基于联邦学习的动态信任评估身份认证方法

石瑞生[1,2], 付 彤[1], 林子丁[1], 兰丽娜[3], 姜 宁[1]

[1]北京邮电大学 网络空间安全学院 北京 中国 100876
[2]北京邮电大学 可信分布式计算与服务教育部重点实验室 北京 中国 100876
[3]北京邮电大学 人文学院 北京 中国 100876

摘要 随着云计算、BYOD(Bring your own device)的流行, 企业信息系统呈现出开放与动态互联的特征, 这种趋势使得基于动态信任评估的零信任安全架构开始取代基于边界信任的一次性身份认证模式, 成为工业界与学术界关注的研究热点。动态信任评估模型为零信任架构提供持续信任评估的能力, 可以对企业信息系统的安全性和隐私性进行有效的保护。然而, 训练动态信任评估模型面临两个现实挑战: 1)很多企业的用户异常登录行为数据很少, 影响模型的训练效果, 导致信任评估模型准确性不高, 不利于身份认证系统的可靠性; 2)用户行为数据中包含着用户的隐私信息, 泄漏用户隐私的法律风险使得企业不愿意共享用户异常登录行为数据。针对这些问题, 本文提出了一种基于联邦学习的动态信任评估身份认证方法, 使得各个平台在不泄漏原始用户数据的情况下达到联合训练模型的目的, 进而提高各平台身份认证系统的安全性。在假设各个平台提供了用户的行为原始数据的前提下, 本方案会根据不同特征的实际含义提取离散型用户行为数据的统计学特征, 并选取与风险用户相关性高的特征。为了保证数据安全性和训练数据的规模, 本方法采用联邦学习技术联合多个企业进行训练, 从而得到动态信任评估层的核心模型, 其误识率和拒识率相较于单一平台有了一定的提升。通过该方案, 身份认证系统可以在不泄露用户敏感信息的情况下, 对用户身份进行有效评估, 进而提升身份认证系统安全性和用户体验。本文还对不同的支持横向联邦学习的机器学习算法应用于动态信任评估模型的效果进行了比较, 实验结果表明了在基于联邦学习的动态身份认证模型中使用 SVM 作为机器学习训练方法的效果优于其他机器学习训练方法。最后, 本文从安全性和隐私性的角度出发还对动态信任评估系统自身以及联邦学习带来的安全性和隐私性的影响做了讨论。

关键词 联邦学习; 动态信任评估; 网络安全; 身份认证
中图法分类号 TP309.2 **DOI 号** 10.19363/J.cnki.cn10-1380/tn.2025.03.03

Research on Dynamic Trust Evaluation Method Based on Federated Learning

SHI Ruisheng[1,2], FU Tong[1], LIN Ziding[1], LAN Lina[3], JIANG Ning[1]

[1] School of Cyberspace Security, Beijing University of Posts and Telecommunications, Beijing 100876, China
[2] Key Laboratory of Trustworthy Distributed Computing and Service, Ministry of Education, Beijing University of Posts and Telecommunications, Beijing 100876, China
[3] School of Humanities, Beijing University of Posts and Telecommunications, Beijing 100876, China

Abstract With the popularity of cloud computing and Bring Your Own Device (BYOD), enterprise information systems exhibit open and dynamic interconnection features. This trend has led to the replacement of the one-time identity authentication mode based on boundary trust with the zero-trust security architecture based on dynamic trust evaluation, which has become a research hotspot in industry and academia. The dynamic trust evaluation model provides continuous trust evaluation for the zero-trust architecture, which can effectively protect the security and privacy of enterprise information systems. However, training dynamic trust evaluation models faces two practical challenges: 1) many enterprises have limited abnormal login behavior data, which affects the training effectiveness of the model and leads to low accuracy of the trust evaluation model, which is not conducive to the reliability of the identity authentication system; and 2) user behavior data contains users' privacy information, and the legal risk of privacy leakage makes enterprises unwilling to share abnormal login behavior data. To address these issues, this paper proposes a dynamic trust evaluation method based on federated learning, which enables various platforms to achieve joint training of the model without leaking the original user data, thereby improving the security of identity authentication systems on various platforms. Assuming that various platforms

通讯作者: 石瑞生, 工学博士, 副教授, Email: shiruisheng@bupt.edu.cn。

本课题得到北京市自然科学基金(No. M21037)、国家重点研发计划项目(No. 2022YFF0902701)、工业与信息化部“2022 年工业互联网公共服务平台-面向工业互联网的虚拟货币挖矿治理公共服务平台项目”、教育部“虚拟货币挖矿行为监管平台研发与应用重大攻关项目”、中国-中东欧国家高校联合教育项目(No. 2022172)、高等学校学科创新引智计划(No. B21049)、北京邮电大学研究生创新创业项目(No. 2025-YC-T020)资助。
收稿日期: 2023-04-02; 修改日期: 2023-07-25; 定稿日期: 2025-01-08

provide users' raw behavioral data, this approach extracts statistical features of discrete user behavior data based on the actual meaning of different features and selects features with high relevance to risky users. To ensure data security and training data scale, this method uses federated learning technology to train multiple enterprises together to obtain the core model of the dynamic trust evaluation layer, achieving 0.205 system false acceptance rate and 0.192 system false rejection rate, with improved accuracy compared to a single platform. Through this approach, the identity authentication system can effectively evaluate user identity without leaking sensitive information, thereby improving system security and user experience. This paper also compares the effects of different machine learning algorithms supporting horizontal federated learning applied to dynamic trust evaluation models. The experimental results show that using SVM as the machine learning training method in the dynamic identity authentication model based on federated learning is more effective than other methods. Finally, this paper discusses the security and privacy impact of the dynamic trust evaluation system itself and the federated learning from the perspective of security and privacy.

Key words federated learning; dynamic trust assessment; network security; identity authentication

1 引言

云计算、BYOD 的流行, 已经改变了企业计算环境, 传统的基于边界信任的一次性身份认证模式面临着巨大挑战, 以账号劫持为代表的攻击模式成为企业信息安全面临的主要威胁之一。基于动态信任评估的零信任安全架构为身份认证提供了新的方向。动态信任评估模型为零信任架构提供持续信任评估能力, 是对抗账号劫持等日益猖獗的有力手段[1]。

谷歌的零信任模型 BeyondCorp 主要包括信任推断系统、设备库存服务、访问控制引擎、访问策略、网关和资源, 其中信任推断系统可以通过执行持续的信任评估来检测设备状态变化和改变设备信任级别[2]。

但是, 对于大多数企业来说, 训练信任评估模型的主要困难在于由于用户的异常行为数据样本在训练数据集中非常稀疏, 难以达到好的训练效果。因此, 这些企业希望通过联合训练来提高模型的性能。

然而, 数据隐私保护的规定使得数据的流通和共享受到较大限制。欧盟于 2016 年 4 月通过了《通用数据保护条例》(General Data Protection Regulation, DGPR), 并于 2018 年 5 月 25 日生效。该条例严格约束了个人隐私数据的收集、传输、保留和处理, 未经用户同意擅自将各方用户数据集中处理的行为被禁止, 这限制了拥有数据的平台相互共享数据。训练动态信任评估模型需要将不同用户的输入数据映射到向量空间中, 并在测试时接收新的用户数据输入。然而, 由于训练时需要直接访问用户的数据, 比如用户姓名、地址、电话, 即使是经过加密, 这些信息仍涉及到用户的隐私敏感信息, 因此这为多方平台联合训练造成了困难。

首先, 训练数据的不足导致模型的误识率受到影响, 导致非法用户的登录请求被错误地接受, 给合法用户的安全造成威胁。同时, 异常行为数据的稀疏性使得机器学习算法难以学习有效的模式, 合法用户被错误地拒绝登录, 进而降低用户的体验。其次, 用户行为数据可能包含噪声, 这些噪声会影响机器学习算法的准确性和模型的性能。此外, 如何将用户行为数据转化为结构化数据并提取有效的特征也是一个具有挑战性的问题。最后, 在身份认证场景中, 选择适合的算法也需要进一步研究。

为解决以上挑战, 本文提出基于联邦学习的动态信任评估方法, 在保证用户数据隐私前提下, 联合多个平台数据建模分析实现企业之间互联互通, 降低对非法用户(非法用户是指使用没有被授权登录的用户)检测的误识率, 并结合身份认证场景对模型的性能指标进行全面的分析。

本文的主要贡献有 3 个方面:

(1) 针对训练数据不足的问题, 提出了一种基于横向联邦学习与多因子认证的动态信任评估方法。在多个平台之间不共享用户行为数据的前提下, 构建了一个安全性更高、可靠性强、用户体验更好的动态身份认证系统。

(2) 针对用户行为数据集的特征提取问题, 设计了离散型行为数据的特征提取方案, 用于完成本地模型的训练, 得到了 0.205 的系统误识率和 0.192 的系统拒识率。

(3) 本研究在真实数据集上通过比较使用不同支持横向联邦学习的机器学习算法在动态信任评估方案的效果, 发现使用 SVM 作为机器学习训练方法比其他方法表现更好。

本文第 2 节介绍了基于机器学习的身份认证和联邦学习的相关工作; 第 3 节详细描述了基于联邦学习的动态信任评估方案; 第 4 节中具体地介绍了实验的过程、使用的数据集以及对离散型数据的处理、采用的方法和使用的评价指标等; 在第 5 节对实验结果进行了总结并对其反映出来的现象进行了分析; 第 6 节对方案的安全性和隐私性进行了讨论。第 7 节对全文进行了总结, 并提出了未来的研究方向。

2 相关工作

2.1 基于机器学习的身份认证

近年来机器学习技术的迅速发展, 为动态信任评估中用户行为特征的认证提供了新方法。我们可以通过机器学习的技术对用户的行为特征进行学习, 从而实现动态信任管理。已有许多基于用户行为认证的研究。我们调研了 2011 年至今的在系统安全, 移动计算, 人机交互或身份验证模式识别领域的国际会议或期刊上有代表性的论文, 共计 28 篇[3-30], 发现在当前研究中还存在以下问题:

(1) 大部分论文(25 篇)所选用的特征都是连续的特征, 而非离散特征。所谓的连续型特征是指一段时间内有序列的特征, 例如加速度传感器特征, 陀螺仪传感器特征或者是步态特征, 眼球特征等。对于这些连续性特征, 虽然能获得较好的认证性能, 但这些特征的获取相对困难, 需要额外的设备, 同时这些特征的通用性也较差, 例如和移动设备相关的特征无法在个人电脑上进行复用。而离散型特征, 例如 IP 地址, 用户设备, 登录时间等具有获取方便, 更具有通用性的特点。

在其余的 3 篇论文中, H. Gomi 等人[29]对雅虎的 1000 名用户的登录行为特征, 提出了一种认证方法, 由于样本容量较小, 并且对数据的处理以及所使用的模型较为粗糙, 准确率较低。Yang Y 等人[30]虽然获得了较高的准确率, 但其中使用了连续特征, 并且某些离散特征会侵犯用户隐私, 例如通话记录。D. Freeman 等人[31]提出了一种仅使用了用户的 IP 地址与用户 Agent 的认证方案, 但由于特征较少, IP 地址与用户 Agent 也容易被获取, 因此存在一定的安全性问题。

因此, 已有方法大部分是依靠连续性行为特征来进行身份认证, 而缺乏仅靠离散型用户的行为特征来进行身份认证, 在特征获取及认证的准确性、数据安全性上还有待提高。

(2) 在身份认证与机器学习结合的过程中, 对于身份认证系统性能指标的分析也是关键工作之一。而在我们调研的论文中, 大部分论文只是展示了自己的方案在不同场景下识别风险用户的实验情况, 并没有对风险用户的行为数据进行分析。因此, 对身份认证系统性能指标的分析有待增强。

2.2 联邦学习

联邦学习在 2016 年被谷歌提出的分布式机器学习框架, 可以支持跨多个参与客户端训练全局共享的模型, 并保证训练数据留在本地。联邦学习旨在从分散的、孤立的数据中训练机器学习模型从而解决数据孤岛问题[42]。联邦学习可以实现数据保留在本地, 这样可以不泄露隐私也不违反法规; 另外利用多个参与方联合建模, 利益共享, 多个参与方贡献各自的数据建立并共享公共模型成果[43]。实际场景中, 联邦学习可以在保护本地数据的同时, 为多个企业建立统一的模型, 帮助企业以数据安全为前提, 共享数据。

Yang 等人根据参与方之间的数据分布不同将联邦学习划分为横向联邦学习、纵向联邦学习和迁移联邦学习, 其中基于样本的联邦学习是横向联邦学习, 其数据集特征往往重叠度大[42]。由于局部数据在相同的特征空间中, 各方可以使用具有相同特征的局部数据来训练局部模型, 再通过平均所有的局部模型来更新全局模型。

因为动态信任评估的场景中多个平台中的用户登录行为有大量相同的数据特征可被应用于模型训练, 因此在该场景中可以使用同构数据, 本文采用横向联邦学习。支持横向联邦学习的主要框架有 FATE、FedML、PySyft 等。其中 PySyft 最早由 Ryffel 等人[37,41]提出, 由 OpenMined 开发, 是一个安全和隐私深度学习 python 库。TFF 是谷歌推出的基于 TensorFlow 深度学习框架的联邦学习框架[39]。而 PySyft 不仅支持 TensorFlow 框架, 还支持 PyTorch 框架。与 FATE、FedML、TFF 相比, PySyft 提供更多种可选的隐私机制, 包括同态加密、安全多方计算和差分隐私。PySyft 同时支持应用在 web、移动设备和边缘网络上。PySyft 开源库有灵活的自由度。基于 GitHub 上的流行程度, PySyft 是目前机器学习社区中最具影响力的联邦学习系统, 可以应用于移动终端机器学习、人工智能、数据中心等场景, 并且该框架支持非独立同分布数据。

3 基于联邦学习的动态信任评估方案

针对单一平台风险用户样本缺少的问题, 本文提出了基于联邦学习的动态信任评估方案。该方案利用联邦学习, 在多个平台之间不共享用户行为数据的前提下, 构建一个误识率低、安全性高的动态身份认证系统。本节将介绍方案的动态信任评估总体结构和联邦学习模型的训练过程。

3.1 动态信任评估方案架构

基于联邦学习的动态信任评估方案总体架构由五部分构成, 分别是动态信任评估层, 二次认证层, 特征抽取层, 特征处理层, 模型训练层和认证结果记录, 如图 1 所示。具体功能如下:

(1) 动态信任评估层: 动态信任评估层是系统认证的核心模块之一。我们将已经训练完的联邦学习模型存储在该层。动态信任评估层对输入的特征进行预测, 并输出预测结果。

(2) 二次认证层: 二次认证层的功能是当认证结果判断该用户行为是不安全行为后, 会启动二次认证, 例如手机验证码, 邮箱验证等显式验证, 通过二次验证来判断用户行为是否合法。

(3) 特征抽取层: 经过二次认证层的数据, 包括用户提交的表单数据, 结合表单提交的时间、IP、设备编号等信息中抽取用户行为特征数据, 同时加入历史特征数据, 共同进入特征处理层。

(4) 特征处理层: 特征处理层是动态信任管理的核心模块之一。特征处理层的功能在于将原始特征进行特征扩展与离散化处理, 作为本地模型的训练输入。

(5) 模型训练层: 模型训练层利用抽取得到的用户行为特征数据进行联邦学习离线训练, 训练得到的模型将用于动态信任评估层, 模型具体介绍见3.2 节。

(6) 认证结果记录: 记录层的功能是记录每次登录的结果, 我们会将认证结果写入数据库, 一方面可以用于分析用户的登录画像, 另一方面作为历史记录, 便于后续用于用户非法登录或非法交易的调查取证。

用户通过登录系统, 将行为特征输入动态信任管理层的全局模型, 并进行预测输出, 系统主动提取用户异常数据的特征, 并对用户特征进行处理与扩展, 再根据结果进行后续处理, 整个过程对于用户而言是透明友好的。

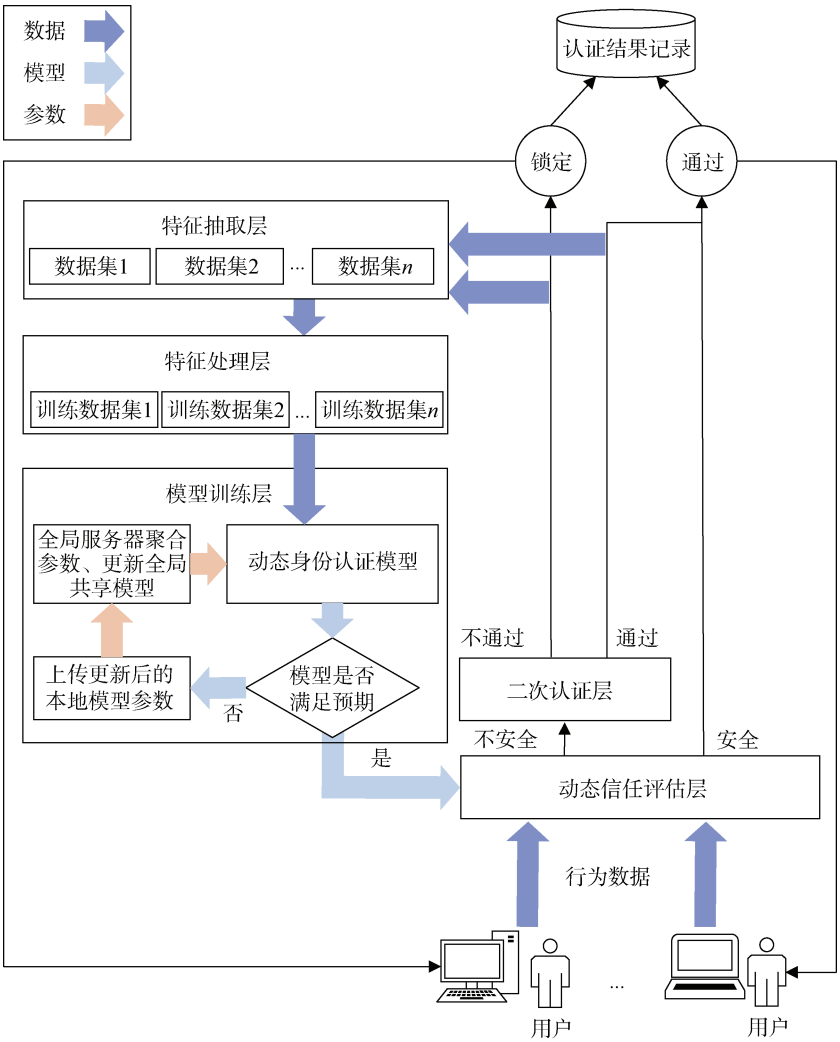


图 1 动态信任评估系统架构图

Figure 1 Dynamic trust evaluation system architecture

系统架构图如图 1 所示。首先用户在平台进行登录等行为, 行为数据进入动态评估层通过训练好的模型评估该用户是否为合法用户, 将对合法用户给予通过, 对不合法的用户做二次验证, 如果二次

认证是合法用户则通过, 如果依旧没有通过则标识为非法用户将该用户锁定。进入二次认证层的数据被认定为异常数据, 该类数据将带有锁定或通过的结果进入特征抽取层, 通过特征获取层提取到登录时间、登录 IP、设备编号等信息, 之后采集到的特征进入特征处理层做特征扩展和离散化处理。经过处理后的数据进入模型训练层, 在该层利用本地数据进行训练之后将训练得到的模型参数上传至中央服务器对模型进行计算、共享和更新, 将在全局更新后的参数传回到本地模型中, 如果该模型误识率满足预期则输出该模型并将该模型存储到动态评估层作为对用户行为动态认证评估核心模型。

3.2 联邦学习模型离线训练过程

本文所使用的数据来自京东公司的真实交易和登录记录[32], 包括交易行为数据集和登录行为数据集[33], 其中记录了大量的异常用户数据。

本文使用模型的训练过程是在模型训练层完成的, 并且训练完成的联邦学习模型将存储在动态信任评估层, 动态信任评估层的核心是联邦学习实验模型。在模型的构建上我们采用横向联邦学习方法。横向联邦学习是一种数据共享模型, 在该模型中不同本地平台共享相同的特征空间, 但有不同的数据样本。实际场景中不同平台的用户行为数据具有多样性且不具有相同的概率分布(即非独立同分布), 但不同平台的用户访问行为特征仍具有相似性(即同构), 比如登陆的次数、登陆设备标志的众数等。因此我们的研究假设在不同的本地平台中的用户行为数据集特征相似, 且使用相同的特征处理方式。这样不同平台用于模型训练的用户行为数据集就是同构且非独立同分布数据集。

联邦学习框架中需要两种类型的模型, 一种是本地模型, 另一种是全局模型, 这两种模型共享着相同的模型网络结构, 但不能相互访问数据。本地模型部署在本地平台(如不同电商网站), 并可以访问各自的私有本地数据, 我们假设所有参与的本地平台都是诚实的。全局模型是在系统认证核心服务平台的联邦学习全局模型部署。我们假设全局模型所在的中央服务器属于一个可信的第三方, 并保证其不会对本地平台进行攻击。

联邦学习一般的训练过程如图 2 所示。首先, 本平台使用本地数据集进行训练, 得到参数 $\hat{w}_m$。每个迭代轮次中, 本地模型的梯度和损失会被加密并发送到全局模型。中央服务器协调梯度计算, 这里使用 FedAvg[44]的梯度平均算法计算平均梯度, 并将聚合后的模型更新广播发送给所有本地模型。以 FedAvg 的梯度平均算法为例子计算平均后的梯度: $w=(\hat{w}_1+\hat{w}_2+\hat{w}_3+\cdots+\hat{w}_m)/m$, 再将聚合后模型的更新广播发送给所有的本地模型, 本地模型解密梯度后再对各自的本地模型进行更新。

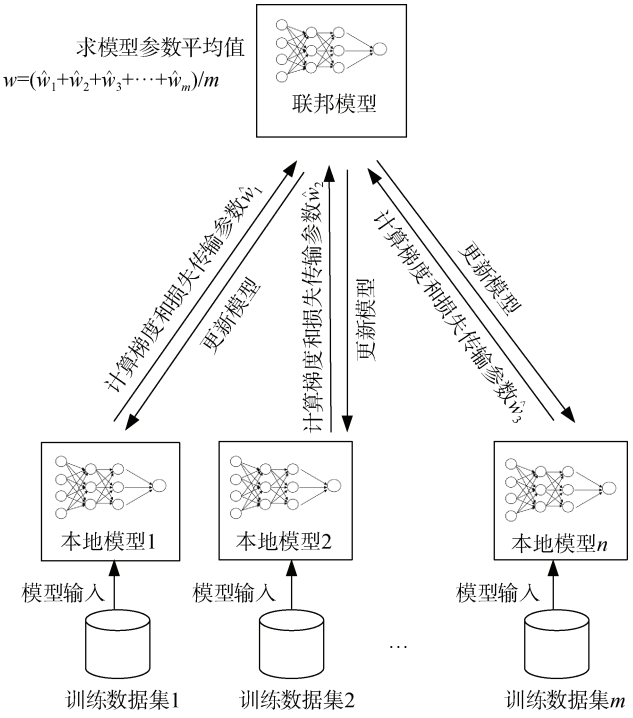


图 2 联邦学习模型训练过程示意图

Figure 2 Schematic of the federated learning model training process

联邦平均(FedAvg)[44]是一种标准的联邦学习算法。FedAvg 是对梯度下降法(SGD)的联邦优化, SGD 中每轮通信只进行一次批量梯度计算。这种方法在计算上是有效的, 但需要非常多轮的训练才能产生好的模型, 因此对 SGD 的变体 FedSGD[44]被提出。当每次使用本地服务器数据集进行训练, 本地训练次数为 1, 之后对梯度进行聚合的联邦学习算法被称为 FedSGD。FedAvg 算法如下所示, 本地模型在其本地数据上采样训练并更新模型, 训练中本地模型对本地数据采样并可以通过 $w \leftarrow w-\eta\nabla l(w;b)$ 来进行多次本地迭代更新, 中央服务器在每一轮聚合中使用 $w_{t+1} \leftarrow \sum_{k=1}^{K}\frac{n_k}{n}w_{t+1}^k$ 来对梯度聚合更新, 并对上传到中央服务器的各个模型的参数做平均。与 SGD 及其变体相比, FedAvg 可以执行更多的本地计算和更少的通信, 从而实现简单且高效的联邦学习算法。

算法 1 FedAvg: K 个客户端以 k_i 为索引; B 为本地训练的批量大小, E 为本地 epoch 数, η 为学习率。C 用来控制全局批量大小

FedAvg:

Server: Initialize w_0

Server: $m \leftarrow \max(C \cdot K,1)$

for each global round $t=1,2,...$ **do**

Server: $S_t \leftarrow$ (random set of m clients)

Server: Send w_t to clients $k \in S_t$

Clients $k \in S_t$ **in parallel do:** $w_{t+1}^k, n_k \leftarrow$ ClientUpdate(k,w_t)

Server: $w_{t+1} \leftarrow \sum_{k=1}^{K} \frac{n_k}{n} w_{t+1}^k$

ClientUpdate(k, w): // Run on client k

$B \leftarrow$ (split P_k into batches of size B)

for each local epoch i from 1 to E do

for batch $b \in B$ do

$w \leftarrow w - \eta \nabla l(w;b)$

return w to server

4 动态信任评估模型训练方法

4.1 数据集

本文的用户行为数据集包括交易行为数据集和登录行为数据集[33]。交易行为数据集中共有 132720 条记录, 其中有 3642 条是风险交易记录; 登录行为数据集中共有 595646 条记录。由于每个用户通常都有多条交易记录和登录记录, 交易记录中被平台记录为风险交易记录和无风险交易记录, 一个用户有多条交易记录, 并且一个用户可能既有无风险交易记录, 又有风险交易记录。在数据预处理过程中, 经过筛选, 我们发现总共有 26044 个用户的相关数据, 其中 24958 个用户的交易记录全是无风险, 1086 个用户的交易记录中有风险记录, 占总用户量的 4.2%。我们将交易记录全是无风险的用户标记为“正常用户”, 将交易记录有风险记录的用户标记为“风险用户”。

本文特征分为登录行为特征与交易行为特征两大类。

登录行为特征主要包括用户登录账号的相关特征, 具体特征如表 1 所示。为了避免噪声数据对模型的影响, 我们没有将登录时间戳(*timestamp*)、日志 id(*log_id*)以及用户 id(*id*)这些无实际含义的特征加入我们的特征集合中。另外, 在实际测试和分析中, 我们发现登录所花费的时长(*timelong*)对于模型的准确性起到了负向作用。经过深入分析, 我们发现该参数受网络影响最大, 例如用户登录的网络可能存在波动, 或者在弱网络进行登录。因此, 我们也没有将该特征加入我们的特征集合中。

表 1 登录特征

Table 1 Login features

	特征	含义
1	*log_id*	日志 id
2	*id*	用户 id
3	*log_from*	登录来源
4	*device*	登录设备标识
5	*timelong*	登录所花费的时长
6	*ip*	登录 ip
7	*city*	登录 ip 归属地
8	*result*	登录结果
9	*timestamp*	登录时间戳
10	*type*	登录类型
11	*is_scan*	是否扫码登录
12	*is_sec*	是否使用安全控件

交易行为特征包括交易时间、交易主键、用户 id 以及是否是风险交易。具体特征如表 2 所示, 由于“*row_key*”(交易主键)和“*id*”(用户 id)没有实际含义, 因此我们也没有将这些特征加入特征集合中, 同时, “*is_risk*”也是我们的目标特征。

表 2 交易特征

Table 2 Trade features

	特征	含义
1	*time*	交易时间
2	*row_key*	交易主键
3	*id*	用户 id
4	*is_risk*	是否是风险交易

4.2 离散型特征提取方法

用户的原始行为数据维度包含大量的冗余数据会对模型的误识率造成很大的影响, 因此需要对本研究的数据进行行为特征处理。通过探索数据的分布、变化等情况, 确定哪些特征对于解决问题最为关键, 以此进行筛选和处理, 去除那些对问题解决没有意义的特征, 提高预测的准确性。

不同于传统手机传感器的连续特征(例如加速度和陀螺仪特征), 本文所选的特征都是离散的, 例如登录 IP、登录城市和登录设备。因此, 我们需要对这些离散型行为特征进行处理。我们采用以下四种数据统计方法的组合方式来进行特征处理, 以最大程度地提取这些离散型行为特征的有效信息:

(1) nunique 值, 即统计唯一值的个数。例如对于

[*a*,*b*,*c*]的 nunique 值就是 3, 因为唯一值有 *a*,*b*,*c* 三个, 对于[*a*,*a*,*b*,*c*]的 nunique 值就是 2, 因为唯一值只有 *b*,*c*。

(2) 众数, 是指在统计分布上具有明显集中趋势点的数值, 代表数据的一般水平。换句话说, 在数据中出现次数最多的数值就是众数, 例如[*a*,*a*,*b*,*c*]的众数就是 *a*。

(3) TopK 值, 是指在数据中出现次数第 K 个多的数值, 例如对于[*a*,*a*,*a*,*b*,*b*,*c*], Top2 就是 *b*, 代表出现次数第 2 多的值为 *b*。

(4) mean、max、min、std 值, 是指一组数值数据的平均值、最大值、最小值以及方差。通过这些指标, 可以刻画这组数据的分布情况。

基于以上的统计学方法, 我们根据不同特征的情况, 进行相应的处理。首先, 我们对用户的常用设备、城市和 IP 进行处理。对于一般用户, 我们认为其最常用的两个设备(一台电脑和一台手机)以及最常用的两个 IP 地址和对应的地区最能代表其常用习惯。因此, 我们取“*device*”(登陆设备标识)、“*city*”(登录 ip 归属地)和“*ip*”(登录 ip)这三个参数的众数和 Top2 值。

其次, 我们对“*is_scan*”(是否扫码登陆)和“*is_sec*”(是否使用安全控件)计算众数。由于这两个参数为二值参数, 即只有 true 和 false 两个值, 因此取其众数可代表用户的最常用习惯。

同时, 我们对“*device*”、“*log_from*”(登录来源)、“*ip*”(登录 ip)、“*city*”(登录 ip 归属地)、“*result*”(登录结果)和“*type*”(登录类型)取 nunique 值, 以表征用户的常用习惯。

其次, 我们对于“*is_scan*”,“*is_sec*”计算众数, 由于“*is_scan*”和“*is_sec*”为二值参数, 即只有 true 和 false 两个值, 因此对于他们取众数可以代表用户的最常用习惯。

同时, 我们对“*device*”,“*log_from*”,“*ip*”,“*city*”,“*result*”,“*type*”取 nunique 值, 同样也是来表征用户的常用习惯。

最后, 针对“*time*”(交易时间)这个特征我们将对同一个 ID 的用户前后交易以小时和秒进行时间差计算, 并且分别计算交易小时时间差的 mean, max, min, std 值。同理也对“*time*”(登录时间)以小时和秒进行时间差计算, 并且计算登录小时时间差的 mean, max, min, std 值, 这样可以来表示用户的登录和交易时间的间隔。例如用户的登录习惯是每天登录一次, 如果用户在一天内登录次数过多或者长时间不登陆再登录的时候, 我们就会认为该用户的登录行为存在异常情况, 对于交易行为也是同样的。

我们还将统计用户的累计登录次数和累计交易次数, 同样, 如果用户在一段时间内的交易或者登录次数过多, 我们也将认为用户的行为存在异常情况。

经过处理后的特征值如表 3 所示:

表 3 处理后的特征

Table 3 Processed features

	特征	含义
1	*log_from_nunique*	登陆来源的 nunique 值
2	*device_nunique*	登录设备标志的 nunique 值
3	*ip_nunique*	登录 IP 归属地的 nunique 值
4	*city_nunique*	登录 IP 归属地的 nunique 值
5	*result_ nunique*	登录结果的 nunique 值
6	*type_ nunique*	登录类型的 nunique 值
7	*is_scan_mode*	是否扫码登陆的众数
8	*is_sec_mode*	使用安全控件的众数
9	*device_mode*	登录设备标志的众数
10	*ip_mode*	登录 IP 的众数
11	*city_mode*	登录 IP 归属地的众数
12	*device_top2*	登录设备标志的 top2 值
13	*city_top2*	登录 IP 归属地的 top2 值
14	*ip_top2*	登录 IP 的 top2 值
15	*login_cnt*	登录次数
16	*trade_cnt*	交易次数
17	*trade_time_diff_seconds*: mean,max,min,std	交易时间差(s)的 mean, max, min, std 值
18	*trade_time_diff_hour*:mean,max,min,std	交易时间差(h)的 mean, max, min, std 值
19	*login_time_diff_seconds*: mean, max,min,std	登录时间差(s)的 mean, max, min, std 值
20	*login_time_hour*:mean, max,min,std	登录时间差(h)的 mean, max, min, std 值

根据以上特征处理, 我们将选择合适的特征作为模型的输入, 并进行实验。Pearson 相关系数可以用来衡量两个连续型变量之间的线性相关程度, 其范围为[–1,1], 正值表示正相关, 负值表示负相关, 绝对值越大表示线性相关程度越高。在本文中, 我们计算处理后的特征与数据集的目标特征(是否有风险)之间的皮尔森系数, 以此给每个特征赋予一个重要性分数, 如图 3 所示。

通过分析用户行为特征和目标特征之间的相关性, 我们发现登录 IP 的归属地数量(如表 3 中的特征 4)、交易时间差(如表 3 中的特征 17、18)、登录结果数量(如表 3 中的特征 5)、登陆来源的 nunique 值(如表 3 中的特征 1)、登录的小时时间差和秒时间差(如表 3 中的特征 19、20)、是否扫码登陆的众数(如表 3

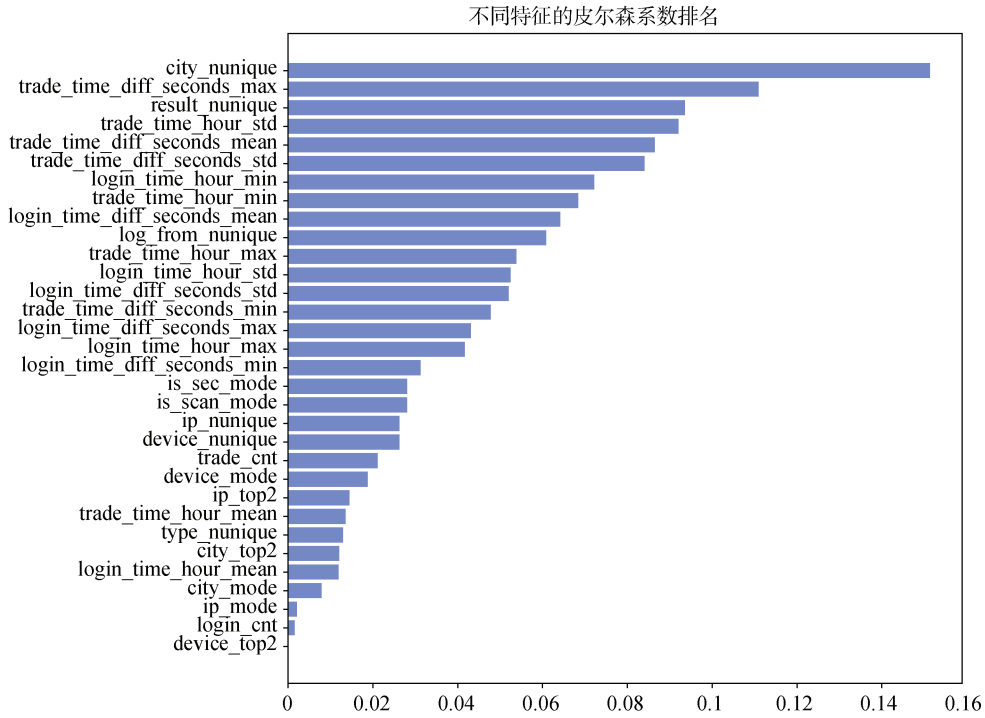


图 3　不同特征的皮尔森系数排名

Figure 3　Ranking of Pearson coefficients for different features

中的特征 7)、登录 IP 归属地的 nunique 值(如表 3 中的特征 3)等特征对于判断用户是否有风险的影响更加显著。我们还使用了皮尔逊相关系数来计算每个特征与目标特征之间的关联程度, 并根据结果对特征进行排序。在这 32 个特征中, 我们选取前 30 个特征, 并排除后两个特征登录次数和登录设备标志的 top2 值(如表 3 中的特征 15、12)。

皮尔森系数与特征对于目标特征的重要性有关, 通过对特征的皮尔森系数排名分析, 我们发现不同行为特征的皮尔森系数相差较大。用户的行为数据中与时间维度相关的行为特征如特征 17-20 排名很靠前, 十分重要, 但是特征 15 和特征 16 即登录的次数和交易次数对判断非法用户的影响不大。具体分析可以看出与交易相关的特征(如表 3 中的特征 17-18)整体要比登录行为相关的特征(如表 3 中的特征 19-20)重要性更高; 另外用户的地理相关特征中特征 4 也就是登录过的城市的个数对验证用户身份十分重要其余如特征 11、13 即用户出现次数最多的城市和用户出现次数第二多的城市对判断用户的风险性并不重要。总体来说, 我们发现时间维度行为特征对于风险用户的判断要比空间维度行为特征更重要, 交易行为特征对于风险用户的判断要比登陆行为特征更重要。

4.3　联邦学习框架和机器学习模型选取

为了保护用户的隐私安全、避免目标样本过于少, 我们使用了联邦学习的方式来训练动态信任核心模型。本节我们将介绍如何以联邦学习的方式训练动态信任核心模型, 包括本地模型和全局联邦模型。

在本实验中, 我们使用了广泛应用的联邦学习框架 PySyft[41]构建了动态信任层的联邦学习系统。PySyft 是一个 Python 库, 提供了客户端和服务器, 可以帮助优化深度学习模型。在 PySyft 框架中, 我们将用户特征数据集随机分配给不同的客户端, 这种数据随机分配的方式不仅模拟了不同平台拥有的数据, 还能增加数据的多样性, 提高训练效果。在训练过程中, 我们从每个客户端随机抽取一批数据, 然后基于当前联邦学习框架下的机器学习模型的副本使用客户端的数据子集进行训练。在每个 epoch 结束时, 中央服务器将聚合所有参与方的参数进行计算和更新, 以更新联邦学习模型。

在机器学习模型的选择中, 首先, 我们采用基于用户行为的动态信任评估方法来判断用户是否为合法用户, 其本质是一个二分类问题。为了实现这一目标, 我们需要选择一个适合的分类器模型作为我们的机器学习模型。其次, 我们使用了身份认证场景的数据集进行监督学习, 其中每个用户都有合法行为或非合法行为的标签。在机器学习领域, 决策树、朴素贝叶斯分类和支持向量机等是常见的监督学习算法。我们还研究了近年来该领域的相关论文, 并选

择了一些代表性的论文进行研究。在该领域中, 支持向量机(SVM)被广泛用作分类模型, 例如在论文[26]、[27]、[30]、[32-36]、[38]中, 使用不同的改进方法对支持向量机进行了应用。因此, 支持向量机(SVM)是基于行为认证方面具有代表性的模型, 我们选择了它作为本文的机器学习模型。

有研究使用了机器学习或者深度学习的动态信任评估方法, 但并不是本研究机器学习方法的最佳选择。例如: Wang[27]使用深度强化学习结合卷积神经网络的人脸识别方法来实现动态信任评估方法。但是深度强化学习适用于与环境交互来学习的最佳决策问题而卷积神经网络适用于图像、视频和音频的分类、分割等任务, 但不适用于本文的数值型数据。Abuhamad 等人[28]提出了基于用户行为的隐式身份认证方法, 使用长短期记忆网络(Long Short-Term Memory, LSTM)对用户不同时间敲击键盘的行为特征进行训练。由于 LSTM 适合时间序列数据集, 而我们的用户行为数据并非基于时间序列, 因此 LSTM 也不适用于本文的场景。Yang 等人[30]在其身份认证系统中除了支持向量机之外还使用了高斯混合模型(GMM), 但是 GMM 属于无监督聚类算法并不适用于本研究使用的有目标特征的数据集, 因此本文也没有采用该机器学习算法。

在选择机器学习模型时, 我们还需要考虑联邦学习框架的限制。目前, PySyft 主要支持的机器学习模型有线性模型、卷积神经网络和浅层神经网络等, 根据本文的数据集和动态信任评估场景, 我们选用支持向量机(SVM)作为机器学习模型之外还与其他模型进行比较。此外, 我们选择了浅层多层感知机(MLP)和浅层 BP 神经网络(BP)作为对照模型。虽然这些模型在联邦学习框架下的应用存在一定局限性, 但我们经过仔细权衡后, 认为这些模型具有代表性和可比性, 能够在一定程度上体现联邦学习系统的性能和稳定性。

最后, 我们将简单地对我们采用的模型进行介绍:

(1) 支持向量机 (Support Vector Machines, SVM)是一种应用最为广泛、可靠且快速的机器学习算法, 尤其在二分类问题中具有极高的准确性和鲁棒性[34]。该算法不仅可以用于线性数据分类, 还可以处理非线性问题。支持向量机利用统计学习理论以及结构风险最小原则, 最小化经验风险和置信范围, 从而实现最佳的分类效果。其基本模型是定义在特征空间上的间隔最大化线性分类器。支持向量机的关键是最大化样本与分类平面间隔, 这个问题可以通过将其转化为一个凸二次规划问题来求解最优解。

(2) 浅层多层感知机(Shallow MLP)是基于前馈人工神经网络(ANN)的一种分类器, 属于多层感知器(MLP)分类器的一种变体。 MLP 由多个节点层组成, 每个层完全连接到网络中的下一层。输入层中的节点表示输入数据。本研究中的浅层多层感知机具有一个输入层和两个隐藏层。每个隐藏层后面都有一个校正线性单元(ReLU)作为激活函数。最后输出层将特征转换为风险用户二分类。

(3) 浅层 BP 神经网络(BP): BP 神经网络是一种使用误差逆传播算法训练的多层前馈网络, 能够减少计算和存储成本, 同时防止模型过度拟合。本文采用的是一个三层全连接网络, 包括输入层和输出层, 以及隐藏层。输入层包含 30 个节点, 输出层仅包含 1 个节点, 而其余层均为隐藏层。每层节点的输出均使用 ReLU 作为激活函数, 并使用均方误差作为损失函数。

5 实验

我们在本节比较了单一平台上使用机器学习训练和利用联邦学习框架在多方平台上训练的实验结果。我们还对 SVM 作为机器学习方法与 MLP 和 BP 方法的实验结果进行了对照, 并使用了 ROC、AUC、ACC、FAR 和 FRR 等评价标准对实验结果进行比较。

5.1 实验配置

在本章节, 我们介绍了使用用户行为数据分别在单一平台上做机器学习训练和在多方平台利用联邦学习框架做训练以及预测的实验细节, 如图 4 所示。

5.1.1 单一平台上机器学习训练

我们首先实现了基于机器学习的动态信任评估, 并检验其可行性。我们使用该实验作为比较对象, 并与下面的联邦学习实验进行比较。在本实验中, 将训练集的 1/3 的数据作为一个参与方数据集整体来使用。我们分别使用了 SVM、MLP 和 BP 机器学习方法来训练模型。实验在 AMD Ryzen 7 5800H 3.2GHz 的 CPU、4GB 内存、Linux Ubuntu 20.4 系统的计算机上进行, 并训练到在第 10 个 epoch 时, 均达到了最佳精度。

5.1.2 联邦学习机器学习训练

实验采用了 AMD Ryzen 7 5800H 3.2GHz 的 CPU, 4GB 内存, Linux Ubuntu 20.4 系统的计算机, 并且使用 python3.7 编程和 PySyft 0.2.4。PySyft 上唯一可用的优化器是 SGD, 我们将 SGD 的 learning rate 设置为 0.01, 批大小设置为 280。

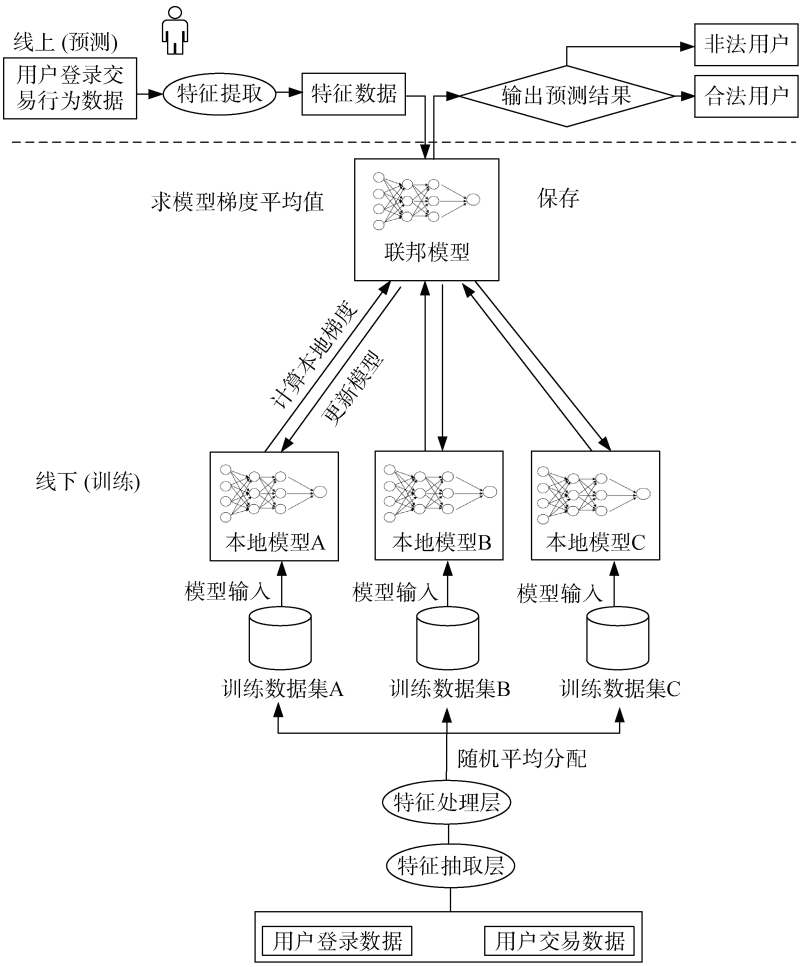


图 4　联邦学习训练与预测示意图

Figure 4　Schematic of federated learning training and prediction

为了检验基于联邦学习的动态信任模型的准确性, 本文从处理后的用户行为数据(共 25589 条数据)中随机选取 2/3 的数据作为训练数据集、1/3 的数据作为测试数据集。我们假设实验中有 3 个参与方参与训练, 重复 3 次, 并取平均值。

(1) 我们首先通过 PySyft 创建 3 个参与方来代表 3 方平台参与联邦学习模型训练, 此外还有一个中央服务器, 它负责对参与方上传的参数做梯度平均。为了保证每个参与方都有相同数量的数据, 我们先将数据集随机平均分配给各个参与方, 然后将相同的初始机器学习模型副本分发给数据集所在的参与方进行训练。

(2) 训练过程中, 在每个 epoch 中, 参与方基于当前的机器学习模型副本使用自己的数据集上训练, 计算梯度并更新模型参数。在每个 epoch 结束时, 各个参与方将模型发送给中央服务器, 该服务器对各个参与方模型的参数求平均值, 然后在该服务器的全局模型中设置其平均参数。在下一个 epoch 开始前, 全局模型副本会传递给各个本地平台, 继续进行下一轮迭代。

(3) 最终经过 10 次迭代后, 我们得到最终的联邦模型, 并将其保存。

通过以上步骤, 我们验证了基于联邦学习的动态信任模型的准确性。在实验中, 我们得到了训练完毕的模型。在预测阶段, 新的输入会通过已经训练好的模型进行预测。具体来说, 当新用户进入系统, 系统将记录其登录和交易的行为, 将其输入到动态信任评估层的模型做预测分类, 并根据输出评估该用户是否为合法用户。

5.2　评价指标

模型的好坏可以从不同的角度来评价。在文献[3]中, 研究者对机器学习中的身份认证性能指标进行了系统的研究, 发现多数论文采用了 ACC、AUC、FAR、TPR 和 FRR 等指标。其中, AUC(Area under the ROC curve)是指 ROC 曲线(Receiver operating characteristic)下方的区域面积, ACC(Maximum accuracy)是指预测的准确率, FAR(False Accept Rate)是指非法用户通过认证的概率, TPR(True positive rate)是指合

法用户通过认证的概率, FRR(False reject rate)是指合法用户认证失败的概率。我们也采用 AUC、ACC、FAR、FRR 四个评价方法作为评价指标, 以更全面地评估和分析我们的模型性能。这些评价方法的具体含义在身份认证系统中已经得到广泛研究[3]:

(1) 准确率(ACC, Accuracy)是指所有预测中预测正确的比例, 其计算公式如公式(1)所示。

$$ACC = \frac{TP + TN}{TP + FN + FP + TN} \tag{1}$$

对于一个二分类问题, 我们可以将样本分成正类(Positive)和负类(Negative)。若一个实样本是正类, 并且被预测为正类, 那么该样本就为真正类(True Positive, TP), 若一个样本是正类, 但是被预测为负类, 那么该样本就为假负类(False negative, FN), 若一个样本是负类, 但是被预测为正类, 那么该样本就为假正类(False Positive, FP), 若一个样本是负类, 并且被预测为负类, 那么该样本就为真负类(True Negative, TN), 其关系如表 4 所示。我们可以设置阈值(Threshold)来判断预测结果是正样本还是负样本。

表 4 混淆矩阵表现形式

Table 4 Confusion matrix

真实值\预测值	正例(通过认证)	反例(认证失败)
正例(用户)	TP	FN
反例(非法用户)	FP	TN

在身份认证系统中, 准确率表示认证系统能够准确识别合法用户和非法用户的比例。但是, 由于正常数据与异常数据是严重不均衡的, 这个指标无法有效评价一个信任评估模型的性能。例如, 考虑一种极端的情况: 由于能够通过口令认证登录的异常行为数据本来就很少, 模型既是简单地将所有样本都判定为正常, 准确率也不会太低。

(2) 误识率(False Acceptance Rate, FAR)是指错误接受率, 为不该接受的样本里但接受的比例, 其计算公式如公式(2)所示:

$$FAR = \frac{FP}{FP + TN} \tag{2}$$

在身份认证系统中, 误识率(FAR)代表非法用户通过认证的比例, 因此 FAR 越小, 代表认证性能越好。

(3)拒识率(False Rejection Rate, FRR)是指错误拒绝率, 为不该拒绝的样本里但拒绝的比例, 其计算公式如公式(3)所示:

$$FRR = \frac{FN}{TP + FN} \tag{3}$$

在身份认证系统中, 拒识率(FRR)代表合法用户认证失败的比例, 因此, FRR 越小, 代表认证性能越好。

(4) 等错误率(Equal Error Rate, EER)是指拒识率(FRR)等于误识率(FAR)时的 FAR 与 FRR 的值, 等错误率(ERR)越低, 代表模型的性能表现越好。在身份认证系统中, 当 FAR 与 FRR 相等时, 代表着用户安全性与便利性的折中, 我们可以通过等错误率来进行折中选择。

在身份认证系统中, 我们不应该过于关注准确率(ACC), 我们应该更关注 FAR 与 FRR 两个指标。因为在实际应用中, FAR 越小代表非法用户通过认证的概率越小, 系统越安全; FRR 越小代表合法用户认证失败的概率越小, 用户体验越好。拒识率高, 会导致用户体验的显著下降。在我们的方案中, 通过二次认证机制, 防止用户无法正常使用系统, 从而保证了用户体验。因此, 在本方案中 FAR 是最体现系统准确性的评价指标。

尽管 FAR 和 FRR 两者的优化往往存在矛盾, 但在身份认证的场景中, 这两个指标的表现可以更加直观地显示出安全性和便利性的关系。FAR 与 FRR 都与阈值有关, 其中 FAR 随着阈值的增大而减小, 而 FRR 则随着阈值的增大而增大。这是因为当阈值增大时, 能通过认证的用户数量也就越少。在身份认证的场景中, FAR 越小代表系统越安全, 但可能会使得 FRR 增加, 也就是使得合法用户认证失败的概率增加, 会影响用户的体验, 用户会进行二次身份认证(例如短信验证码)。在身份认证系统中, 应该在平衡安全性和便利性的前提下, 根据具体的场景需求进行权衡。对于对安全性要求较高的场景, 应优先满足 FAR 越小越好, 这样系统的安全性更有保障; 而对于用户要求更加便利的、安全性不高的场景, 应优先满足 FRR 越小越好, 以提升用户的使用体验。因此, 在实现身份认证系统时, 应该考虑到特定应用场景的需求, 找到平衡点, 从而满足安全性与便利性需求。

(5) AUC(Area under the ROC Curve)是 ROC 曲线下的面积。ROC 曲线是指接收器操作特性曲线(Receiver Operating Characteristic, ROC), 它描述了二值分类器系统在识别阈值变化时的性能。ROC 曲线通过以 FPR 假阳性率(1-特异度)为横坐标, TPR 真阳性率(灵敏度)为纵坐标, 通过计算不同阈值后得到的所有坐标对 (FPR, TPR) 的连线而得。ROC 曲线具有当测试集中的正负样本的分布变化而 ROC 曲线能够保持不变的特点, 因此利用 ROC 曲线可以解决

在实际的数据集中经常会出现的数据不平衡现象。在身份认证系统中, AUC 的值代表合法用户通过认证的概率大于非法用户通过认证的概率的比例, 因此 AUC 越大, 代表身份认证的性能越好。综合考虑以上指标结合身份认证场景, 我们通过实验结果对其进行全面的分析。

5.3 实验结果

(1) 对于 FAR 的实验结果:

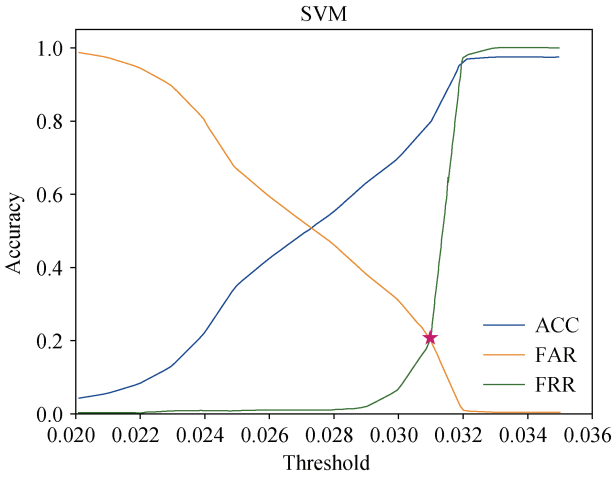
由于 ACC, FAR 和 FRR 都是和阈值(Threshold)有关的值, 因此我们将不同模型的 ACC、FAR 与 FRR 的值随着阈值变化的曲线进行了绘制, 如图 5 所示, 横坐标为阈值, 纵坐标为阈值对应的值。

一方面, 我们看到 FAR 值变小的同时 FRR 值和 ACC 值越大。另一方面, 我们也可以发现, 在身份认证场景中, 安全性与便利性总是折中的。这也体现了在身份认证场景中, 安全性和便利性需要平衡考虑的情况。通过实验结果中的曲线图, 我们可以更加直观地了解这种平衡。FAR 代表非法用户认证通过的概率, FAR 值越小, 安全性越高; 而 FRR 代表合法用户验证失败的概率, FRR 值越小, 便利性越高。因此, 对于某些涉及敏感数据的服务, 我们应该优先考虑安全性, 而对于某些普通服务, 我们可以优先考虑便利性。在动态信任管理的场景中, 我们优先考虑保证 FAR 足够低的情况下 FRR 尽可能低的阈值, 以达到最佳效果。

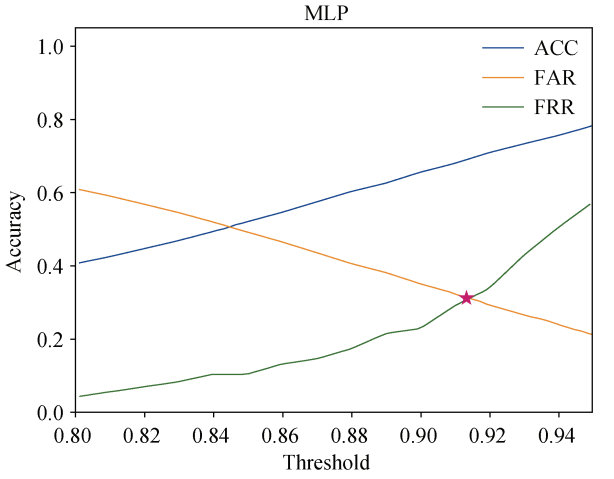
本研究中, 我们选择 FAR 与 FRR 相接近的阈值作为最佳阈值, 具体数值可见表 5, 其中使用单一平台数据(整体数据的 1/3), 没有联邦学习的单一平台训练结果记为 Partial data NOFL, 使用了联邦学习的实验结果记为 Federated Learning:

我们在基于联邦学习的 SVM 模型下选择得到最佳结果的阈值为 0.031, 在基于联邦学习的 MLP 模型最佳结果的阈值选择 0.916, 基于联邦学习的 BP 模型最佳结果的阈值选择 0.566。

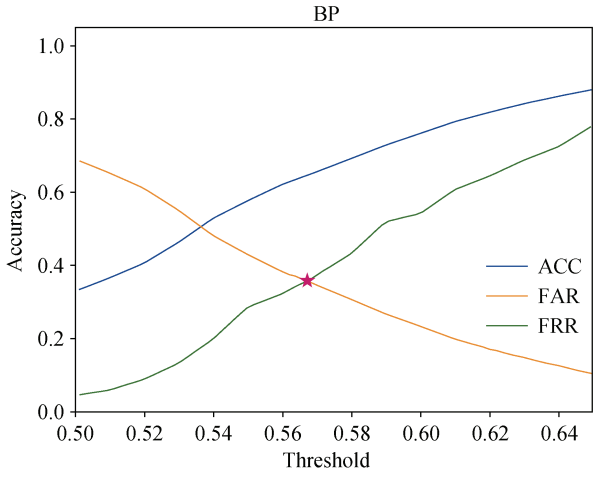
最后, 综合考虑了 FAR 和 FRR 相等的情况下针对基于联邦学习的 SVM、MLP 和 BP 模型的实验结果进行对比, 我们可以看到使用 SVM 作为机器学习模型的单一平台数据 FAR 为 0.203, 而联邦学习身份认证模型 FAR 为 0.192; 而在 MLP 机器学习模型下, 单一平台传统机器学习模型的 FAR 为 0.333, 联邦学习身份认证模型下的 FAR 为 0.309; 在 BP 模型的条件下, 单一平台传统机器学习模型的 FAR 为 0.384, 联邦学习身份认证模型下的 FAR 为 0.357。可以看出在相同的机器模型下, 我们提出的联邦学习身份认证模型都要明显优于在单一平台上传统机器学习模



(a) 基于联邦学习以SVM为机器学习方法的模型
(a) Model based on federated learning using SVM as machine learning method



(b) 基于联邦学习以MLP为机器学习方法的模型
(b) Model based on federated learning using MLP as machine learning method



(c) 基于联邦学习以BP为机器学习方法的模型
(c) Model based on federated learning using BP as machine learning method

图 5 基于联邦学习的模型 ACC、FAR、FRR 实验结果

Figure 5 Experimental results of models ACC, FAR, and FRR based on federated learning

表 5 选定阈值下的实验结果

Table 5 Experimental results with selected thresholds

模型	数据	ACC	FAR	FRR	阈值
SVM	Partial data NOFL	0.779	0.220	0.203	0.060
	Federated Learning	**0.794**	**0.205**	**0.192**	**0.031**
MLP	Partial data NOFL	0.661	0.329	0.333	0.791
	Federated Learning	**0.692**	**0.307**	**0.317**	**0.916**
BP	Partial data NOFL	0.626	0.373	0.384	0.905
	Federated Learning	**0.642**	**0.357**	**0.357**	**0.566**

型。虽然从实验数值上来, 使用联邦学习后 FAR 从 22%降低到了 20.5%, 从数值上来看只降低了 1.5%, 但是京东活跃用户为 5.805 亿人, 意味着可以多识别其中 1.5%即 32 万次非法用户的入侵, 大大降低了账户被劫持的威胁, 对动态评估系统的安全性的提升效果是非常显著的。

为了更明显地看到各个机器学习模型下的实验结果, 我们对每个机器学习方法下联邦学习和单一平台机器学习的训练结果对比, 如图 6 所示。使用 SVM 模型获得了最低的误识率。结合表 5 的数据, 相比于其他两种机器学习方法, SVM 在误识率最低的同时, 其拒识率最低、准确率最高。这些实验结果表明, 在基于联邦学习的动态身份认证模型中, 使用 SVM 作为机器学习模型的联邦学习训练结果性能最好, 因此我们建议在实践中使用 SVM 模型来提高系统的准确性和效率。

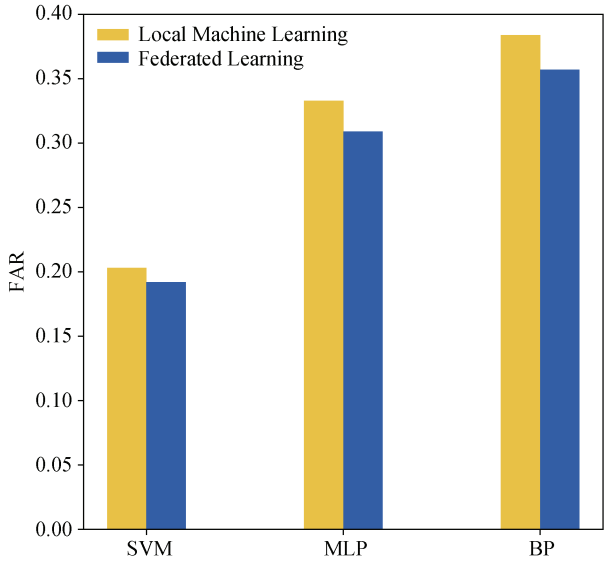


图 6 基于联邦学习和本地机器学习的 FAR 实验结果

Figure 6 Experimental results of FAR based on federated learning and local machine learning

(2) ROC 与 AUC 的实验结果:

在本研究中, 我们对不同机器学习模型下使用联邦学习和单一平台使用传统机器学习的 ROC 结果进行了对比, 如图 7 所示。联邦学习中使用 SVM 模型的 AUC 值为 0.85, MLP 模型的 AUC 值为 0.74, BP 模型的 AUC 值为 0.71; 而单一平台下的 SVM 模型的 AUC 值为 0.84, MLP 模型的 AUC 值为 0.73, BP 模型的 AUC 值为 0.69。根据实验结果, 我们发现在单一平台下, SVM 机器学习模型的性能相对较好; 而在联邦学习中, 也是 SVM 的性能最好, MLP 和 BP 次之。此外, 总体测试效果表明, 联邦学习的性能优于单一平台的性能。因此, 我们可以认为在离散型用户行为特征的身份认证场景下, 使用联邦学习方法可以增强身份认证系统的性能。

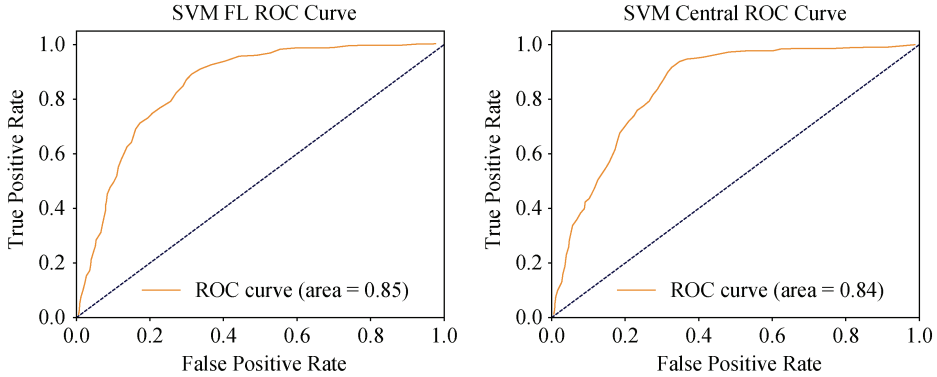
(3) 不同模型的训练时间的实验结果

我们分别对实验的训练时间进行了记录, 使用不同的模型在相同的实验环境下, 25589 个用户的数据集下在每个模型中分别训练 10 次, 记录每次的训练时间, 并计算平均训练时间。其中基于联邦学习的 MLP 的平均训练时间为 26.254 秒, BP 的平均训练时间为 25.214 秒, SVM 的平均训练时间为 410.151 秒, 单一平台的 MLP 的平均训练时间为 2.620 秒, BP 的平均训练时间为 2.930 秒, SVM 的平均训练时间为 55.503 秒。从训练时间的角度, 我们可以发现联邦学习在可接受的范围内牺牲一定的训练时间性能表现的情况下, 降低了身份认证的误识率。

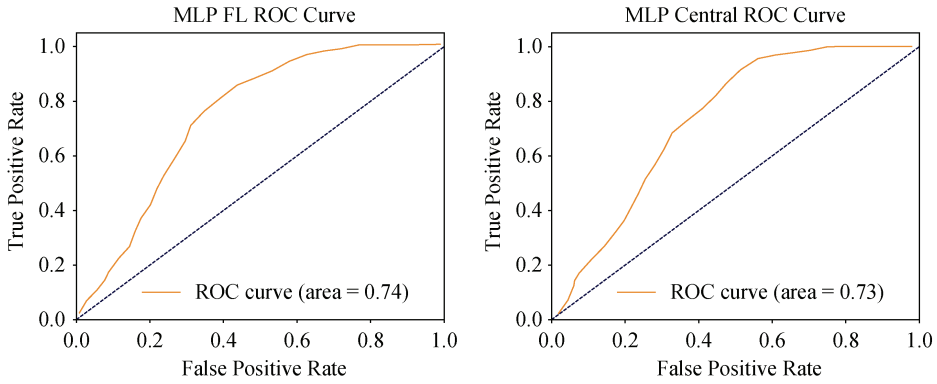
(4) 异常样本均衡对实验结果的影响

我们还探究了异常样本均衡度对联邦学习训练结果的影响, 我们将异常数据的 81.68%分配给第一个参与方, 9.21%分配给第二个参与方, 9.21%分配给第三个参与方, 三个参与方获得不同均衡程度的数据的方式进行训练。这样, 每个参与方拥有的异常数据所占比例从原来的均为 4.5%变为不同的均衡程度: 现在第一个参与方拥有的数据中异常数据占 10.2%, 第二个、第三个参与方拥有的数据中异常数据占 1.2%。我们通过基于不同均衡程度的数据训练三个参与方的方式探究异常样本均衡度对联邦学习训练结果的影响。

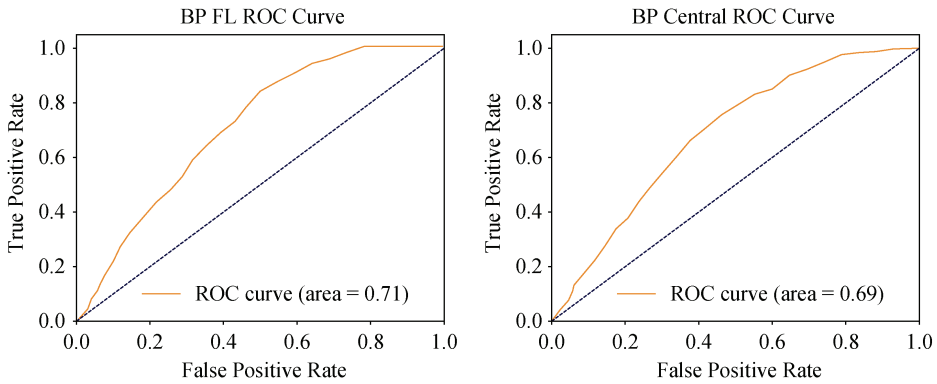
我们分别采用这些不同均衡程度的数据对联邦学习模型进行训练, 并比较其性能表现。实验结果如图 8 所示, 基于误识率为衡量标准, 我们采用 SVM、MLP 和 BP 等机器学习模型对数据集进行训练。我们发现, 当采用异常样本不均衡的数据集, 基于联邦学习的动态身份认证模型略优于单一平台的机器学习模型训练结果, 但其性能提升效果还是明显差于采用异常样本均衡的数据集的联邦学习模型性能结果, 可以知道使用异常样本均衡数据的联邦学习



(a) 以SVM为机器学习方法的模型
(a) Model using SVM as machine learning method



(b) 以MLP为机器学习方法的模型
(b) Model using MLP as a machine learning method



(c) 以BP为机器学习方法的模型
(c) Model using BP as a machine learning method

图 7　基于联邦学习和单一平台的 ROC 与 AUC 的实验结果

Figure 7　Experimental results of ROC and AUC based on federated learning and single platform

效果更好。这是因为异常样本在所有参与方之间均衡分布，使得每个参与方都可以得到足够数量的异常样本进行训练。在这种情况下，联邦学习模型可以更好地学习异常样本的特征，并在动态身份认证场景中表现出更好的性能。值得注意的是，MLP 机器学习模型下，采用不均衡异常样本数据联邦学习的误识率不仅高于采用异常样本均衡数据联邦学习模型，还高于单一平台的机器学习模型的训练结果，这表明了使用 MLP 机器学习模型对异常样本的均衡十分敏感。这进一步强调了使用异常样本均衡数据进行联邦学习训练的重要性，尤其对于那些对异常样本分布较为敏感的模型。

6　讨论

联邦学习虽然在动态身份认证领域可以解决用户异常登陆数据稀缺和用户隐私泄露的问题，本文假设参与方式诚实可信的，但参与方如果是恶意的，那么联邦学习自身的安全性和隐私性问题也会对动态信任评估系统也有非常大的影响。

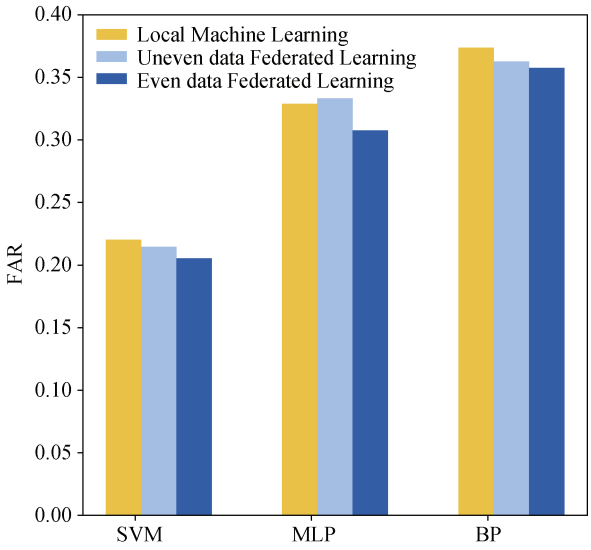


图 8　异常样本均衡度不同的联邦学习模型和单一平台机器学习模型的误识率实验结果

Figure 8　The FAR of federated learning models and single platform machine learning models with different equilibrium degrees of abnormal sample

6.1　动态信任评估系统的安全性及联邦学习自身的安全性对动态信任评估系统的影响

联邦学习主要面临的安全威胁会破坏联邦学习中的完整性和可用性, 主要包括投毒攻击(Poisoning attack)、对抗样本攻击、搭便车攻击(Free-riding attacks)。

投毒攻击就是参与训练的参与方中如果有一方是恶意的, 就可以利用投毒攻击来建立后门, 从而达到误导训练模型在恶意参与方的控制下按照恶意参与方预先设定的目标进行分类训练。本系统中, 参与方 A 参与方 B 和参与方 C 联合训练一个模型, 参与方 A 是恶意参与方, 通过训练中毒数据来生成有毒的本地模型, 从而降低动态信任评估系统性能。

对抗样本攻击会干扰联邦学习训练或推理过程, 影响联邦学习训练时的收敛速度和推理结果。对抗样本攻击可以根据攻击者拥有的信息分为白盒攻击和黑盒攻击。白盒攻击中恶意的参与方能够获取机器学习算法以及模型参数, 并根据这些已知信息去制作对抗样本, 影响模型训练。在我们的动态信任评估系统中参与方中的一方可以拿到样本的模型, 在输入的样本中加入轻微的扰动, 进而导致动态信任评估系统对用户输出错误的信任评估结果, 导致误识率变高。

在搭便车攻击中, 在联邦学习的数据收集阶段和训练阶段, 如果有一方利用了各个参与方共同训练的全局模型却不贡献或者只少量贡献自己的数据和计算资源。这样的攻击行为会使得全局模型的训练付出额外的计算资源。在本方案的场景中, 参与方 A 参与动态信任评估系统中的全局模型的训练但只是向中央服务器发送随机参数, 而不是发送利用本平台用户的真实数据训练得到的本地模型参数。参与方 A 假装为联邦学习全局模型的训练更新提供了贡献, 这样会降低全局模型的性能从而增加动态信任评估方案核心模型的训练轮数, 进而消耗训练资源, 降低系统总体的可用性, 也会影响其他参与方的安全性。

6.2　动态信任评估系统的隐私性及联邦学习自身的隐私性对动态信任评估系统的影响

本系统在模型训练阶段使用的联邦学习框架应用了同态加密来加密参数, 并将其上传至中央服务器解密, 保护了模型训练中用户的隐私性。另外, 相比于 Yang 等人所提出的经典身份认证方案使用的连续型特征数据, 例如通话记录、蓝牙连接日志、Wifi 接入点等数据[30], 从中可以挖掘到用户的使用习惯、兴趣爱好和行为特征等隐私信息, 本研究使用的离散型数据, 包括登录频率、交易频率等模糊信息, 相比之下用户隐私受到的威胁更弱。

联邦学习虽然能够帮助其参与方在保护用户数据隐私的同时联合训练模型, 但其面对的隐私威胁也会破坏联邦学习的机密性, 以下是对该段话进行纠正的建议: 虽然联邦学习能够在保护用户数据隐私的同时联合训练模型, 但它面临着隐私威胁, 这可能会破坏其机密性, 进而影响动态信任评估系统的隐私性。本文使用的横向联邦学习主要面临的隐私威胁是推理攻击。由于联邦学习下每个参与方和中央服务器随着训练迭代更新相同的全局模型, 参与方可以推断出模型的信息。相比于投毒攻击, 这种方法是被动获取模型泄露信息。在动态身份认证场景中, 参与方 A 进行推理攻击可能会在模型训练过程中或者模型训练完成后, 推断出其他平台的用户训练数据集的属性信息, 包括与数据集无关的属性, 例如用户的年龄分布等信息, 极大地威胁本系统中用户的隐私性。

7　结束语

本文提出了基于联邦学习的动态信任评估身份认证方法, 在单一平台异常数据比较少的情况下, 通过充分利用离散型用户行为数据, 可以一定程度上降低了身份认证的误识率, 提高了动态信任评估系统的可靠性, 同时保证用户数据的隐私安全。与已有的方案相比本文使用联邦学习框架中的同态加密

保护了用户数据的隐私, 使用离散型数据使得用户与传统的机器学习动态信任评估方法相比, 本文所提出的方法进一步扩大了数据集, 同时能够更好地保护用户数据的隐私性和安全性, 提升了训练和学习效果。

通过对真实的 26044 个用户的登录与交易数据处理得到离散型用户行为数据, 本文进行了三方平台联邦学习和单一平台机器学习训练实验。实验结果表明, 本文所提出的方法可以在可接受的范围内牺牲一定的训练时间性能表现, 降低身份认证的误识率, 并且相较于单一平台, 基于联邦学习的动态信任评估方案扩展了数据集并表现了更好的性能。

另外, 我们用三种不同的机器学习模型对提出的方案进行了对比, 发现 SVM 机器学习模型可以完成更加高效准确的分类任务。综上, 我们所提出的基于联邦学习的动态信任评估方法, 可以有效地解决单一平台异常数据少的问题, 同时保证用户行为数据的隐私性, 整个认证过程高效可靠, 但也需要注意异常样本的均衡性。

未来研究中, 我们会考虑进一步挖掘以下几个方向:

(1) 本文仅考虑了用户登录行为数据和交易行为数据, 但在实际业务中, 不同类型的异常数据可能会对方案的训练效果造成影响。因此, 未来的研究可以考虑在数据预处理和特征提取方面进行优化, 以提高方案对于各种异常数据的适应能力。

(2) 本研究主要关注交易平台身份认证领域, 但身份认证方案可以应用于更多的业务场景, 未来可以考虑在这些领域中探索方案的应用, 为实际应用场景提供更有效的解决方案。

参考文献

[1] Ward R, Betsy B. BeyondCorp: A New Approach to Enterprise Security[J]. *login Usenix Mag*, 2014: 39.

[2] Osborn B, McWilliams J, Beyer B, et al. Beyondcorp: Design to deployment at google[J]. *login Usenix Mag*, 2016: 26-34.

[3] Sugrim S, Liu C, McLean M, et al. Robust Performance Metrics for Authentication Systems[C]. *Proceedings 2019 Network and Distributed System Security Symposium*, 2019.

[4] Trojahn M, Ortmeier F. Toward Mobile Authentication with Keystroke Dynamics on Mobile Phones and Tablets[C]. *2013 27th International Conference on Advanced Information Networking and Applications Workshops*, 2013: 697-702.

[5] Frank M, Biedert R, Ma E, et al. Touchalytics: On the Applicability of Touchscreen Input as a Behavioral Biometric for Continuous Authentication[J]. *IEEE Transactions on Information Forensics and Security*, 2013, 8(1): 136-148.

[6] Li L, Zhao X, Xue G. Unobservable Re-authentication for Smartphones[C]. *Network and Distributed Systems Security*, 2013, 56: 57-59.

[7] Feng T, Liu Z Y, Kwon K A, et al. Continuous Mobile Authentication Using Touchscreen Gestures[C]. *2012 IEEE Conference on Technologies for Homeland Security*, 2012: 451-456.

[8] Zheng N, Bai K, Huang H, et al. You Are how You Touch: User Verification on Smartphones *via* Tapping Behaviors[C]. *2014 IEEE 22nd International Conference on Network Protocols*, 2014: 221-232.

[9] Serwadda A, Phoha V V, Serwadda A, et al. When Kids' Toys Breach Mobile Phone Security[C]. *The 2013 ACM SIGSAC Conference on Computer & Communications Security*, 2013: 599-610.

[10] Xu Z, Bai K, Zhu S C. TapLogger: Inferring User Inputs on Smartphone Touchscreens Using On-Board Motion Sensors[C]. *The Fifth ACM Conference on Security and Privacy in Wireless and Mobile Networks*, 2012: 113-124.

[11] Ehatisham-ul-Haq M, Awais Azam M, Naeem U, et al. Continuous Authentication of Smartphone Users Based on Activity Pattern Recognition Using Passive Mobile Sensing[J]. *Journal of Network and Computer Applications*, 2018, 109: 24-35.

[12] Zhu T T, Qu Z Y, Xu H T, et al. RiskCog: Unobtrusive Real-Time User Authentication on Mobile Devices in the Wild[J]. *IEEE Transactions on Mobile Computing*, 2020, 19(2): 466-483.

[13] Amini S, Noroozi V, Pande A, et al. DeepAuth: A Framework for Continuous User re-Authentication in Mobile Apps[C]. *The 27th ACM International Conference on Information and Knowledge Management*, 2018: 2027-2035.

[14] Lee W H, Lee R B. Multi-sensor Authentication to Improve Smartphone Security[C].*2015 International conference on information systems security and privacy*, 2015: 1-11.

[15] Yang L, Guo Y, Ding X, et al. Unlocking Smart Phone through Handwaving Biometrics[J]. *IEEE Transactions on Mobile Computing*, 2015, 14(5): 1044-1055.

[16] Buthpitiya S, Zhang Y, Dey A K, et al. N-Gram Geo-Trace Modeling[M]. Pervasive Computing. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011: 97-114.

[17] Riva O, Qin C, Strauss K, et al. Progressive Authentication: Deciding when to Authenticate on Mobile Phones[C]. *USENIX Security Symposium*, 2012.

[18] Agadakos I, Hallgren P, Damopoulos D, et al. Location-Enhanced Authentication Using the IoT: Because You Cannot Be in Two Places at Once[C]. *The 32nd Annual Conference on Computer Security Applications*, 2016.

[19] Mare S, Markham A M, Cornelius C, et al. ZEBRA: Zero-Effort Bilateral Recurring Authentication[C]. *2014 IEEE Symposium on Security and Privacy*, 2014: 705-720.

[20] Vhaduri S, Poellabauer C. Multi-Modal Biometric-Based Implicit Authentication of Wearable Device Users[J]. *IEEE Transactions on Information Forensics and Security*, 2019, 14(12): 3116-3125.

[21] Zhang Y T, Hu W, Xu W T, et al. Continuous Authentication Using Eye Movement Response of Implicit Visual Stimuli[J]. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous*

Technologies, 2018, 1(4): 1-22.

[22] Lee W H, Lee R B. Implicit Smartphone User Authentication with Sensors and Contextual Machine Learning[C]. *2017 47th Annual IEEE/IFIP International Conference on Dependable Systems and Networks*, 2017: 297-308.

[23] Li H Y, Yu J N, Cao Q. Intelligent Walk Authentication: Implicit Authentication when You Walk with Smartphone[C]. *2018 IEEE International Conference on Bioinformatics and Biomedicine*, 2018: 1113-1116.

[24] Sun J C, Zhang R, Zhang J X, et al. TouchIn: Sightless Two-Factor Authentication on Multi-Touch Mobile Devices[C]. *2014 IEEE Conference on Communications and Network Security*, 2014: 436-444.

[25] Yang Y Y, Sun J Y, Guo L K. PersonaIA: A Lightweight Implicit Authentication System Based on Customized User Behavior Selection[J]. *IEEE Transactions on Dependable and Secure Computing*, 2019, 16(1): 113-126.

[26] Lee W H, Liu X C, Shen Y L, et al. Secure Pick Up: Implicit Authentication when You Start Using the Smartphone[C]. *The 22nd ACM on Symposium on Access Control Models and Technologies*, 2017: 67-78.

[27] Wang P, Lin W H, Chao K M, et al. A Face-Recognition Approach Using Deep Reinforcement Learning Approach for User Authentication[C]. *2017 IEEE 14th International Conference on e-Business Engineering*, 2017: 183-188.

[28] Abuhamad M, Abuhmed T, Mohaisen D, et al. AUToSen: Deep-Learning-Based Implicit Continuous Authentication Using Smartphone Sensors[J]. *IEEE Internet of Things Journal*, 2020, 7(6): 5008-5020.

[29] Gomi H, Yamaguchi S, Tsubouchi K, et al. Towards Authentication Using Multi-Modal Online Activities[C]. *The 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers*, 2017: 37-40.

[30] Yang Y Y, Sun J Y. Energy-Efficient W-Layer for Behavior-Based Implicit Authentication on Mobile Devices[C]. *IEEE INFOCOM 2017 - IEEE Conference on Computer Communications*, 2017: 1-9.

[31] Freeman D, Jain S, Duermuth M, et al. Who Are You? A Statistical Approach to Measuring User Authenticity[C]. *Proceedings 2016 Network and Distributed System Security Symposium*, 2016: 21-24.

[32] JD data set[DB]. https://github.com/pjgao/jddir, Nov. 2017.

[33] Cortes C, Vapnik V. Support-Vector Networks[J]. *Machine Learning*, 1995, 20(3): 273-297.

[34] Ke G L, Meng Q, Finley T, et al. LightGBM[C]. *The 31st International Conference on Neural Information Processing Systems*, 2017: 3149-3157.

[35] Khan H, Atwater A, Hengartner U. Itus: An Implicit Authentication Framework for Android[C]. *The 20th Annual International Conference on Mobile Computing and Networking*, 2014: 507-518.

[36] Ryffel T, Trask A, Dahl M, et al. A Generic Framework for Privacy Preserving Deep Learning[EB/OL]. 2018: 1811.04017. https://arxiv.org/abs/1811.04017v2.

[37] FederatedAI/ FATE[EB/OL]. https://github.com/FederatedAI/FATE. Sep. 2022.

[38] FedML-AI/ FedML[EB/OL]. https://github.com/FedML-AI/FedML. Oct. 2022.

[39] tensorflow/ federated[EB/OL]. https://github.com/tensorflow/federated. Oct. 2022.

[40] OpenMined/ Pysyft[EB/OL]. https://github.com/OpenMined/PySyft. Sep. 2020.

[41] He C Y, Li S Z, So J, et al. FedML: A Research Library and Benchmark for Federated Machine Learning[EB/OL]. 2020: 2007.13518. https://arxiv.org/abs/2007.13518v4.

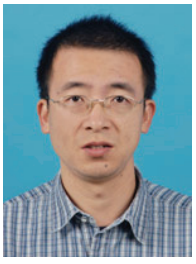
[42] Yang Q, Liu Y, Chen T J, et al. Federated Machine Learning: Concept and Applications[J]. *ACM Transactions on Intelligent Systems and Technology*, 2019, 10(2): 1-19.

[43] Geng J H, Kanwal N, Jaatun M G, et al. DID-eFed: Facilitating Federated Learning as a Service with Decentralized Identities[C]. *Evaluation and Assessment in Software Engineering*, 2021: 329-335.

[44] McMahan H B, Moore E, Ramage D, et al. Communication-Efficient Learning of Deep Networks from Decentralized Data [EB/OL]. 2016: 1602.05629. https://arxiv.org/abs/1602.05629v4.

[45] 京东 2022 年第一季度财报[EB/OL]. https://ir.jd.com/system/files-encrypted/nasdaq_kms/assets/2022/05/17/18-38-12/JD.com%20Announces%20First%20Quarter%202022%20Results.pdf. Mar. 2022

[46] Xiao X, Tang Z, Xiao B, et al. A Survey on Privacy and Security Issues in Federated Learning[J]. *Chinese Journal of Computers*, 2023, 46(5): 1019-1044.
(肖雄, 唐卓, 肖斌, 等. 联邦学习的隐私保护与安全防御研究综述[J]. *计算机学报*, 2023, 46(5): 1019-1044.)



石瑞生 博士, 北京邮电大学副教授, 硕士生导师, 主要研究方向为网络空间安全、信息安全、物联网、服务计算等。Email: shiruisheng@bupt.edu.cn

付彤 于 2021 年在北京工业大学软件工程专业获得学士学位。现在北京邮电大学计算机技术专业攻读硕士学位。研究领域为网络安全。研究兴趣包括: 网络空间安全、信息安全。Email: futong2021140994@bupt.edu.cn

林子丁 于 2023 年在北京邮电大学电信工程及管理专业获得学士学位。现在北京邮电大学网络与信息安全专业攻读硕士学位。研究领域为网络安全。研究兴趣包括: 网络空间安全、区块链、隐私保护。Email: linziding@bupt.edu.cn

姜宁 于 2021 年在北京邮电大学获得硕士学位, 研究领域为网络安全、动态信任管理、单点登录。Email: jiangning@bupt.edu.cn

兰丽娜 博士, 北京邮电大学副教授, 硕士生导师, 主要研究方向为物联网、服务计算、大数据分析。Email: lanlina@bupt.edu.cn