控制与决策
Control and Decision

面向分布式系统标签噪声的时间序列分类方法

林子谦, 张坤, 樊重俊, 杨夏洁

引用本文:
林子谦, 张坤, 樊重俊, 等. 面向分布式系统标签噪声的时间序列分类方法[J]. 控制与决策, 2024, 39(12): 4118-4126.

在线阅读 View online: https://doi.org/10.13195/j.kzyjc.2023.1576

面向分布式系统标签噪声的时间序列分类方法

林子谦[1], 张 坤[2], 樊重俊[1†], 杨夏洁[1]

(1. 上海理工大学 管理学院，上海 200093；2. 上海财经大学 信息管理与工程学院，上海 200433)

摘 要: 时间序列数据广泛存在于工业、医疗等应用领域的分布式边缘设备中,由于其往往具备人类不可识别的特征,基于现实数据的时间序列分类任务中普遍存在数据“孤岛”和标注错误等问题. 为解决分布式数据环境下这一困难,提出一种联邦时序过滤框架,该框架充分考虑自监督对比学习在提取复杂时序数据表征的优越性,并结合联邦学习方法来解决分布式系统的隐私安全问题,同时降低通信成本. 首先,通过在服务器上维护一套基准样本,使用基于区别对比损失和预测对比损失的时序增强预监督策略,通过预训练-微调方法获得一个高泛化时间序列表征能力的预监督模型;然后,引入一种新的标签噪声过滤的方法,利用由预监督模型指导的伪标签与本地标注的标签协同过滤设备中的噪声数据,并将干净数据集用于全局模型的训练;最后,根据各种标签噪声下对框架进行有效性验证,验证不同基准数据比例对于所构造框架的影响,并通过消融实验验证预监督模型各损失的过滤效果.

关键词: 联邦学习；自监督学习；时间序列分类；标签噪声；分布式系统

中图分类号: TP391 **文献标志码:** A

DOI: 10.13195/j.kzyjc.2023.1576

引用格式: 林子谦,张坤,樊重俊,等. 面向分布式系统标签噪声的时间序列分类方法[J]. 控制与决策, 2024, 39(12): 4118-4126.

Time series classification method for distributed system label noise

LIN Zi-qian[1], *ZHANG Kun*[2], *FAN Chong-jun*[1†], *YANG Xia-jie*[1]

(1. Business School，University of Shanghai for Science and Technology，Shanghai 200093，China；2. School of Information Management and Engineering，Shanghai University of Finance and Economics，Shanghai 200433，China)

Abstract: Distributed edge devices in the industrial, healthcare, and other application fields frequently contain time series data. Due to the often unrecognizable features it possesses, there are common issues in time series classification tasks based on real-world data, such as ‘data islands’ and labeling errors. To address this difficulty in distributed data environments, a federated temporal filtering framework is proposed. It incorporates the advantages of self-supervised contrastive learning in extracting complex temporal data representations and is combined with the federated learning approach to tackle the privacy and security issues of distributed systems, while also reducing the communication cost. By maintaining a set of benchmark samples on the server, this paper employs a time-series augmented pre-supervised strategy that relies on distinguishing contrast loss and predicting contrast loss. A pre-supervised model with a high-capacity for generalizing time-series characterizations is achieved through a pre-training and fine-tuning methodology in this approach. Meanwhile, a new approach for label noise filtering is introduced, which utilizes pseudo-labels guided by the pre-supervised model to filter the noisy data in the device in concert with local dataset labels, and uses the clean dataset for the training of the global model. Finally, this paper validates the framework’s effectiveness across different types of labeling noise, examines the impact of varying baseline data ratios on the constructed framework, and confirms the filtering effects of each loss in the pre-supervised model through ablation experiments.

Keywords: federated learning；self-supervised learning；time series classification；label noise；distributed system

0 引 言

随着工业互联网的普及,可穿戴设备和机器传感器可通过“端-边缘-云”系统获取大量时间序列数据[1]. 通过获取到的大规模、多源数据,提供高质量的智能服务和产品成为了可能,如智能医疗系统中监测患者的脑电图序列[2],智能工业控制系统中故障诊断的传感器读数[3]等. 但是在数据采集过程中,物联网设备往往会收集和处理大量的私有数据,企业主体和个人数据的隐私问题引发了各界关注. 如何更好地利用“孤岛”形式的分布数据,进而推动数据科学和智能系统技术的发展成为学界亟需解决的问题[4]. 联邦学习 (federated learning, FL) 是解决上述问题的一

收稿日期: 2023-11-11；**录用日期:** 2024-03-26.

责任编委: 李少远.

†通讯作者. E-mail: fan_chj@163.com.

*本文附带电子附录文件,可登录本刊官网该文“资源附件”区自行下载阅览.

类机器学习方法,通过多个客户的数据来协同训练得到一个全局模型,FL可在充分利用多源数据的同时,保证分散设备上的数据私有性[5-6]. 然而,多源私有数据同时意味着数据的质量良莠不齐,数据安全性和模型性能成为了FL实际应用中面临的权衡,因此,在保证安全性的前提下,更高效地利用多源、不均衡和不稳健数据集来保证全局模型的性能,是FL领域近年来的研究热点.

分布式物联网环境中的标签噪声问题十分普遍,因为实际场景下常常面临设备多样性、标注专家技能的差异、潜在的偏见以及恶意篡改问题[7],已有许多研究者针对标签噪声问题展开研究,主要可分为两类:一类是通过设计的训练损失来减少噪声样本的影响或正则化,主要是根据噪声样本对损失的贡献而后进行适应性的重新加权[8-9];另一类是通过额外的干净数据集,利用一种数据增强的方法获得额外的监督信息,这种方法已应用于许多机器学习问题,如监督学习、半监督学习、自监督对比学习等[10-11]领域. 其中:后者更具有可操作性,通过获得额外数据获得可靠监督模型,处理噪声数据集且不损害数据隐私. 考虑到时间序列中的特征为时序的依赖与因果关系,想要利用可靠监督模型纠正噪声标签,必须解决以下两个关键问题: 1) 在时间序列分类任务中,获得一个预先定义和完美标签的基准数据集的成本很高,如何在有限的时间序列数据集中训练一个稳健的监督模型[12]; 2) 在边缘设备噪声标签过滤的过程中,如何结合本地标签与额外的监督模型采样出干净的数据集.

鉴于上述问题的描述,可从两个方面直观考虑: 1) 增强时间序列数据和对比学习 (contrastive learning, CL) 的方法,可在少量的数据集中训练出一个稳健的监督模型[13]. 2) 额外的监督模型产生的伪标签与边缘设备标注的标签互补能够得到真实可靠的标签数据. 具体而言,本文提出一种两阶段的FL框架,称之为联邦时序过滤算法 (federated timing filtering, FedTF). 在第1阶段,结合数据增强和CL的方法,提出区分对比损失与预测对比损失. 其中:区分对比损失旨在最大化相同样本间的相似性,同时最小化不同样本间的相似性;预测对比损失是利用一个增强数据的潜在特征来预测原始数据的未来,使得模型通过更难的预测任务来学习稳健的表征. 此外,为了加深监督模型对时间序列类别的理解,将基准数据对预监督模型微调. 第2阶段主要提出一种噪声过滤的FL方法,在训练开始,通过预监督模型生成的伪标签以及与边缘设备原有标注标签相互比对,建立一组可信任的数据集索引,用于后续全局模型的训练. 本文的主要内容如下: 1) 设计一个两阶段的FL框架. 在FedTF框架能够利用服务器中产生的预监督模型对设备中数据进行干净数据的采样,保证数据安全与FL模型的泛化能力和性能. 2) 设计两种针对时间序列预监督训练的对比损失. 基于增强数据与原始数据的关系,设计区分对比损失和预测对比损失,实现在少量的数据基准中获取有效的外监督模型. 3) 提出一种在时间序列分类中标签噪声过滤的方案并应用于FL框架. 主要是通过边缘设备标签与预监督模型产生的伪标签的互补过滤,实现干净数据的采样,并用于全局模型的更新. 4) 在3个实际物联网数据集上测试FedTF框架. 实验中不仅在不同噪声下验证模型的有效性,同时通过分析不同基准数据的比例以及消融实验表明所提出方法对于全局模型性能的稳健性和可行性.

1　背景知识

1.1　时间序列对比学习

在图像识别领域,CL已被广泛应用,主要思路是在对比损失学习中,通过数据增强提取特征向量来提高模型的识别性能[14]. 与图像数据不同,对序列数据实施数据增强时要特别关注新生成数据间的时序关系. 在时间序列领域通常利用翻转、拉伸或添加噪声等方式进行数据增强实现CL[15]. 时间序列对比模型如图1所示:CL可以最大化相同序列间表示的特征相似性,并最小化不同序列间的特征相似性. 对于一个未标记的输入序列x,与一个随机噪声变换生成数据增强序列x^+,将两个数据用于自监督模型,则有$Z_i=f(\theta,x_i)$、$Z_i^+=f(\theta,x_i^+)$分别为原始序列及其增强序列的特征映射,各自包含了N个不同特征集. 其优化目标可表示为

$$\arg\min_{\theta}\sum_{i=1}^{N}\mathcal{L}(f(\theta,x_i),\{f(\theta,x_m)\}_{m=1,i\neq m}^{2N}). \tag{1}$$

令$Z_M=\{f(\theta,x_m)\}_{m=1,i\neq m}^{2N}$为除$Z_i$外的所有特征集,共$M=2N-1$个,则$\mathcal{L}_{\mathrm{CL}}$为对于第$i$个样本的损失,计算方式如下所示:

$$\mathcal{L}_{\mathrm{CL}}(Z_i,Z_M)=-\log\frac{\exp(Z_i\cdot Z_i^+/\tau)}{\sum_{m=1}^{M}\exp(Z_i\cdot Z_m/\tau)}. \tag{2}$$

其中:最大化Z_i与其正例Z_i^+间的特征相似性,同时最小化所有其他$2N-2$个负例样本与一个正例样本(用于归一化)特征的相似度总和;参数τ为温度系数. 通过调整模型参数θ最小化整个训练集上的损失$\mathcal{L}_{\mathrm{CL}}$,以此提高模型对数据的特征表示能力及其在复

杂数据集上的泛化性.

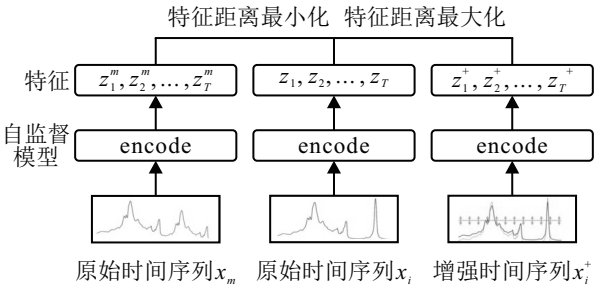


图 1 时间序列对比模型

1.2 联邦学习

经典的FL框架主要由一个服务器和多个设备组成. 其中:广泛应用的方法为联邦平均算法(federated averaging, Fedavg)[16],其模型的训练是通过循环进行的,每轮通过重复本地学习和模型聚合过程来实现,直至收敛. 在一轮中:服务器激活部分边缘设备,并将最新的模型θ_G发送给设备,被激活的设备在本地数据学习训练. 具体而言,每个设备$k \in K$同时对其本地模型参数θ_k执行随机梯度下降的参数优化,在各自私有数据集D_k上进行训练,使得本地损失最小化,其计算公式为

$$\theta_k^{t+1} = \theta_k^t - \eta \nabla_\theta \mathcal{L}_k(\theta_k^t). \tag{3}$$

其中:η为本地学习率,$\mathcal{L}_k$为本地的交叉熵损失. 中央服务器通过以下方式汇总本地模型:

$$\theta_G^{t+1} = \sum_{k \in K} \frac{|D_k|}{|D|} \theta_k^{t+1}. \tag{4}$$

这里:$|D|$为所有设备的总数据量,$|D_k|$为k设备的数据量. 持续以上过程,直至收敛. 在经典的FL框架中假定设备对所有数据标签均为真实且标准的. 然而,这一假设在时间序列分类中并不现实,因为准确标注需要足够的专业知识. 因此,要将FL应用于时间序列分类时,需要一种方法有效利用有限的标签,以达到较高的模型准确性.

2 问题描述

假设智能物联网系统由多个边缘设备和云服务器组成. 每个边缘设备通过本地传感器收集本地数据,并由本地专家或传感器的反馈进行标注. 时间序列的分类任务由云端服务器发起,其拥有一个小的基准数据集$D_B = (x_i, y_i)_{i=1}^N$. 这个数据集可根据云端服务器自身对学习任务的了解或专家的权威生成和标注,但是其规模不足以训练高度精确的模型. 边缘设备收集数据后由本地专家进行标注,由于标注专家技能的差异、潜在的偏见和恶意篡改问题,本地设备存在大量的标签噪声. 具体而言,服务器的任务是执行时间序列分类问题. 其中:c为类别数量,$\mathcal{X}$为输入空间,$\mathcal{Y} = [0,1]^c$为标签空间. 第k个边缘设备通过本地传感器收集一组时间序列样本,且在标注过程中可能会出现错误的标签或标注,y_i为由本地标注的标签,$\hat{y}_i$为真实的标签,在边缘设备中实际拥有的数据集为$D_k = (x_i, y_i)_{i=1}^N$. 该智能系统的目标是找到一个由深度学习模型θ定义的映射函数$f_k(\theta; x) : \mathcal{X} \to [0,1]^c$,以尽可能满足$f_k(\theta; x_i) = \hat{y}_i$. 为了实现这一目标,优化目标可表示为

$$\arg\min_\theta \sum_{i=1}^N \mathcal{L}_k(f_k(\theta; x_i), \hat{y}_i), \tag{5}$$

其中$\mathcal{L}_k$为第k个设备的交叉熵损失函数. 若有K个边缘设备,其数据集表示为$\{D_1, D_2, \ldots, D_k\}$,服务器的目标是在不交换原始数据的情况下,学习一个机器学习模型θ_G,则服务器任务的优化目标是解决以下最优化问题:

$$\arg\min_{\theta_G} \mathcal{L}(\theta_G) = \sum_{k=1}^K \frac{|\mathcal{D}_k|}{|\mathcal{D}|} \mathcal{L}_k(f(\theta_G; x_i), \hat{y}_i). \tag{6}$$

3 FedTF框架

所提出FedTF框架的概述如图2所示,主要由两个阶段6个步骤组成.

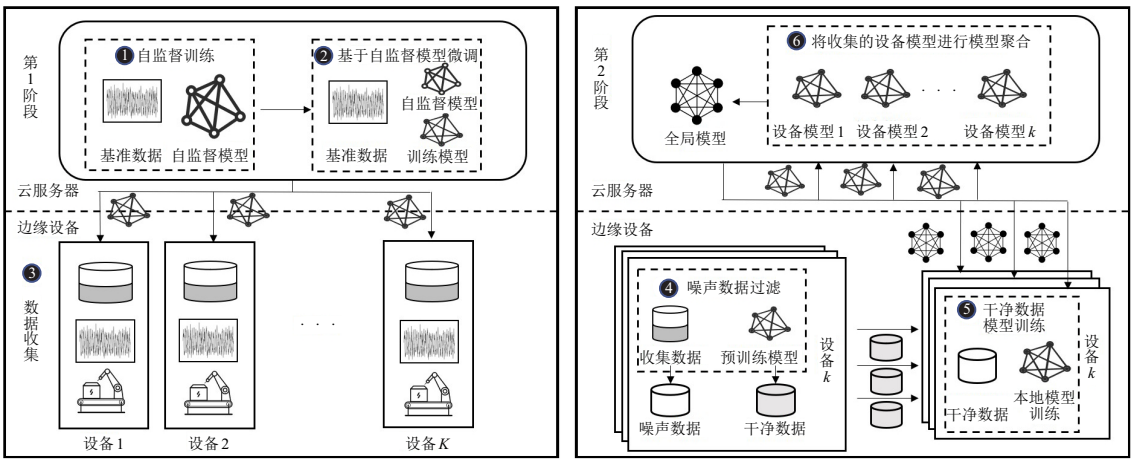


图 2 FedTF框架流程

step 1: 服务器使用有限的D_B基准数据通过最小化预监督损失$\mathcal{L}_{\text{unsup}}$,获得预监督模型$\theta_{\text{pre}}$.

step 2: 服务器利用D_B基准数据和预监督模型θ_{pre},基于微调的模型损失$\mathcal{L}_{\text{semi}}$,生成用于设备干净数据集采样的外监督模型$\hat{\theta}_{\text{pre}}$. 然后,将外监督模型$\hat{\theta}_{\text{pre}}$和初始化训练模型$\theta_G^0$发送至设备.

step 3: 边缘设备k同时接收训练任务和预监督模型,然后根据任务需求收集时间序列数据集X,随后专家对该数据集标注生成Y,并建立私有的本地数据集D_k.

step 4: 边缘设备k利用外监督模型$\hat{\theta}_{\text{pre}}$对本地私有数据$D_k$噪声过滤,以筛选出干净的标签$\hat{D}_k$数据集.

step 5: 在接收到全局模型θ_G后,边缘设备k利用式(3)对其本地的干净数据集$\hat{D}_k$进行训练.

step 6: 服务器使用式(4)聚合边缘设备在第t轮训练时上传的模型θ_k^t,并生成当前的全局模型θ_G^t,将其发送给边缘设备下一轮迭代. 这一过程在全局模型θ_G^t收敛前反复执行step 5和step 6.

3.1　第1阶段:外监督模型构建

如图3所示,预监督模型训练分为4个步骤. 接下来,将详细介绍步骤中的细节.

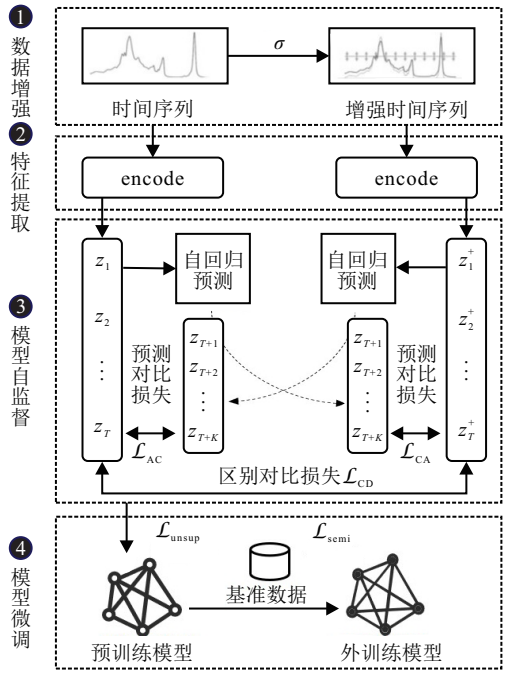


图 3　外监督损失构建

3.1.1　数据增强和特征提取

本文通过将时间序列分为随机的片段,然后重新组合新的时间序列的排列数据增强方式,使得训练的模型可识别在不同时间顺序下保持不变的特征,提高模型对于时间扰动的鲁棒性. 这对于对比学习中特征提取的应用尤为重要. 在给定原始数据x_i,基于排列的增强数据为x_i^+. 因此,在FedTF框架中,在服务器存在的少量基准数据为$D_B = (x_i, y_i)_{i=1}^N$. 其中:$x_i = [t_1, t_2, \ldots, t_T]$包含$T$个有序的实数,通过数据增强生成$x_i^+ = [t_1^+, t_2^+, \ldots, t_T^+]$. 通过优化预监督模型$\theta_{\text{pre}}$(详见后文第4.1节)得到有效的特征$Z_i = f(\theta_{\text{pre}}; x_i)$和$Z_i^+ = f(\theta_{\text{pre}}; x_i^+)$. 其中:$Z_i = [z_1, z_2, \ldots, z_T]$, $Z_i^+ = [z_1^+, z_2^+, \ldots, z_T^+]$. 因此,基准数据通过数据增强和特征提取得到两个特征集,分别为$Z = [Z_1, Z_2, \ldots, Z_N]$和$Z_i^+ = [Z_i^+, Z_2^+, \ldots, Z_N^+]$,用于下一步的对比损失计算.

3.1.2　外监督模型损失

区分对比损失. 本文采用非线性投影头对原始特征和增强特征进行非线性变换,类似于先前时间序列数据增强的方法[17],目的是增加预监督模型的区分性,使其在表示空间$\mathcal{X}$中更易区分. 已知基准数据中产生共$2N$个特征,其中预监督模型将(Z_i^+, Z_i)视为一对正样本. 与此同时,来自同一批次中其他输入样本的剩余$(2N-2)$个数据集被视为Z_i的负样本. 基于此,区分对比损失的目标是最大化样本与其正样本间的相似性,同时最小化样本与其负样本间的相似性,使得最终的表示更具区分性. 在式(2)的基础上增加余弦相似度,具体如下所示:

$$l_d(Z_i, Z_M) = -\log \frac{\exp(\text{sim}(Z_i, Z_i^+)/\tau)}{\exp\Big(\sum_{m=1}^{M} \text{sim}(Z_i, Z_m)\Big)\Big/\tau}, \quad (7)$$

$$\mathcal{L}_{\text{CD}} = \frac{1}{N}\sum_{i=1}^{N} l_d(Z_i, Z_M). \quad (8)$$

其中:$\text{sim}(\cdot,\cdot)$为余弦相似度,τ为一个温度参数,Z_M为除Z_i外的所有特征集.

预测对比损失. 为了捕捉时间维度上的潜在结构,本文尝试探索时间序列中预测的时间关系[18]. 本节通过自回归模型预测两个特征向量未来的特征,然后以交互的方式最大化特征相关性,使其提取的特征更具有鲁棒性和有用的信息. 具体而言,基于通过使用自回归模型生成未来的时间序列$Q_i \in [z_{T+1}, z_{T+2}, \ldots, z_{T+k}]$. 因此,通过最大化预测增强数据的特征$Q_i^+$与同样本的真实特征$Z_i$间的相似度,使其对相同样本的特征表示更加一致. 同时,通过最小化不同样本预测特征间的相似度,促使预监督模型更好地区分不同样本间的差异,从而增强特征的可区分性. 因此在样本为N时,噪声对比损失为$\mathcal{L}_{\text{AC}}$和$\mathcal{L}_{\text{CA}}$,有

$$\mathcal{L}_{\text{CA}} = -\frac{1}{N}\sum_{i=1}^{N} \log \frac{\exp(I(Q_i^+, Z_i))}{\exp\Big(\sum_{m=1}^{N} I(Q_i^+, Z_m)\Big)}, \quad (9)$$

$$\mathcal{L}_{\mathrm{AC}}=-\frac{1}{N}\sum_{i=1}^{N}\log\frac{\exp(I(Q_i,Z_i^+))}{\exp\Big(\sum\limits_{m=1}^{N}I(Q_i,Z_m^+)\Big)},\quad(10)$$

其中$I(\cdot,\cdot)$为互信息相关性.值得注意的是,为了使得编码器更具泛化能力,在区分性损失和预测对比损失中分别使用余弦相似度和互信息相似性,这是两种不同的衡量方法. 因此,总的预监督对比损失具体为

$$\mathcal{L}_{\mathrm{unsup}}=\lambda_1\cdot(\mathcal{L}_{\mathrm{CA}}+\mathcal{L}_{\mathrm{AC}})+\lambda_2\cdot\mathcal{L}_{\mathrm{CD}},\quad(11)$$

这里λ_1和λ_2为固定的标量超参数. 预监督训练的迭代方式如算法1中第1行～第4行所示.

算法1 联邦过滤时序算法伪代码.

输入: 全局模型θ_G^0,基准模型D_B,设备数据D_k;

输出: 全局模型θ_G^{T}.

1. 服务器:将基准数据D_B数据增强生成$\hat{D}_B$
2. for $t=1,2,\ldots,O$ do //预监督模型训练
3. $\theta_{\mathrm{pre}}^{t+1}=\theta_{\mathrm{pre}}^{t}-\eta\nabla_\theta\mathcal{L}_{\mathrm{unsup}}(\theta_{\mathrm{pre}}^{t},D_B)$
4. end
5. for $t=1,2,\ldots,O$ do //微调监督模型训练
6. $\hat{\theta}_{\mathrm{pre}}^{t+1}=\hat{\theta}_{\mathrm{pre}}^{t}-\eta\nabla_\theta\mathcal{L}_{\mathrm{semi}}(\hat{\theta}_{\mathrm{pre}}^{t},\theta_{\mathrm{pre}},D_B)$
7. end
8. for $t=1,2,\ldots,T$ do //全局模型训练
9. for $k=1,2,\ldots,K$所有设备平行 do
10. $\theta_k=$设备端$(\hat{\theta}_{\mathrm{pre}},\theta_G,D_k)$
11. end
12. $\theta_G^{t+1}=\sum\limits_{k\in K}\dfrac{|D_k|}{|D|}\theta_k^{t+1}$
13. end
14. return θ_G^{t+1}
15. 设备端$k(\hat{\theta}_{\mathrm{pre}},\theta_G,D_k)$: //本地模型训练
16. 设备k根据式(12)和(13)采样干净数据集$|\hat{D}|$
17. for $t=1,2,\ldots,\mathcal{T}$ do
18. $\theta_k^{t+1}=\theta_k^{t}-\eta\nabla_\theta\mathcal{L}_k(\theta_k^{t})$
19. end
20. return θ_k^{t+1}

微调损失. 在微调阶段,为了更好地满足任务需求,对已训练好的预监督模型进行基准数据微调,以加深外监督模型对时间序列特征的理解[19]. 为此,基于预监督对比损失的式(8)可改写为

$$\mathcal{L}_{\mathrm{SCD}}=\frac{1}{2N}\sum_{i=1}^{N}\sum_{j=1}^{N}\delta_{ij}l_d(Z_i,Z_M).\quad(12)$$

其中:δ_{ij}表示当$y_i=y_j$时取值为1;否则,为0. 因此,总微调模型损失可表示为

$$\mathcal{L}_{\mathrm{semi}}=\lambda_3\mathcal{L}_{\mathrm{CE}}+\lambda_4\mathcal{L}_{\mathrm{SCD}}.\quad(13)$$

这里:λ_3和λ_4为固定的标量超参数,$\mathcal{L}_{\mathrm{CE}}$为交叉熵损失. 微调的迭代方式在算法1中第5行～第7行.

3.2 第2阶段:全局模型训练

FedTF的第2阶段由step 4～step 6组成. 每轮开始时,所选客户端会从服务器获取全局模型参数和预监督模型参数,以便进行本地的噪声过滤和更新. 随后,进行全局模型的更新,主要包括对各客户端的模型进行聚合并将更新后的模型传回客户端,然后反复迭代此过程,直至模型收敛.

边缘设备本地更新. 在本地更新前,所选客户端会下载外监督模型参数,并使用下文式(14)在其本地数据集上对模型进行训练和噪声过滤. 需要注意的是,数据选择过程中不应完全信任提供的标签,因为它们可能没有被准确标注. 同样,也不应完全依赖于预监督模型建立的伪标签,因为这可能会导致由于相似性而出现的错误标签. 因此,本节提出了基于伪标签和本地标注的标签互补使用,有助于找到可信的数据集向量. 通过边缘设备k中的本地标注和预监督模型建立的伪标签,可得到用于干净数据向量$m_k(i)$,即

$$m_k(i)=\begin{cases}1,\ \tilde{y}_{k,i}=y_{k,i};\\ 0,\ \tilde{y}_{k,i}\neq y_{k,i}.\end{cases}\quad(14)$$

其中:$y_{k,i}$为第k个边缘设备本地自己标注的标签;$\tilde{y}_{k,i}$为由外模型在第k个边缘设备中建立的伪标签,具体生成方式表示为

$$\tilde{y}_{k,i}=f_k(\hat{\theta}_{\mathrm{pre}},x_{k.i}),\quad(15)$$

这里$\hat{\theta}_{\mathrm{pre}}$为从服务器中接收的外监督模型参数. 在第1轮中确定好的干净数据集$m_k(i)$样本索引后,本地模型基于干净数据的索引训练本地模型,并上传至服务器,其因此根据式(5)中的损失,加上干净数据向量,目标是最小化损失函数,第k个设备的优化模型为

$$\arg\min_{\theta}\sum_{i=1}^{N}m_k\mathcal{L}_k(f_k(\theta;x_i),y_i).\quad(16)$$

全局模型聚合更新. 在第1轮中边缘设备根据式(16)将噪声数据集D_k转为干净数据集$\hat{D}_k=(x_i,\hat{y}_i)_{i=1}^{n}$,其中$n\in N$. 同时,在干净数据进行本地更新训练后,边缘设备会上传模型参数θ_k^t至服务器. 然后,根据Fedavg的加权平均聚合的方法,服务器将上传的本地模型参数聚合如下所示:

$$\theta_G^{t+1}=\sum_{k\in K}\frac{|\hat{D}_k|}{|\hat{D}|}\theta_k^{t+1}.\quad(17)$$

其中:θ_G^{t+1}为下一轮的全局参数,$|\hat{D}|$和$|\hat{D}_k|$分别为所有设备干净数据总数和第k个客户的干净数据数量. 全局的迭代方式在算法1中第8行～第20行表示.

4 实验结果

4.1 实验装置

实验环境:所有实验均在英伟达3090 GPU上执行,使用PyTorch 1.4.0和Python 3.8.13,同时采用PFL

平台[20]. 本文仅考虑数据噪声对FL模型的影响,最终结果取最后10轮的平均准确率. 除非另有说明,FL中用于外监督的基准数据占20%,它与训练数据为互斥关系,剩余数据将会分给30个边缘设备. 每个设备的本地迭代周期被设置为5次,每轮从30台设备中选择,共进行100轮,预监督和外监督阶段共进行100轮. 所有实验中的批量大小均设置为128,学习率为0.01,权重衰减为0.01,FL中采用SGD优化器,动量为0.9,预监督的优化器Adam的超参数设置为$\beta_1=0.9$和$\beta_2=0.99$,数据增强的比率设置为1.1,超参数设定为$\lambda_1=1$、$\lambda_2=0.7$,在区别对比实验中将τ设置为0.2.

实验模型. 在本文中,模型主要有自监督模型和监督模型两种:自监督模型用于预训练,监督模型用于微调和全局训练. 1) 自监督模型为一个4层全连接网络构成的投影头,其主要作用是对序列变换器的输出进行有效降维. 该投影头融入ReLU激活函数和批量归一化操作,利用这两种策略的结合显著提升了模型在捕捉和表达特征方面的能力,增强其在处理复杂数据时的稳定性和鲁棒性. 该模型的设计灵感来源于BERT架构,特别是在处理序列信息的方式上,采用了类似于BERT中的[CLS]标记[21]的策略,通过引入一个专门的类标记来捕捉整个序列的综合信息. 此外,变换器的设计不仅使得模型能够全面处理整个序列,且能够提取并利用全局上下文信息,揭示复杂的时间依赖关系,有效地分析时间序列内在关系. 2) 监督模型基于3层卷积神经网络结构,其核心结构由3个卷积块组成,每个卷积块包含以下4层:一个卷积层,用于特征提取;一个归一化层,确保特征处理的稳定性;激活函数,引入非线性来增强模型的表达能力;一个池化层,减少特征的维度并提高计算效率. 最终产生的logits向量专门用于交叉熵损失计算,确保针对特定特征提取任务的优化. 通过其独特的架构,能够有效捕捉数据中的短期和长期依赖性,专注于提取时间序列结构中的关键特征. 表1为两个模型在3个数据集中对应的模型结构. 其中:时间对比模型包括线性层(步长×维度×输出通道数)、投影头(线性层维度×归一层维度×线性层维度)、序列变化层(输出通道数×维度×层数×投影头);监督模型包括3个卷积层(卷积核数×归一化层维度×内核尺寸)和全连接层(维度×渠道数×类别)以及步长.

表1　数据集相关参数

数据信息	训练样本数	测试样本数	样本长度	传感器通道数	类别数
HAR	7352	2947	128	9	6
Sleep-EDF	25126	8910	3000	1	5
FD	8184	2728	5120	1	3

预监督模型	线性层	摄影头层	序列变换器层
HAR	50×64×128	64×64×32	100×64×64×4×4
Sleep-EDF	6×100×128	100×64×64	128×64×64×4×4
FD	50×60×128	64×64×32	128×64×64×4×4

训练模型结构	卷积块1	卷积块2	卷积块3	全连接层	步长
HAR	9×32×8	32×64×8	64×128×8	128×18×6	1
Sleep-EDF	1×32×25	32×64×8	64×128×8	128×127×5	3
FD	1×32×32	32×64×8	64×128×8	128×162×3	4

实验数据. 为了全面评估所提出模型,使用3个不同的真实世界数据集:1) 人类活动识别(HAR)数据集[22]. UCI HAR数据集包含3位受试者在执行行走、上楼、下楼、站立、坐和躺下6种活动时的传感器读数. 2) 睡眠阶段分类(sleep-EDF)数据集[23]. 该数据集包括20位受试者的整夜多导睡眠图记录. 在记录中沿用了之前的研究,使用单个脑电图通道. 睡眠阶段分类指的是将输入的脑电信号分为5个等级:清醒、非快速眼动(3个等级)和快速眼动. 3) 故障诊断(FD)[24]. 故障诊断数据集是从轴承机器在4种不同工作条件下的传感器读数中收集的. 每个工况均可视为一个单独的域,因为它具有不同于其他工况的特性. 每个域有3个类别:内部故障、外部故障和一个健康类别,具体如表1所示.

数据仿真. 本文仅考虑标签噪声,初始30个设备的数据集的数据均为独立同分布,且数据量大小一致,噪声类型为对称翻转[25]. 本节针对标签噪声设计两个参数τ_1和τ_2,分别为噪声客户端的噪声水平下界和上界. 设备端的本地噪声水平通过均匀分布$U(\tau_1,\tau_2)$抽样,图4为HAR数据集噪声分布,圆圈大小反应了噪声的实际大小.

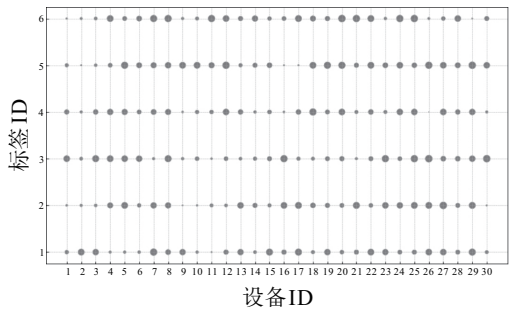


图4　HAR数据集噪声区间为$\tau_1=0,\tau_2=1$的分布

基准方法. 除Fedavg外,实验中还使用其他前沿基线对比评估: 1) FedAvg 和 Coteaching[25] 组合. Coteaching同时训练两个深度神经网络,使得它们相互教导,选择可能为干净标签的数据. 2) Fedavg和对称SCE组合[26]. SCE通过交叉熵与一个嘈杂标签鲁棒性的替代方法相结合. 3) FedLSR为一种局部自规范化方法,有效地减轻了噪声标签对性能的影响,并通过减小原始实例与增强实例间的模型输出差异来实现[27]. 4) Fedcorr为动态地识别噪声客户,通过利用模型预测的维度子空间的维度,在所有客户端上独立测量识别出噪声客户上的错误标签[28].

评估方法. 本文进行多项实验来验证FedTF的性能,使用两个指标来衡量性能,即准确率(acc)和宏F1得分(MF1-score),其中MF1-score为

$$\mathrm{MF1}=\frac{1}{C}\sum_{i=1}^{C}\frac{2\times\mathrm{Precision}_i\times\mathrm{Recall}_i}{\mathrm{Precision}_i+\mathrm{Recall}_i}. \tag{18}$$

这里: $\mathrm{Precision}_i=\dfrac{\mathrm{TP}_i}{\mathrm{TP}_i+\mathrm{FP}_i}$, $\mathrm{Recall}=\dfrac{\mathrm{TP}_i}{\mathrm{TP}_i+\mathrm{FN}_i}$, TP_i、FP_i和FN_i分别为第i类的真正例、假正例和假负例; N为样本总数; C为数据集中的类别总数. 本文均以百分比形式展示.

4.2 全局模型在噪声下性能的对比

表2为在不同噪声水平数据集上的测试精度. 实验结果表明: FedTF能够在不同噪声水平下显著抵御噪声标签的影响,且能够训练出高性能和稳健的全局模型. 相比于其他先进的方法,FedTF通常能够在大多数情况下达到更优的性能. 在HAR和Sleep-EDF领域,尽管在标签低噪声情况下,FedTF稍逊于Fedcorr,但是在高噪声环境下,FedTF表现卓越. 这表明: 尽管在使用本地的正则损失函数时设备能够有效地避免噪声的影响,但是使用额外监督模型来引导噪声过滤具有更高的稳定性和可行性; 当数据噪声复杂且特征多样时,使用损失或蒸馏方法并不是最佳选择,而使用外监督模型仍然能够实现相对出色的效果. 此外,由于FedTF方法为采样部分干净训练,相比于采样所有数据进行训练降低了一定的通信成本,表3为训练时间与收敛轮数对比. 由表3可见: 在HAR中相比于Fedavg,在100轮的训练中时间减少了1.14倍,收敛更快至1.73倍,消耗的通信成本为23.54 MB,相比于集中训练的通信成本减小了1.8倍. 在FD中相比于Fedavg,在100轮的训练中时间减少了2.45倍,收敛更快至1.45倍,消耗的通信成本为28.02 MB,相比于集中训练的通信成本减小了1.14倍. 在Sleep-EDF中相比于Fedavg,在100轮的训练中时间减少了2.31倍,收敛更快至1.59倍,消耗的通信成本为39.16 MB,相比于集中训练的通信成本减小了1.14倍.

表 2 在不同噪声水平数据集上的测试精度

数据集	HAR				Sleep-EDF				FD			
噪声比率	$\tau_1=0,\tau_2=1$		$\tau_1=0.5,\tau_2=1$		$\tau_1=0,\tau_2=1$		$\tau_1=0.5,\tau_2=1$		$\tau_1=0,\tau_2=1$		$\tau_1=0.5,\tau_2=1$	
方法	acc	MF1	acc	MF1	acc	MF1	acc	MF1	acc	MF1	acc	MF1
Fedavg	54.6+0.6	54.1+0.5	20.4+0.3	20.6+0.3	34.1+0.5	29.3+0.4	36.4+0.2	28.4+0.2	49.6+7.1	46.8+6.1	22.3+0.6	21.9+0.9
Fedavg+SCE	58.1+1.1	58.6+1.2	20.4+0.4	20.5+0.4	28.1+6.1	23.1+2.3	25.4+3.4	23.2+3.4	65.1+0.2	62.2+0.2	25.7+0.7	24.1+0.7
Fedavg+Coteaching	64.2+0.3	63.2+0.4	30.8+0.2	29.5+0.2	31.2+0.2	30.2+0.3	29.5+0.4	27.5+3.2	57.7+0.1	48.7+0.2	24.8+0.5	23.4+0.4
FedLSR	87.5+0.1	87.4+0.1	64.5+0.3	63.1+0.3	35.6+0.2	31.4+0.2	36.8+0.2	34.4+0.2	64.5+0.1	58.4+0.4	**37.2+0.1**	**38.4+0.2**
Fedcorr	**91.4+0.1**	**91.4+0.1**	57.3+0.2	56.1+0.3	**69.6+0.1**	60.1+2.1	28.8+9.1	24.5+8.4	70.9+0.2	69.4+0.2	26.1+0.6	26.2+0.5
FedTF	91.3+0.2	91.4+0.2	**89.3+0.2**	**88.2+0.1**	69.1+0.3	**62+8+0.4**	**55.6+0.4**	**47.6+0.4**	**75.9+0.1**	**73.1+0.2**	35.1+0.2	30.6+0.2

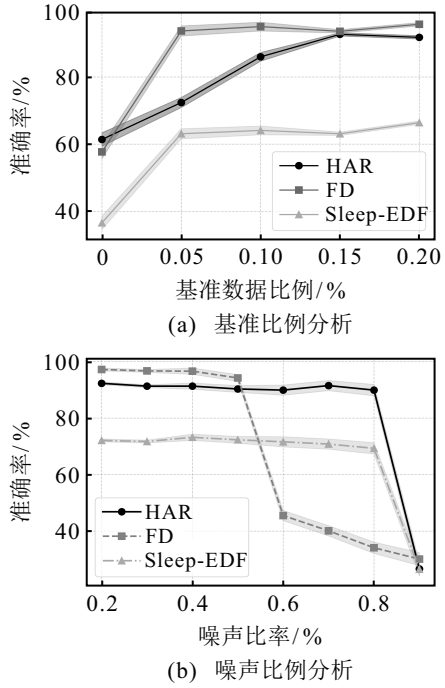
表 3 训练时间与收敛轮数对比

数据集	HAR		FD		Sleep-EDF	
方法	t/s	收敛轮数	t/s	收敛轮数	t/s	收敛轮数
Fedavg	177	80	982	78	714	76
Fedavg+SCE	166	66	495	75	509	71
Fedavg+Coteaching	210	90	444	86	531	84
FedLSR	266	88	684	89	498	87
Fedcorr	180	48	734	56	344	49
FedTF	155	46	426	49	291	48

4.3 全局模型鲁棒性区间分析

图5(a)为3个数据集在不同基准数据比例下FedTF的性能. 由图5(a)可见: 在没有基准数据的情况下,FedTF是没有噪声过滤能力的,因此在噪声环境下,性能明显较低. 当只有5%的基准数据时,相较于没有基准数据在HAR数据集中的性能提高了10%,FD数据集提高了30%,Sleep-EDF数据集提高了30%. 当基准数据占比为10%、15%、20%时,在3个数据集中,性能没有明显提升. 这表明,当基准数据占比约为10%时,外监督模型已具有良好的噪声过滤效果. 可得出如下结论: FedTF在过滤噪声方面是有效的,但是想要到达较为良好的过滤效果需要基准数据为10%左右. 图5(b)中: 在数据结构相对简单的HAR数据集中,当所有噪声均超过0.9时,模型准确率

才出现了明显的下降;而在Sleep-EDF和FD数据集中,噪声分别为0.5和0.8时,性能开始下降. 进一步分析发现,这是由于噪声数量过多,导致干净的数据集严重不足,在本地训练过程中无法提取有效的信息.



(a) 基准比例分析

(b) 噪声比例分析

图 5　基准数据比例和噪声程度线性图

4.4　外监督模型消融实验

本节研究FedTF中每个组成部分对噪声过滤的有效性,具体而言,实验中单独训练区别对比损失、预测对比损失以及总的无监督损失,并对预监督模型进行微调. 表4为不同损失与数据增强的消融实验对比. 由表4可见:无论是加入区别损失还是预测损失,对外监督模型过滤的性能均有一定的提升. 由预测对比损失的结果可见:预测任务生成了稳健的特征,在HAR上的准确率提高了2%以上,在SleepEDF上的准确率提高了4%,但是在FD中并没有明显的提升,原因在于FD数据集中的样本长度、复杂度较高,仅使用增强可能无法完成艰巨的时序预测任务. 值得注意的是,在预监督模型设置下,总损失并没有很好的过滤效果,原因可能如下:与预监督学习相比主要用于特征的提取,若只通过特征的提取没有很好地理解对应标签的外监督模型是不具有能力进行预测的. 对于微调后的总损失得出如下结论:模型在HAR过滤效果中提升了3%,在FD过滤效果中提升了14%,在Sleep-EDF过滤效果中提升了6%. 由此表明,FedFT能够有效利用现有的少量标记数据训练出具有良好过滤效果的外监督模型. 此外,本文进一步探讨了不同数据增强技术的效果,包括放缩、拉伸和旋转等方法. 通过对比实验发现:所选的排列方式在性能上表现最佳,排列方法在保持数据物理意义的同时,能够更有效地增强数据的多样性. 与表4中仅进行微调的模型相比,其他增强方法同样能够带来一定程度的性能提升,这一点验证了数据增强技术在提高外监督模型滤除噪声方面的有效性.

表 4　不同损失与数据增强的消融实验对比

数据集	HAR		FD		Sleep-EDF	
组件	acc	MF1	acc	MF1	acc	MF1
排列+预监督+预测对比损失	**52.7**	**34.0**	**53.2**	**49.6**	**43.2**	19.9
排列+预监督+区分对比损失	25.3	13.7	45.8	42.1	36.5	**23.2**
排列+预监督+总预监督损失	44.6	27.9	47.5	43.9	33.0	21.1
微调	92.9	86.4	84.7	82.8	85.2	78.1
排列+预测对比损失+微调	94.4	88.7	84.9	82.7	90.5	85.7
排列+区分对比损失+微调	93.2	86.7	95.2	94.4	88.2	83.0
排列+总监督损失+微调	**95.9**	**90.6**	**98.7**	**98.4**	**91.4**	**88.8**
缩放+总监督损失+微调	93.9	88.1	88.4	86.2	89.6	84.8
拉伸+总监督损失+微调	93.4	86.8	88.3	86.6	90.3	85.4
旋转+总监督损失+微调	93.7	87.1	89.3	88.7	90.6	82.8

5　结　论

本文围绕以下重要问题展开,即如何在分布式、多来源和不可信数据集设置下得到稳健的时间序列分类模型. 针对实际工业场景下的时序数据标签噪声问题,本文设计了一类FL框架,旨在通过边缘设备上的学习干净数据集,为时间序列分类任务提供稳健的FL训练结果. 在FL训练初期,通过增强时序数据挖掘时序依赖特征,提出了使用外监督模型用于纠正边缘设备的错误标签,克服了错误标签影响FL的全局模型稳健性的挑战. 所提出方法设计了时序区别对比损失和预测对比损失的预监督训练方法,提升了外监督模型特征提取的泛化性能,通过在服务器少量的基准数据上进行微调,实现能够有效过滤标签噪声的外监督模型. 然后,根据外监督模型设计了一种新颖的伪标签与本地边缘设备标签相融合的干净数据集采样方法,提高了FL框架性能. 最后,本文通过3个真实的时间序列分类数据集,验证了FedTF的方法能够达到比现有技术更好的精确度和稳定性. 此外,通过对不同数量的基准数据和不同比例的噪声数据进行深入分析,从而验证了FedTF方法的有效性.

参考文献(References)

[1] Yang L, Liao Y W, Cheng X, et al. Efficient edge data management framework for IIoT via prediction-based data reduction[J]. IEEE Transactions on Parallel and Distributed Systems, 2023, 34(12): 3309-3322.

[2] Raza A, Tran K P, Koehl L, et al. Designing ECG monitoring healthcare system with federated transfer learning and explainable AI[J]. Knowledge-Based Systems, 2022, 236: 107763.

[3] 代伟, 黄金昊, 王聪, 等. 基于多特征融合的工业气动调节阀快速自学习故障诊断方法[J]. 控制与决策, 2023, 38(10): 2934-2942.

(Dai W, Huang J H, Wang C, et al. Fast self-learning fault diagnosis method for industrial pneumatic control valves based on multi-feature fusion[J]. Control and Decision, 2023, 38(10): 2934-2942.)

[4] Mothukuri V, Parizi R M, Pouriyeh S, et al. A survey on security and privacy of federated learning[J]. Future Generation Computer Systems, 2021, 115: 619-640.

[5] 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.

[6] McMahan H B, Moore E, Ramage D, et al. Communication-efficient learning of deep networks from decentralized data[J/OL]. 2016, arXiv: 1602.05629.

[7] Angluin D, Laird P. Learning from noisy examples[J]. Machine Language, 1988, 2(4): 343-370.

[8] Lyu Y M, Tsang I W. Curriculum loss: Robust learning and generalization against label corruption[J/OL]. 2019, arXiv: 1905.10045.

[9] Feng L, Shu S L, Lin Z Y, et al. Can cross entropy loss be robust to label noise?[C]. Proceedings of the 29th International Joint Conference on Artificial Intelligence. Yokohama, 2021: 2206-2212.

[10] Wu P X, Zheng S Z, Goswami M, et al. A topological filter for learning with label noise[C]. Proceedings of the 34th International Conference on Neural Information Processing Systems. Vancouver, 2020: 21382-21393.

[11] Yan J X, Luo L, Deng C, et al. Adaptive hierarchical similarity metric learning with noisy labels[J]. IEEE Transactions on Image Processing: A Publication of the IEEE Signal Processing Society, 2023, 32: 1245-1256.

[12] Ching T, Himmelstein D S, Beaulieu-Jones B K, et al. Opportunities and obstacles for deep learning in biology and medicine[J]. Journal of the Royal Society, Interface, 2018, 15(141): 20170387.

[13] 姚家琪, 宋鹏宇, 沈萌, 等. 面向少样本故障诊断的知识自监督深度表征学习方法[J]. 控制与决策, DOI: 10.13195/j.kzyjc.2023.0334.
(Yao J Q, Song P Y, Shen M, et al. Knowledge-aided self-supervised deep representation learning method for few-shot fault diagnosis[J]. Control and Decision, DOI: 10.13195/j.kzyjc.2023.0334.)

[14] Ho C H, Vasconcelos N. Contrastive learning with adversarial examples[C]. Proceedings of the 34th International Conference on Neural Information Processing Systems. Vancouver, 2020: 17081-17093.

[15] Eldele E, Ragab M, Chen Z H, et al. Self-supervised contrastive representation learning for semi-supervised time-series classification[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023, 45(12): 15604-15618.

[16] McMahan H B, Moore E, Ramage D, et al. Communication-efficient learning of deep networks from decentralized data[J/OL]. 2016, arXiv: 1602.05629.

[17] Chen T, Kornblith S, Norouzi M, et al. A simple framework for contrastive learning of visual representations[C]. Proceedings of the 37th International Conference on Machine Learning. New York, 2020: 1597-1607.

[18] van den Oord A, Li Y Z, Vinyals O. Representation learning with contrastive predictive coding[J/OL]. 2018, arXiv: 1807.03748.

[19] Yuan Y, Lin L. Self-supervised pretraining of transformers for satellite image time series classification[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2020, 14: 474-487.

[20] Tan A Z, Yu H, Cui L Z, et al. Towards personalized federated learning[J]. IEEE Transactions on Neural Networks and Learning Systems, 2023, 34(12): 9587-9603.

[21] Devlin J, Chang M W, Lee K, et al. BERT: Pre-training of deep bidirectional transformers for language understanding[J/OL]. 2018, arXiv: 1810.04805.

[22] Anguita D, Ghio A, Oneto L, et al. A public domain dataset for human activity recognition using smartphones[C]. The European Symposium on Artificial Neural Networks. Bruges, 2013: 437-442.

[23] Goldberger A L, Amaral L A N, Glass L, et al. PhysioBank, PhysioToolkit, and PhysioNet[J]. Circulation, 2000, 101(23): e215-e220.

[24] Lessmeier C, Kimotho J K, Zimmer D, et al. Condition monitoring of bearing damage in electromechanical drive systems by using motor current signals of electric motors: A benchmark data set for data-driven classification[J]. PHM Society European Conference, 2016, 3(1): 47-49.

[25] Tan A Z, Yu H, Cui L Z, et al. Towards personalized federated learning[J]. IEEE Transactions on Neural Networks and Learning Systems, 2023, 34(12): 9587-9603.

[26] Wang Y S, Ma X J, Chen Z Y, et al. Symmetric cross entropy for robust learning with noisy labels[C]. IEEE/CVF International Conference on Computer Vision. Seoul, 2019: 322-330.

[27] Jiang X F, Sun S, Wang Y W, et al. Towards federated learning against noisy labels via local self-regularization[C]. Proceedings of the 31st ACM International Conference on Information & Knowledge Management. Atlanta, 2022: 862-873.

[28] Xu J Y, Chen Z H, Quek T Q S, et al. FedCorr: Multi-stage federated learning for label noise correction[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, 2022: 10174-10183.

作者简介

林子谦 (1994–), 男, 博士生, 主要研究方向为智慧医疗、边缘计算、联邦学习, E-mail: linziqian001@163.com;

张坤 (1996–), 男, 博士生, 主要研究方向为统计学习、优化理论, E-mail: Kun.zhang.sufe@outlook.com;

樊重俊 (1963–), 男, 教授, 博士生导师, 主要研究方向为智慧医疗、智慧机场、电子商务, E-mail: fan_chj@163.com;

杨夏洁 (2000–), 女, 硕士生, 主要研究方向为智慧机场、联邦学习, E-mail: a15727160779@163.com.