与模型训练。

热度：热度机制是一种类似于 BitTorrent 中 tit-for-tat 的数据交换协议[21]，即奖励相互贡献的节点，惩罚那些没有参与训练的节点。为了获得有利于本节点的模型，节点需要分享本节点的数据，以获得在数据集中更大的占比。训练出最优模型的节点会获得大量的热度，对于热度高的节点，其他节点会趋向于与其交换数据。热度机制可以减少寄生节点的数据在数据集中的占比，从而训练出对高热度节点有利的模型。因此，节点不进行模型训练就难以得到对本节点有利的模型。

3.5 模型链的安全问题

根据以上论述，模型链面临的安全威胁主要有以下几个方面。

(1)账本分析

文献[8]指出，可以通过追踪单节点在账单中的交易记录等方式推断出用户，因而侵犯了用户的隐私。在比特币公开账本中追踪单一公钥的所有交易是可行的。在模型链中，由于每个数据存在多个公钥，根据公钥查找对应的数据是不可行的。因此通过分析模型链账本推断出用户身份的难度要远大于普通的区块链。对于单个数据而言，使用 n 个公钥进行签名(其中包括本节点的公钥)，攻击者猜出数据所属公钥的概率为 $1/n$。对于链中所有的数据而言，存在通过对整个账本分析得到数据和节点关系的风险，因此我们给出签名建议：节点对数据事务环签名时，对不同的数据事务应使用多个新公钥进行环签名，让数据集中数据环签名的公钥集交叉相关，以期模型链中区块的数据集达到“理想状态”，从而减少上述风险。

例 1　模型链中有 2 个节点分别分享了 2 组数据，每组数据 d_i 用本节点的私钥和另一个其他节点公钥进行环签名，环签名的公钥集合为 $\{pk_i, pk_j\}$。在数据集 D 中，环签名在理想状态下有如下数据和公钥集：

$$P=\{\langle d_1,\{pk_1,pk_2\}\rangle,\langle d_2,\{pk_2,pk_3\}\rangle,\langle d_3,\{pk_3,pk_4\}\rangle,\langle d_4,\{pk_1,pk_4\}\rangle\}$$

在非理想状态下有：

$$P=\{\langle d_1,\{pk_1,pk_2\}\rangle,\langle d_2,\{pk_1,pk_2\}\rangle,\langle d_3,\{pk_3,pk_4\}\rangle,\langle d_4,\{pk_3,pk_4\}\rangle\}$$

对于集合中的数据，在非理想状态下可以猜测数据 d_1 和 d_2 可能同属于 pk_1 或 pk_2，数据 d_3 和 d_4 可能同属于 pk_3 或 pk_4。而在理想状态下则不存在这种关系。

(2)脏数据攻击

脏数据是一种攻击分享链的方法，即攻击者通过分享大量无意义的数据来破坏数据集。在无激励机制的情况下，节点进行攻击付出的代价极小。在具有激励机制的情况下，攻击者则需要付出大量代价。例如在热度机制下，节点分享的数据是有限制的，攻击者需要大量的算力训练出满足基准的模型，只有提高恶意节点的热度才有可能对数据集进行破坏。恶意节点为获得热度而训练合法模型的行为反而在一定程度上对模型链是有利的。同时由于链中数据和模型存在时效性，这使得模型链具有自我恢复能力，提高了污染数据集的难度。

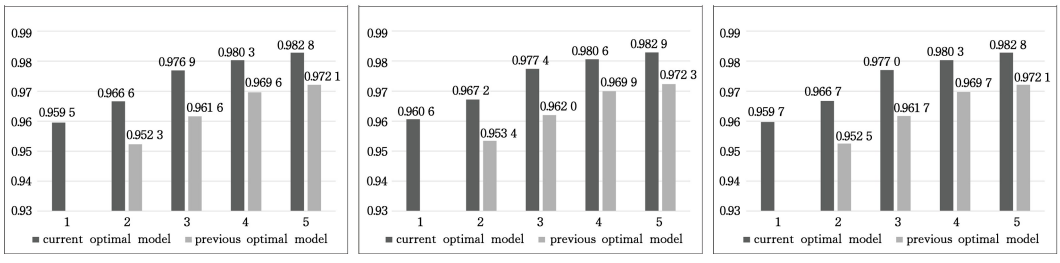
(3)欺骗投票

攻击者有可能通过构造特定的测试集、减少训练模型的工作量来进行欺骗投票。节点的自私性让节点倾向于选择本节点数据占比较多的数据集，环签名混淆了节点和数据间的关系，加大了攻击者构造选票的难度。在良好的网络环境、理想环签名的情况下，攻击者成功控制一个区块需要拥有多于诚实节点的合法选票。

4 实验和评价

本文使用 JAVA 实现提出的模型链，并将其应用于数字识别。由于模型链需要基于用户的时序化数据集，而现有的数据集如 mnist[22] 无法满足要求，我们使用 5500 种计算机字体生成数据集。将每种计算机字体视作一个用户字体，字体的拉伸、倾斜、缩放等变换视作用户手写体的噪声。在每轮区块生成中，对 2 200 种字体进行 25 次变换，每个数字生成 55000个 28 * 28 像素的单通道图片，其中 85% 作为训练数据，15% 作为测试数据。新区块产生后，使用全新的字体替换掉一半旧字体，再次生成新一轮的数据集，模拟模型链中用户的动态变化。

节点模型训练使用 deeplearning4j 框架，参考文献[23]设置了一个卷积神经网络作为骨架网络，同时设置了 10 个节点进行探索训练。每轮区块生成后选出最优模型，在下一轮模型训练时，节点会参考上一个区块的训练结果对模型进行调整并重新开始训练。实验总共进行了 5 轮区块生成，采用了 Accuracy，Precision，F1 Score 作为评价指标，图 5 为每个区块中最优模型在测试集上的评价。



(a)Accuracy　(b)Precision　(c)F1 Score

图 5　在 5 轮区块生成中，最优模型和上一区块最优模型在每个区块测试集上的评价

Fig. 5　Evalution of optimal model and previous block optimal model in each block test set in 5 rounds of block generation

从图中可以看到，随着数据集的改变(即实际情况下用户的动态变化)，上一个区块的最优模型在新数据集中的预测结果明显下降，最优模型的评价较前一个区块最优模型的评价有明显提升。如在第二个区块中，第一个区块的最优模型准确率从 96.06% 下降到95.34%，而得益于基于骨架神经网络模型的探索训练，此区块中的准确率提升到 96.72%。

在时间效率方面，每个区块最优模型的训练时间分别为 391s，471s，420s，515s，605s。模型验证平均耗时为 4.52s，平

均每轮投票共识共计花费 90.40 s。由于模型竞争阶段持续时间较长，因此节点有足够的时间进行模型训练和投票。目前单节点对整个测试集进行验证，如果多节点合作，每个节点分别验证测试集的一部分，则可以大幅减少模型验证耗时，提高共识效率。

在实际应用中，模型链中的节点不断探索新模型来替代之前产生的模型，当链中用户群体变化不大时，链中的模型趋于稳定。当链中用户群体变化较大时，链中的模型精度会先下降(由最优模型的泛化性决定)，然后随着模型链中的区块不断增多逐渐趋于稳定。在 5 轮区块的生成过程中，最优模型的评价不断提高，说明模型链中多节点模型探索有较好的效果，不仅能适应数据集的动态变化，还提高了模型的评价。

结束语 本文提出了一种可用于分享数据和模型的模型链，通过节点分享数据模型实现了不依赖于第三方的模型更新。带工作量证明的投票共识保证了模型链的安全，链中时序化的数据和基于骨架网络的探索训练可以适应实际场景下发生的用户迁移。目前，模型链适用于存在丰富用户群体、面向个人用户的软件应用场景，对于企业级的用户，可以通过引入智能合约来提高其通用性。本文下一步的工作是研究节点合作训练模型的可能性，以训练更加庞大复杂的神经网络模型。同时我们将进一步完善模型链的激励机制，研究 IPFS 存储和激励机制相结合的可能性，即让部分节点付出存储代价，而另一部分节点付出计算代价。

参 考 文 献

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

[2] LINDEN G, SMITH B, YORK J. Amazon.com recommendations: Item-to-item collaborative filtering[J]. IEEE Internet computing, 2003 (1): 76-80.

[3] ZHISONG P, SIQI T, JUNYANG Q, et al. Survey on Online Learning Algorithms[J]. Journal of Data Acquisition and Processing, 2016, 31(6): 1067-1082.

[4] NAKAMOTO S. Bitcoin: A peer-to-peer electronic cash system [EB/OL]. Bitcoin. -URL: https://bitcoin.org/bitcoin.pdf.

[5] ZHENG Z, XIE S, DAI H, et al. An overview of blockchain technology: Architecture, consensus, and future trends[C]// 2017 IEEE International Congress on Big Data (BigData Congress). IEEE, 2017: 557-564.

[6] KURTULMUS A B, DANIEL K. Trustless machine learning contracts: evaluating and exchanging machine learning models on the ethereum blockchain[J]. arXiv: 1802.10185, 2018.

[7] ANDROULAKI E, KARAME G O, ROESCHLIN M, et al. Evaluating user privacy in bitcoin[C]// International Conference on Financial Cryptography and Data Security. Berlin, Heidelberg: Springer, 2013: 34-51.

[8] CONTI M, KUMAR E S, LAL C, et al. A survey on security and privacy issues of bitcoin[J]. IEEE Communications Surveys & Tutorials, 2018, 20(4): 3416-3452.

[9] RIVEST R L, SHAMIR A, TAUMAN Y. How to leak a secret [C]// International Conference on the Theory and Application of Cryptology and Information Security. Berlin, Heidelberg: Springer, 2001: 552-565.

[10] TIEYAN L, WEI C, TAIFENG W, et al. Machine Learning: Distributed Algorithms, Theory, and Practice[M]. Beijing: China Machine Press, 2018.

[11] SHIZHAO S, WEI CH, JIANG B, et al. Slim-DP: A Multi-Agent System for Communication-Efficient Distributed Deep Learning [C]// Proceeding of the 17th International Conference on Autonomous Agents and Multiagent Systems (AAMAS). 2018.

[12] HAN S, POOL J, TRAN J, et al. Learning both weights and connections for efficient neural network[C]// Advances in Neural Information Processing Systems. 2015: 1135-1143.

[13] YUAN Y, NI X C, ZENG S, et al. Blockchain consensus algorithms: the state of the art and future trends[J]. Acta Automatica Sinica, 2018, 44(11): 2011-2022.

[14] DOUCEUR J R. The sybil attack[C]// International workshop on peer-to-peer systems. Berlin, Heidelberg: Springer, 2002: 251-260.

[15] BENET J. Ipfs-content addressed, versioned, p2p file system [J]. arXiv: 1407.3561, 2014.

[16] CHEN Y, LI H, LI K, et al. An improved P2P file system scheme based on IPFS and Blockchain[C]// 2017 IEEE International Conference on Big Data (Big Data). IEEE, 2017: 2652-2657.

[17] SCHOENMAKERS B. A simple publicly verifiable secret sharing scheme and its application to electronic voting[C]// Annual International Cryptology Conference. Springer, Berlin, Heidelberg, 1999: 148-164.

[18] SYTA E, JOVANOVIC P, KOGIAS E K, et al. Scalable Bias-Resistant Distributed Randomness[C]// 2017 IEEE Symposium on Security and Privacy (SP). IEEE, 2017.

[19] KIAYIAS A, RUSSELL A, DAVID B, et al. Ouroboros: A provably secure proof-of-stake blockchain protocol[C]// Annual International Cryptology Conference. Springer, Cham, 2017: 357-388.

[20] BONNEAU J, CLARK J, GOLDFEDER S. On Bitcoin as a public randomness source[J]. IACR Cryptology ePrint Archive, 2015, 2015: 1015.

[21] COHEN B. Incentives build robustness in BitTorrent [C]// Workshop on Economics of Peer-to-Peer Systems. 2003: 68-72.

[22] LECUN Y, BOTTOU L, BENGIO Y, et al. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11): 2278-2324.

[23] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. Imagenet classification with deep convolutional neural networks[C]// Advances in neural information processing systems. 2012: 1097-1105.

YAN Kai-lun, born in 1994, postgraduate. His main research interests include blockchain and machine learning.

ZHANG Ji-lian, born in 1977, Ph.D, associate professor, is a member of China Computer Federation. His main research interests include data management, information security and machine learning.