doi:10.3969/j.issn.1001-4616.2024.03.015

基于联邦迁移的跨项目软件缺陷预测

宋慧玲[1,2],李　勇[1,2,3],张文静[1,2]

(1.新疆师范大学计算机科学技术学院,新疆 乌鲁木齐 830054)
(2.新疆电子研究所软件事业部,新疆 乌鲁木齐 830010)
(3.南京航空航天大学高安全系统的软件开发与验证技术工信部重点实验室,江苏 南京 211106)

[**摘要**]　跨项目软件缺陷预测基于已标注的多源项目数据构建模型,可以解决软件历史数据不足和标注代价高的问题. 但在传统跨项目缺陷预测中,源项目数据持有者为了保护软件数据的商业隐私,而导致的“数据孤岛”问题直接影响了跨项目预测的模型性能. 本文提出基于联邦迁移的跨项目软件缺陷预测方法(FT-CPDP). 首先,针对数据隐私泄露和项目间特征异构问题,提出基于联邦学习与迁移学习相结合的模型算法,打破各数据持有者间的“数据壁垒”,实现隐私保护场景下的跨项目缺陷预测模型. 其次,在联邦通信过程中添加满足隐私预算的噪声来提高隐私保护水平,最后构建卷积神经网络模型实现软件缺陷预测. 基于 NASA 软件缺陷预测数据集进行实验,结果表明与传统跨项目缺陷预测方法相比,本文提出的 FT-CPDP 方法在实现软件数据隐私保护的前提下,模型的综合性能表现较优.

[**关键词**]　软件缺陷预测,联邦学习,迁移学习,差分隐私,卷积神经网络
[**中图分类号**]TP181;TP311.5　[**文献标志码**]A　[**文章编号**]1001-4616(2024)03-0122-07

Cross-project Software Defect Prediction Based on Federated Transfer

Song Huiling[1,2],Li Yong[1,2,3],Zhang Wenjing[1,2]

(1.College of Computer Science and Technology,Xinjiang Normal University,Urumqi 830054,China)
(2.Software Development Department,Xinjiang Electronic Research Institute,Urumqi 830010,China)
(3.Key Laboratory of Ministry of Industry and Information Technology for Safety-critical Software Development and Verification,
Nanjing University of Aeronautics and Astronautics,Nanjing 211106,China)

Abstract:Cross-project software defect prediction is based on labeled multi-source project data to build a model,which can address the problem of insufficient software historical data and high labeling cost. However,in traditional cross-project defect prediction,the problem of “data-island” caused by source project data holders to protect the business privacy of software data directly affects the model performance of cross-project prediction. Therefore,in this paper,we propose a cross-project software defect prediction method based on federated transfer(FT-CPDP). Firstly,to address the problem of data privacy leaking and feature heterogeneity between projects,this paper presents a model algorithm based on the combination of federal learning and migratory learning to break down the “data barrier” among data holders,and to achieve cross-project defect prediction model in the privacy protection scenario. Secondly,in the federal communication process,the level of privacy protection is increased by adding noise that satisfies the privacy budget. Finally,a convolution neural network model is built to realize software defect prediction. Experiments based on NASA software defect prediction dataset show that compared with traditional cross-project defect prediction methods,FT-CPDP method achieves better comprehensive performance on the premise of software data privacy protection.

Key words:software defect prediction,federated learning,transfer learning,differential privacy,convolutional neural network

随着计算机软件应用范围的不断扩大,软件需求的多样化导致各行业对软件质量标准的要求也越来越严格,提前发现软件中存在的缺陷,可避免造成不必要的财产损失. 因此,软件缺陷预测是提高软件测试效率,保证软件可靠性的重要途径[1].

传统的软件缺陷预测由一个数据分析者进行集中式的模型训练,但这种方法容易造成数据隐私泄

收稿日期:2022-09-20.
基金项目:新疆维吾尔自治区天山青年计划项目(2020Q019)、新疆师范大学博士科研启动基金项目(XJNUBS1905).
通讯作者:李勇,博士,副教授,研究方向:机器学习与实证软件工程. E-mail:liyong@ live.com

露. 因此,各开发数据持有者开始重视数据隐私而不愿意共享敏感数据,数据的碎片化导致模型性能下降. 因此,数据隐私问题成为软件缺陷预测领域中研究的热点. 为解决这一问题,联邦学习顺应而起. 联邦学习(Federated Learning)是特殊的分布式深度学习方法,与传统的训练模式不同,联邦学习实现了"数据不动,模型动"的训练方式,是一种可信联邦学习,对隐私保护起到了重要作用.

根据上述问题,对联邦学习和隐私计算方法进行分析和研究,提出基于联邦迁移的跨项目软件缺陷预测方法(cross project software defect prediction based on Federated transfer,FT-CPDP). 将联邦学习与数据保护机制结合对软件缺陷进行预测,可达到在不共享数据的前提下,各数据持有方获得理想的模型效果. 具体来讲,各项目都可作为参与方协同训练模型,为解决各项目间数据异构的特点,通过特征选择方法使得各项目存在相同度量元. 在可信联邦体系中,存在客户端和服务器端两个角色. 客户端的各项目"闭关训练",互不干扰. 本文采用真实公开数据集,使用卷积神经网络(convolutional neural networks,CNN)模型进行软件缺陷预测,并与传统的跨项目软件缺陷学习方法比较. 结果显示,基于联邦迁移的跨项目软件缺陷预测方法不仅保护了数据隐私,而且具有很好的效果,在软件缺陷预测的研究上有很好的实际意义.

1 相关工作

1.1 软件缺陷预测研究

传统的软件缺陷预测方法可分为项目内软件缺陷预测和跨项目软件缺陷预测,大部分的软件缺陷预测基于 SVM、决策树等机器学习方法构建模型. 项目内软件缺陷预测方法要求目标数据必须有历史积累数据[2],刘文英等[3]对软件缺陷数据中类不平稳问题提出了新的 RUS-RSMOTE-PCA-Vote 分类方法,降低软件缺陷预测中数据不平衡影响,提高分类器性能. 在软件版本演变的进程中,数据集中容易出现标记不同而特征相同的情况,针对这种跨版本的类重叠问题,曲豫宾等[4]提出采用混合式最近邻清理策略方法,以深度学习为基模型,将语义向量作为基本输入,自动学习语义特征并改进数据抽样策略,提高了模型性能. 针对训练数据的选择,盖金晶等[5]提出基于 JS 散度和相对密度的跨项目软件缺陷预测方法.

但在实际应用中,新的项目往往历史数据不足或者没有,为解决这一问题,有研究者提出跨项目软件缺陷预测方法(cross-project defect prediction,CPDP). 倪超等[6]提出将迁移学习用于软件缺陷预测,实现特征和实例迁移的跨项目软件缺陷预测,解决项目间的特征异构和数据不足问题. 李勇等[7]使用多源迁移学习算法,通过将多源项目的数据信息转移到目标项目,实现跨项目软件数据的缺陷预测.

上述文献根据软件缺陷数据中存在的历史数据以及特征等问题分别进行研究. 但都没有考虑到敏感数据的隐私问题,敏感数据的泄露容易造成用户财产损失以及安全问题. 联邦学习与隐私计算的提出为软件缺陷预测研究提供了新的思路.

1.2 联邦学习研究

针对软件缺陷预测中的数据隐私问题,Chen 等[8]首次提出将差分隐私用于软件缺陷预测,通过连续特征离散化来分配优化隐私预算. 但这一方法仍然需要数据持有者将数据共享. 联邦学习的提出解决了这一问题.

在医疗健康方面,Chen 等[9-10]在保护用户数据的前提下,对数据建模实现了个性化联邦. 针对领域转移和数据隐私问题,Zhang 等[11]提出不同用户使用不同的模型实现个性化模型,并在模型通信阶段采用深度对抗网络解决联邦迁移问题,但该研究的局限是假设所有客户机的机器健康状态集都相同,无法解决来自不同测试平台的源数据和目标数据. 面对数据异构问题,Wang 等[12]提出通过迁移学习和微调相结合的方法构建每个参与者的初始私有模型,知识蒸馏获得度量信息. 但上述文献均使用同态加密方法对数据进行保护,增加了通信代价. Sharma 等[13]在保护机制上与之前基于同态加密(homomorphic encryption,HE)的方法相比,采用多方计算(multi-party computation,MPC)实现联邦模型,在半诚实安全设置下提高了一个数量级的效率,但缺乏主动安全的机器学习协议. 在脑电波信号分类问题中,Ju 等[14]提出利用单次试验协方差矩阵,利用领域自适应技术从多学科脑电图数据中提取常见的鉴别信息. 但在联邦通信过程中可能会遭受恶意攻击造成敏感数据的泄露. 孔秀平等[15]通过在隐私保护的前提下,对车辆轨迹数据进行联邦建模,但缺乏对异构数据的考虑.

基于以上分析,本文提出了基于联邦迁移的跨项目软件缺陷预测算法,一方面引入联邦学习解决数据

持有者不愿共享敏感数据问题,并结合迁移学习解决实际情况中项目间存在的特征异构[16],以及历史数据不足问题. 另一方面,在联邦通信过程中采用差分隐私安全机制,既保护模型参数,又减小了通信代价[17]. 在隐私保护的场景下,实现对软件缺陷预测模型性能的提高.

2　基于联邦迁移的跨项目缺陷预测

本章节首先介绍基于联邦迁移的跨项目软件缺陷预测(FT-CPDP)解决的问题,接下来分别从本地模型的更新和中央服务器聚合模型两个方面对上述 FT-CPDP 算法进行详细介绍.

2.1　FT-CPDP 算法

软件缺陷预测由于存在敏感属性,为防止商业隐私的泄露,各软件开发商形成数据孤岛局面. 而联邦学习是满足各参与方需求的一种新型的模型训练方式,它打破了传统的集中式学习思维,保护了数据隐私. 本文将联邦学习引入软件缺陷预测,每个软件开发商作为一个客户端,选择一个半诚实的第三方服务器聚合模型参数. 由于各客户端存在特征异构,采用特征选择方法将特征对齐,并将全局参数传递给参与方(步骤 1-3). 本地训练时(步骤 4),参与方根据损失函数计算梯度并更新模型参数权重,公式 1 为二分类的交叉熵损失函数(Cross Entropy Loss). 最后,服务器端聚合扰动后的模型参数(步骤 5-6). FT-CPDP 伪代码如算法 1 所示.

$$L=-[y\log\hat{y}+(1-y)\log(1-\hat{y})], \tag{1}$$

式中,L 为交叉熵损失函数,y 代表真实值,$\hat{y}$ 表示预测值.

算法 1　联邦迁移的跨项目软件缺陷预测算法(FT-CPDP)

输入:全局模型 CNN,公共数据集 public data,本地数据集 $D_1\cdots D_k$

输出:损失函数 loss,模型参数 $\bar{w}$

1.各 client 端数据集与公共数据集进行特征对齐

2.server 端随机选择 n 个参与方并训练 public data 得到全局参数 w^0

3. server 端将 w^0 下发到选中的 client

4.每个 client 使用 D_i 执行随机梯度下降(SDP),得到局部参数 w_i^k 和损失函数 loss,并添加噪声发送给 server

5.server 采用联邦平均算法计算聚合的模型参数 $\bar{w}$,计算损失函数

6.判断是否大于全局轮数,若没有,则重复 3-5

7.结束

为避免模型参数在向服务器的传递过程中遭到恶意攻击,导致参与方敏感数据遭到泄露,本文在 client 端传递模型参数前引入保护机制,增强隐私保护程度. 在联邦学习中常用的保护机制有同态加密、秘密共享、差分隐私等. 差分隐私(differential privacy,DP)是联邦学习的一种保护方法,通常在联邦通信阶段给模型参数添加噪声达到增强 FL 隐私保护程度的目的. 与差分隐私相比,同态加密和秘密共享机制的通信代价高且隐私保护程度低[18-19]. 因此,本文结合高斯机制提高数据保护水平,实现模型参数的个性化保护. 高斯噪声满足如下公式:

$$M(d)\triangleq f(d)+N(0,s_f^2\cdot\sigma), \tag{2}$$

式中,$M(d)$ 表示噪声相加后的查询结果,$N(0,s_f^2\cdot\sigma)$ 表示高斯机制产生的噪声,其均值为 0,标准差为 $s_f^2\cdot\sigma$,s_f 为灵敏度调整算子. 联邦迁移的跨项目软件缺陷预测框架如图 1 所示.

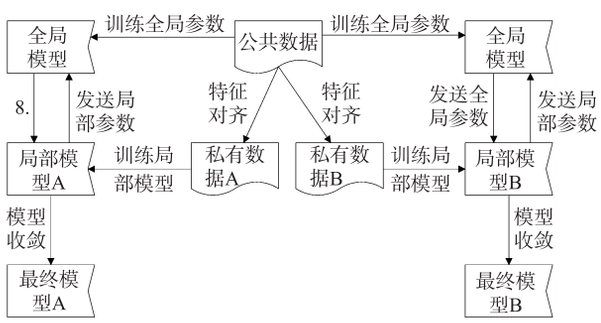


图 1　基于联邦迁移的跨项目软件缺陷预测框架

Fig. 1　Cross project software defect prediction based on Federated transfer

2.2　FT-CPDP 本地化参数更新算法

客户端在接收到服务器端发送的全局参数 w^0 后,使用本地数据集对局部模型 CNN 进行训练,在训练

过程中采用随机梯度下降(SGD)更新局部模型参数 w_i^k,并将扰动后的模型参数 $\tilde{w}_i^k$ 传递给服务器[20-21]. 在这个过程中,各客户端数据不出本地,也不会知道别的客户端数据信息. FT-CPDP 的本地化参数更新伪代码如算法 2 所示.

算法 2　FT-CPDP 的本地化参数更新算法

输入:客户端 client:$(D_1 \cdots .. D_k)$,批量大小:B,初始全局模型参数 w^0,隐私预算 ε
输出:局部模型参数 w_t^k,损失函数 loss
1.for client i in range(K) do
B← Random(D_i)
loss = softmax(y,y')
2. 根据损失函数计算梯度并更新
3. for each $b \in B$ do //每一个小批量数据进行随机梯度下降更新参数
$w_{b+1,i} \leftarrow w_{b,i} - \eta \nabla L(w_{b,i};b)$
end for
4.对模型参数进行扰动
5.$\tilde{w}_{b+1,i} \leftarrow w_{b+1,i} + M(\varepsilon,\delta)$
$w_t^k \leftarrow \tilde{w}_{b+1,i}$
6. 每个 client 将扰动后的模型参数传递给服务器端
7.结束

2.3　FT-CPDP 服务器聚合算法

在每一轮迭代中,服务器接收到各客户端传递过来的局部模型参数 w_i^k,对模型参数进行加权平均得到聚合后的全局参数[22]. 并对损失函数进行计算,经过 N 轮迭代后停止迭训练[23-24]. *FT-CPDP* 的服务器聚合更新伪代码如算法 3 所示.

算法 3　FT-CPDP 的服务器聚合算法

输入:全局训练轮数 N,公共数据集 pubilc data,局部模型参数 w_i^k
输出:聚合后的模型参数 $\bar{w}_{t+1}$
1.计算全局初始参数 w^0
2. for epoch t in range(N) do //每一轮都进行参数聚合

$$\bar{w}_{t+1} \leftarrow \sum_{k=1}^{k} \frac{n_k}{n} w_t^k$$

$N' = N+1$
If N'大于 N //判断是否大于全局轮数
Else 服务器将作为全局参数下发到每一个 client 端并执行算法 2

3　结束

3.1　实验

实验采用美国航天航空局 NASA 公开数据集,NASA 数据是软件缺陷相关数据集,其数据描述如表 1 所示[25].

表 1　NASA 数据集表

Table 1　NASA data set table

项目	语言	样例	属性	预测率	项目	语言	样例	属性	预测率
CM1	C	505	38	48	MW1	C	375	38	7.5
JM1	C	7720	22	20.9	PC1	C	919	38	6.5
KC1	C++	1162	22	25.3	PC2	C	5589	37	0.41
KC3	Java	324	40	13.0	PC3	Java	1409	38	10.5
MC1	C	1952	39	1.8	PC4	C++	1270	38	13.9
MC2	C++	155	40	32.9	PC5	C	1694	39	27.0

软件缺陷预测为二分类问题,即 0 和 1 问题. 其中 1 为正例,代表有缺陷样例;0 为负例,代表无缺陷

样例. TP(真正例)表示正确预测有缺陷模块的数量,FP(假正例)表示错误预测无缺陷模块为有缺席模块,FN(假反例)表示错误预测有缺陷模块为无缺陷模块,TN(真反例)表示正确预测无缺陷模块的数量. 其混淆矩阵如表 2 所示.

表 2　混淆矩阵表

Table 2　Confusion matrix table

	预测结果	
	正例	负例
正例	TP	FN
负例	FP	TN

(1)ACC 是预测正确的结果占总样本的百分比,被称之为准确率,其公式如下:

$$ACC=\frac{TP+TN}{TP+YN+FP+FN}. \tag{3}$$

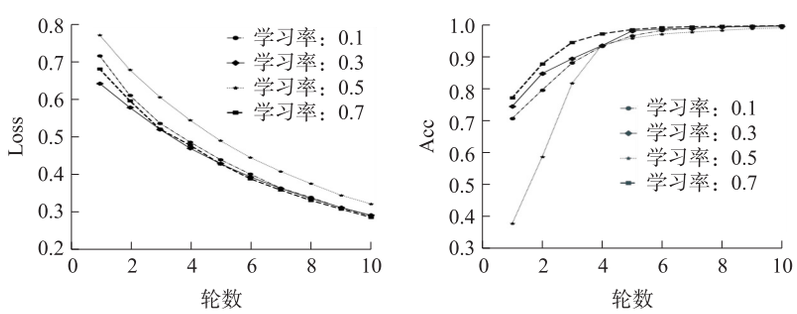
(2)AUC 是衡量"二分类问题"中机器学习算法性能的一种性能指标,AUC 是 ROC(Peceiver Operating Charateristic)曲线下的面积. 当 AUC 大于 0.5 时,模型性能越好,当小于等于 0.5 时,模型性能不佳. 考虑到软件缺陷数据存在严重的类不平衡问题,AUC 不会受到阀值和类不平衡问题的影响,本文采用 AUC 和 ACC 作为评价指标.

3.2　实验设置

本文基于迁移联邦学习方法实现软件缺陷预测任务,为保证实验结果的可参考性,选择 NASA 软件缺陷预测公开数据集进行实验. 其中 NASA 数据集中包含 12 个软件数据项目,每个项目间相互独立. 为模拟联邦学习场景,将不同软件项目作为若干本地参与方,并使用 CNN 全局模型进行模拟[26-28]. 实验在 NVIDIA GTX 950 4 GB 和 12 GB 内存的 Windows10 环境下运行,且所有程序均由 Python 编程语言和 Pytorch1.7 实现.

3.3　学习率对 FT-CPDP 方法的影响

在本节分析不同的学习率对 FT-CPDP 方法的影响. 采用控制变量法保持参与的客户端数量、隐私预算和训练轮数不变,只改变学习率大小进行实验. 图 2 是不同学习率对 FT-CPDP 的性能影响. 图 2(a)为在不同学习率下,随着全局轮数的增加损失函数的变化,图 2(b)为不同学习率下准确率的比较. 结果显示,当学习率为 0.5 时其表现效果最差,需要更多的训练轮数才能达到好的效果. 当学习率为 0.7 时,FT-CPDP 的 loss 和分类准确性有好的表现. 在不同的场景下学习率有所不同,需要进行多次实验. 适合某种场景下的学习率有利于模型的快速收敛和提高模型的分类性能.



(a) 损失函数图　　(b) 准确率图

图 2　不同学习率下的损失函数和准确率比较

Fig. 2　Comparison of loss function and accuracy under different learning rates

为更直观的显示模型性能,将三种算法的 AUC 结果用折线图进行展示. 图 3 列出不同算法下的 AUC 指标折线图,横坐标表示不同的目标数据集,依次为上述表中数据集,纵坐标表示模型性能 AUC. 折线图中直观显示本文提出的 FT-CPDP 算法比传统的跨项目软件预测方法在 AUC 表现上具有更好的性能,跨项目软件缺陷方法中 LOP 算法比 Peters 过滤法相对较好,但需要一个可信的第三方,否则无法保证数据隐私. 相较于跨项目软件缺陷预测方法,本文提出的 FT-CPDP 算法在软件缺陷预测上不仅提供了隐私保护效果,还提高了模型性能.

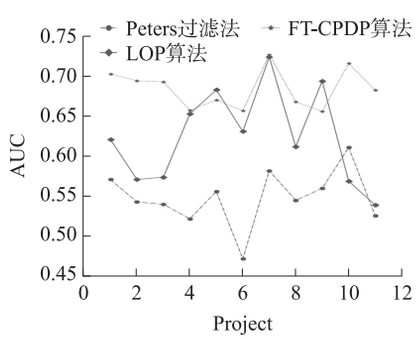


图 3　不同算法下的 AUC 比较

Fig. 3　AUC comparison under different algorithms

表 3　AUC 比较

Table 3　AUC comparison

	CM1	JM1	KC1	KC3	MC1	MC2	MW1	PC1	PC2	PC3	PC4
过滤法	0.57	0.542	0.539	0.521	0.555	0.471	0.581	0.544	0.559	0.61	0.525
LOP	0.62	0.57	0.573	0.652	0.682	0.63	0.723	0.611	0.693	0.568	0.538
FLT-SDP	0.702	0.693	0.691	0.656	0.669	0.656	0.725	0.667	0.655	0.715	0.681

4　结论

本文提出了一种基于联邦迁移学习的软件缺陷预测方法(FT-CPDP),该方法基于 CNN 构建模型,采用迁移学习方法解决源数据集和目标数据集特征异构问题,在隐私预算的前提下,提供了对传递参数的保护. 首先,分析了现有软件缺陷预测中存在的隐私问题,以及同态加密技术使得通信代价变高的问题. 其次详细描述了本文提出的 FT-CPDP 方法. 最后在 NASA 数据集下进行多场景实验,并与传统的跨项目软件缺陷预测方法进行比较. 结果表明,FT-CPDP 方法不仅提高了模型性能,还在通信代价和隐私保护上具有很好的效果.

本文提出的 FT-CPDP 方法可以在数据异构的场景下实现项目间的联合建模,然而低质量数据会降低最终的模型性能. 在未来的研究中将侧重于如何减少低质量数据对模型影响,以及如何聚合服务器端参数,为以后的软件缺陷预测研究提供新的思路.

[参考文献]

[1] 李勇,黄志球,房丙午,等. 代价敏感分类的软件缺陷预测方法[J]. 计算机科学与探索,2014,8(12):1442-1451.

[2] 李勇,黄志球,王勇,等. 数据驱动的软件缺陷预测研究综述[J]. 电子学报,2017,45(4):982-988.

[3] 刘文英,林亚林,李克文,等. 一种软件缺陷不平衡数据分类新方法[J]. 山东科技大学学报(自然科学版),2021,40(2):84-94.

[4] 曲豫宾,陈翔,李龙,等. 可缓解类重叠问题的跨版本软件缺陷预测方法[J]. 吉林大学学报(理学版),2021,59(2):372-378.

[5] 盖金晶,郑尚,于化龙,等. 一种跨项目缺陷预测的源项目训练数据选择方法[J]. 南京师大学报(自然科学版),2022,45(1):110-117.

[6] 倪超,陈翔,刘望舒,等. 基于特征迁移和实例迁移的跨项目缺陷预测方法[J]. 软件学报,2019,30(5):1308-1329.

[7] 李勇,黄志球,王勇,等. 基于多源数据的跨项目软件缺陷预测[J]. 吉林大学学报(工学版),2016,46(6):2034-2041.

[8] CHEN X,ZHANG D,CUI Z Q,et al. DP-share:privacy-preserving software defect prediction model sharing through differential privacy[J]. Journal of computer science and technology,2019,34(5):1020-1038.

[9] CHEN Y, QIN X, WANG J, et al. FedHealth: a federated transfer learning framework for wearable healthcare[J]. IEEE intelligent systems,2020,35(4):83-93.

[10] CHEN Y, LU W, WANG J, et al. Federated learning with adaptive batchnorm for personalized healthcare[J/OL]. arXiv Preprint arXiv:2112.00734,2021.

[11] ZHANG W,LI X. Federated transfer learning for intelligent fault diagnostics using deep adversarial networks with data privacy[J]. IEEE/ASME transactions on mechatronics,2022,27(1):430-439.

[12] WANG A, ZHANG Y, YAN Y, et al. Heterogeneous defect prediction based on federated transfer learning via knowledge distillation[J]. IEEE access,2021,9:29530-29540.

[13] SHARMA S, CHAOPING X, LIU Y, et al. Secure and efficient federated transfer learning[C]//2019 IEEE International Conference on Big Data. New York:IEEE,2019.

[14] JU C,GAO D,MARE R,et al. Federated transfer learning for EEG signal classification[C]//IEEE Engineering in Medicine and Biology Society Conference Proceedings. Motreal,Canada:IEEE,2020.

[15] 孔秀平,陆林. 隐私保护下的车辆轨迹联邦嵌入学习与聚类[J]. 南京师范大学学报(工程技术版),2022,22(2):80-86.

[16] TANG S,HUANG S,ZHENG C,et al. A novel cross-project software defect prediction algorithm based on transfer learning[J]. Tsinghua science and technology,2022,27(1):41-57.

[17] WU X, ZHANG Y, SHI M, et al. An adaptive federated learning scheme with differential privacy preserving[J]. Future

generation computer systems,2022,127:362-372.
[18] 叶青青,孟小峰,朱敏杰,等. 本地化差分隐私研究综述[J]. 软件学报,2018,29(7):25.
[19] GF A,RS B. On the behavioral implications of differential privacy[J]. Theoretical computer science,2020,841:84-93.
[20] LI H,HYB C,LANG L,et al. MHAT:an efficient model-heterogenous aggregation training scheme for federated learning[J]. Information sciences,2021.
[21] 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.
[22] ZHANG W,LI X,MA H,et al. Federated learning for machinery fault diagnosis with dynamic validation and self-supervision[J]. Knowledge-based systems,2021,213(1):106679.
[23] 张泽辉,富瑶,高铁杠. 支持数据隐私保护的联邦深度神经网络模型研究[J]. 自动化学报,2022,48(5):1273-1284.
[24] 杨庚,王周生. 联邦学习中的隐私保护研究进展[J]. 南京邮电大学学报(自然科学版),2020,40(5):204-214.
[25] RODRIGUEZ D,HERRAIZ I,HARRISON R. On software engineering repositories and their open problems[C]//First International Workshop on Realizing Artificial Intelligence Synergies in Software Engineering(RAISE'12). New York:IEEE,2012.
[26] FANG H,QIAN Q,CHEN M L,et al. Privacy preserving machine learning with homomorphic encryption and federated learning[J]. Future internet,2021,13.
[27] 贾峰,李世豪,沈建军,等. 采用深度迁移学习与自适应加权的滚动轴承故障诊断[J]. 西安交通大学学报,2022,56(8):1-10.
[28] MG A,KE P A,YU X A,et al. Preserving differential privacy in deep neural networks with relevance-based adaptive noise imposition-ScienceDirect[J]. Neural networks,2020,125:131-141.

[责任编辑:杜忆忱]