doi:10.3969/j.issn.1672-1292.2023.01.003

基于联邦知识蒸馏的多站点脑疾病诊断方法

杨启鸣[1],朱　旗[1],王明明[1],孙　凯[2],朱　敏[3],邵　伟[1],张道强[1]

(1.南京航空航天大学计算机科学与技术学院,江苏 南京 211106)
(2.深圳市华赛睿飞智能科技有限公司,广东 深圳 518063)
(3.南京航空航天大学公共实验教学部,江苏 南京 211106)

[**摘要**]　多中心疾病诊断方法通过整合不同医疗机构的样本信息到一台服务器上,集中训练来提高预测的准确性,有效解决了医疗领域小样本的问题. 但仍存在两个问题:不同医疗机构的数据分布不同以及无法保护病人的隐私. 基于此,设计了一种应用在多站点脑疾病诊断领域中隐私保护的联邦知识蒸馏算法. 首先,设计了服务器端基于批标准化的加权平均算法,帮助联邦模型提取各个医疗机构数据分布无关的特征. 之后,在客户端设计了联邦教师模型-本地学生模型的框架,部署了本地分类器,利用蒸馏损失保证模型提取本地化特征,利用分类损失保证模型性能稳定. 实验结果表明,该算法在自闭症及精神分裂症数据集上均优于现有的其他算法.

[**关键词**]　联邦学习,知识蒸馏,脑疾病诊断
[**中图分类号**]TP391　[**文献标志码**]A　[**文章编号**]1672-1292(2023)01-0018-07

Multi-Site Brain Disease Diagnosis Method Based on Federal Knowledge Distillation

Yang Qiming[1],Zhu Qi[1],Wang Mingming[1],Sun Kai[2],Zhu Min[3],Shao Wei[1],Zhang Daoqiang[1]

(1.College of Computer Science and Technology,Nanjing University of Aeronautics and Astronautics,Nanjing 211106,China)
(2.Shenzhen Huasai Ruifei Intelligent Technology Co.,Ltd.,Shenzhen 518063,China)
(3.Public Experimental Teaching Department,Nanjing University of Aeronautics and Astronautics,Nanjing 211106,China)

Abstract:The multi-site disease diagnosis method can improve the accuracy of prediction by integrating the sample information of different medical institutions into one server,which effectively solves the problem of small sample size in the medical field. However,most of these approaches have two problems in the medical field which being the different distribution of data in different medical institutions and the inability to protect patient privacy. Based on these,we design a federal knowledge distillation algorithm for privacy protection in multi-site brain disease diagnosis. Firstly,a weighted average algorithm based on batch standardization is designed on the server to help the federated model to extract the distribution independent feature of each medical institution. Then,the framework of federated teacher model-local student model is designed on the client,and local classifier is deployed. The distillation loss guarantee model is used to extract localized features,and the classification loss is used to ensure the stable performance of the model. Experimental results show that the proposed algorithm is superior to other existing algorithms in autism and schizophrenia datasets.

Key words:federated learning,knowledge distillation,brain disease diagnosis

多中心疾病诊断已成功地应用于医学图像的病理检测和诊断[1],其整合了多个医疗机构的样本信息并利用这些样本得到鲁棒性更强的模型. 尽管这类方法取得了很好的效果,但需要不同站点的医疗机构把自己的数据集公开给模型训练方,大大增加了数据泄露的风险. 此外,若不同的医疗机构所拥有的数据分布差异很大,仅仅集合这些数据训练模型也无法拥有很好的泛化能力.

随着大数据技术和移动互联网的蓬勃发展,数据安全问题已成为困扰全世界的棘手难题,其不仅事关个人生活及企业商业隐私安全,更威胁到了国家信息安全. 近年来,世界各国及国际组织相继出台了保护数据安全的相关法律法规,种种法律法规也导致了各客户端上的数据不能像以往一样自由“流通”,即出

收稿日期:2022-09-15.
通讯作者:朱旗,博士,副教授,研究方向:机器学习、模式识别、脑疾病诊断. E-mail:zhuqinuaa@163.com

现数据“孤岛”现象. 在医疗领域,由于医疗数据的隐私性较强,数据“孤岛”问题尤为突出. 现代医学研究工作需依赖多样本、多维度的大数据作为支撑,若不能有效解决数据“孤岛”问题,研究工作将很难开展.

在此背景下,谷歌的 McMahan 等[2]提出了一种名为联邦学习[3-5]的分布式机器学习方法,其核心思想是多个拥有数据集的客户端在只交换模型参数而不交换数据的前提下共同训练从而得到所需要的模型. 联邦学习一经提出就引发了广泛的关注,众多研究机构和企业都在不断提出新方法以提高联邦学习的安全性和准确性. 联邦学习在医疗领域也有巨大的应用前景,不同医疗机构之间通过联邦学习可在不泄露病人隐私的前提下扩大样本数量和特征维度,助力各机构在数据安全的前提下,充分实现数据共享,更好地完成医学研究工作.

尽管联邦学习保护了分布在各个医疗机构的数据的安全,但不同医疗机构之间普遍存在着数据的非独立同分布(not identically and independently distributed,Non-IID)问题,严重制约了联邦学习的模型精度. 知识蒸馏(knowledge distillation)由 Hinton 等[6]提出,是模型压缩的常用方法. 其核心思想是利用性能较好的大模型的监督信息来训练小模型,以达到模型压缩和知识迁移的目的. 这恰好可运用于解决联邦学习所面临的 Non-IID 问题. 在具体的应用中,Vielzeuf 等[7]通过知识蒸馏将多模态网络特征迁移到单个模型中. Wang 等[8]设计了一个私有模型压缩框架 RONA,在数据不外泄的条件下完成了大模型到小模型的特征迁移. Vongkulbhisal 等[9]将一组无法共享数据且具有不同体系结构和目标的分类器通过知识蒸馏训练为单个分类器.

基于此,本文设计了一种针对医学图像分类领域的联邦知识蒸馏算法(federated knowledge distillation algorithm,FKDA). 首先,服务器端在联邦平均算法(federated average,FedAvg[2])的基础上构建基于批标准化层(batch normalization layer,BN 层)的加权平均算法,通过提取本地模型的 BN 层数据,缓解不同站点之间数据分布不均的问题;其次,在客户端使用教师-学生模型,将联邦模型作为教师模型,本地模型作为学生模型,通过蒸馏联邦模型知识保证本地参与方性能的稳定. 本文在真实的自闭症数据集和精神分裂症数据集上进行实验验证,结果证明了该模型在医学图像分类中的可行性和有效性.

1 相关工作

联邦学习是一种分布式机器学习技术,应用于医学研究领域可避免数据隐私泄露问题. 联邦学习不要求各医疗机构将数据直接共享到一个集中的数据存储平台中以构建机器学习模型,而是在各孤立的数据站点上进行模型的训练,在保持数据本地化的同时通过模型参数的传递训练全局模型.

Sheller 等[10]首次将联邦学习应用于多机构图像语义分割研究,无需共享患者数据即可实现深度学习建模. 为了提高通信效率和模型性能,Zhang 等[11]提出了一种基于动态融合的联邦学习方法,应用于医学诊断图像分析以检测 COVID-19 病毒. 这些算法成功地将联邦学习应用到医学领域中,取得了非常好的效果.

但基于同分布数据的联邦学习并不能很好地解决 Non-IID 问题. 研究者们提出了很多解决方案,可大致分为基于数据、基于模型、基于算法和基于框架 4 种思想[12]. 在基于框架的思想下,知识蒸馏技术得到了较好地应用.

知识蒸馏是压缩模型的一种常用方法,利用性能更好的大模型的监督信息来训练小模型,以期达到更好的性能和精度. 将大规模数据训练的模型应用于本地模型,可显著提升本地模型的训练效果. Jeong 等[13]发现知识蒸馏也可用于解决联邦学习中的 Non-IID 问题,于是提出了联邦蒸馏算法,以期减少模型的通信开销,并降低 Non-IID 问题对模型造成的负面影响. Jiang 等[14]提出了一种基于知识蒸馏的分布式联邦训练方法,每个客户端都引入一个个性化模型来适应本地数据以提高局部性能,通过知识蒸馏技术,即使在全局模型难以适应局部数据集的情况下,也能提高全局模型的性能和稳定性. Cha 等[15]提出了一种分布式强化学习框架,即联邦强化蒸馏(FRD),将强化学习和知识蒸馏的思想同时应用到联邦学习中. Itahara 等[16]提出了一种基于蒸馏的半监督联邦学习算法(DS-FL),在各客户端之间交换本地模型输出而非模型参数. 上述研究成果表明,基于知识蒸馏的联邦学习算法可有效减少 Non-IID 问题的负面影响,并提高客户端通讯效率.

2 联邦知识蒸馏算法

假设有 N 个站点 $\{C_1, C_2, \cdots, C_N\}$ 参加联邦学习,每个站点作为联邦中的一个客户端,都有自己的数据

集$\{D_1,D_2,\cdots,D_N\}$. 用D_i表示每个站点的数据集,令$D_i=\{(x_i,y_i)\}_{i=1}^{n_i}$,其中,$x_i$表示数据集中第$i$个样本,$y_i$表示第$i$个样本的标签,$n_i$表示$D_i$中的样本数量. 同时,每个站点的数据集都分成训练集和测试集. 为每个站点都训练一个模型$\{f_i\}_{i=1}^{N}$,总体的损失函数可表示为:

$$F_{\text{Server}}=\sum_{i=1}^{N}\alpha_i f_i, \tag{1}$$

式中,F_{Server}表示经加权平均后得到的联邦模型;α_i表示每个本地模型参与联邦平均的权重.

本文所设计的 FKDA 算法结构图如图 1 所示. 其核心思想是将客户端的任务分解成特征提取和分类两个模块,利用联邦学习,将本地客户端的特征提取模块在服务器端进行加权平均,得到一个能够提取站点无关特征及鲁棒性更强的特征提取器;同时,在客户端使用基于特征的知识蒸馏损失及针对具体任务的分类损失,保证应用在客户端模型的稳定性和准确性.

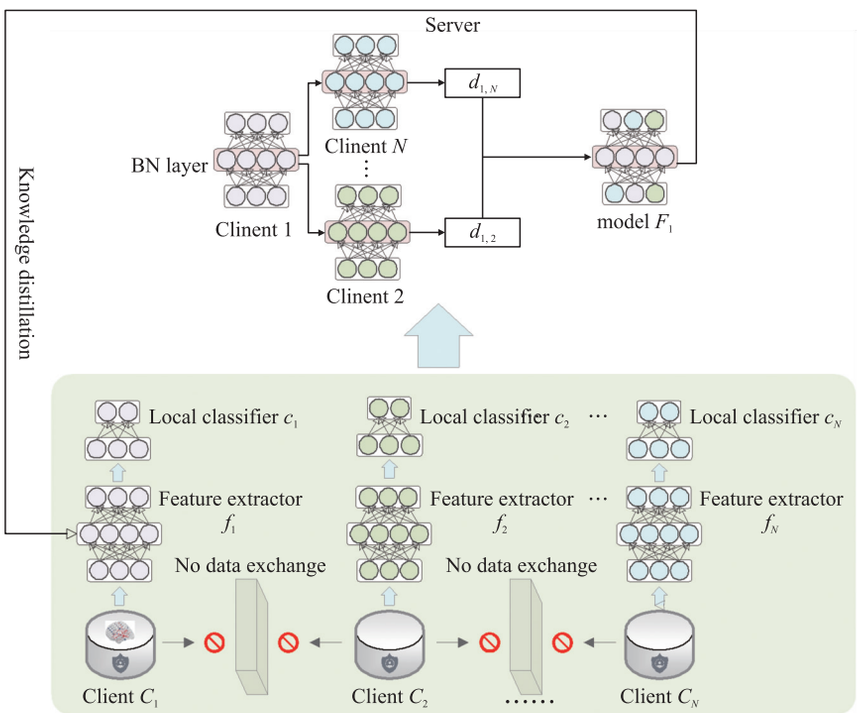

图 1　FKDA 结构图

Fig. 1　The diagram of the FKDA method

2.1　基于批标准化的联邦加权算法

联邦学习的核心问题就是求得式(1)中的权重α_i. 早期的 FedAvg 算法已证明,把所有的本地客户端模型f_i的参数求和取平均值,即$\alpha_i=1/N$,已在很多领域取得了很好的结果. 但医院图像数据大多是 Non-IID 的,FedAvg 算法在这些异构医学数据上表现不佳. 为解决这个问题,本文在联邦模型中添加了 BN 层,BN 层首先计算上一层特征的均值和方差:

$$\mu=\frac{1}{m}\sum_{i=1}^{m}x_i, \tag{2}$$

$$\sigma^2=\frac{1}{m}\sum_{i=1}^{m}(x_i-\mu), \tag{3}$$

式中,m为神经网络训练的批次大小;μ为均值;σ^2为方差. 通过均值和方差对特征x_i归一化:

$$\hat{x}_i=\frac{x_i-\mu}{\sigma}. \tag{4}$$

得到归一化参数后,引入可学习重构参数γ和β对归一化特征$\hat{x}_i$重构:

$$\text{BN}_{\gamma,\beta}(x_i)=\gamma\hat{x}_i+\beta. \tag{5}$$

批标准化不仅可让神经网络训练更快、更稳定,还可使学到的特征泛化性更强. 在本文中,来自不同

客户端模型中的 BN 层是通过不同分布的站点特征训练得到,更具站点代表性. 而后,计算各客户端中不同 BN 层参数之间的欧几里得距离:

$$d_{i,j}=|\gamma_i,\gamma_j|+|\beta_i,\beta_j|, \tag{6}$$

式中,$d_{i,j}$为客户端 i 和客户端 j 之间的距离;$|\cdot,\cdot|$为欧几里得距离. 距离越近,认为两个客户端的分布一致性越高. 针对每个客户端 i,都可得到一组和其他客户端的距离:

$$d_i=\{d_{i,1},d_{i,2},\cdots,d_{i,N}\}. \tag{7}$$

本文基于模型之间的距离,得到每个客户端的定制化模型. 当客户端 i 和其余客户端差异较大时,给予客户端 i 一个 0 到 1 的可调节权重 λ,其余客户端 j 的权重为:

$$\alpha_j=\frac{d_{i,j}}{\mathrm{sum}(d_i)}(1-\lambda). \tag{8}$$

最终,考虑到模型的 BN 层中带有各客户端的本地信息,因此不参与模型的加权平均过程:

$$F_i=\lambda f_i+\sum_{j=1,j\neq i}^{N}\alpha_j f_j. \tag{9}$$

2.2 基于特征的本地客户端知识蒸馏

经服务器聚合后的联邦模型能够更好地提取站点无关特征,但本地模型和联邦模型之间仍存在着一定偏移. 为了消除这个偏移,本文使用基于特征的知识蒸馏来优化本地客户端模型:将来自服务器的定制化联邦模型视为教师模型,经教师模型提取特征 f^{T};将驻留本地客户端的本地模型视为学生模型,经学生模型提取特征 f^{S}.

得到这两种特征后,通过优化 f^{T} 和 f^{S} 之间的最大均值差异(maximum mean discrepancy,MMD)损失使本地模型学习到不失本地特性并更具一般性的特征. 最大均值差异是迁移学习中使用最广泛的一种损失函数,其主要作用是用来度量两个不同但相关的随机变量的分布距离,其表达式为:

$$\mathrm{MMD}=\left\|\frac{1}{n_i}\sum_{j=1}^{n_i}\phi(f_j^{\mathrm{T}})-\phi(f_j^{\mathrm{S}})\right\|_H, \tag{10}$$

式中,$\phi(\cdot)$为映射函数. 本文的蒸馏损失为:

$$\mathrm{KD}=\mathrm{MMD}(f^{\mathrm{T}}\|f^{\mathrm{S}}). \tag{11}$$

为了保证模型的稳定性,本文使用交叉熵损失来约束分类器,最终的损失函数为:

$$L_i=\mathrm{KD}_i+l_c(c(f(D_i^{\mathrm{train}}))), \tag{12}$$

式中,l_c 为交叉熵损失;$c(\cdot)$为客户端 i 的分类器;D_i^{train} 为客户端 i 的训练集.

2.3 算法流程

首先,服务器进行初始化设定模型参数、学习率、迭代次数等,并将其分发给各客户端. 在第一次训练时,客户端先训练本地模型,并将模型发送给服务器,暂不进行知识蒸馏. 服务器得到每一个参与方的模型后,通过基于批标准化的加权平均算法,为每一个客户端计算得到一个联邦模型,并将模型下发给客户端,客户端收到联邦模型后通过知识蒸馏算法更新本地模型参数. 重复该过程迭代训练模型,直至模型收敛,保存模型并结束本次训练. 具体算法流程如下:

算法 1　联邦知识蒸馏

输入:全局权重,本地数据$\{D_1,D_2,\cdots,D_N\}$,学习率,迭代次数
1:初始化原模型参数 F_0,将原始模型参数广播给所有参与方
在客户端执行(仅在开始训练时执行)
2:For 所有参与方 C 并行 do
3:本地更新模型参数:$f_i,c_i\leftarrow D_i$
4:更新后的模型参数 f_i 发送给服务器,分类器 c_i 驻留本地
5:End for
在服务器端执行
6:For 每一个参与方 C do
7:通过式(7)得到 d_i

8:通过式(8):$\alpha_j \leftarrow d_i$
9:将式(9)聚合得到的 F_i 分发给各个客户端
10:End for
在客户端执行
11:For 所有参与方 C 并行 do
12:蒸馏损失:KD$\leftarrow f_i, F_i, D_i$
13:分类损失:$l_c \leftarrow f_i, D_i$
14:　更新模型参数:$f_i, c_i \leftarrow$KD, l_c
15:更新后模型参数 f_i 发送给服务器,分类器 c_i 驻留本地
16:End for

3　实验

3.1　实验材料

为验证所提出的方法,本文在真实的自闭症大脑成像(autism brain imaging data exchange,ABIDE)[17]数据集以及来自 5 家医院的精神分裂症数据集上进行了实验. ABIDE 数据集是一个合格的多站点异构数据集,包括来自 17 个不同地点的 1 112 名受试者的静息态功能磁共振成像(resting-state functional magnetic resonance imaging,rs-fMRI)和临床数据. 考虑到一些站点的样本量较小,本文选择了其中 5 个样本量超过 50 名受试者的站点参与实验:Leuven,USM,UCLA,UM 和 NYU. 这些站点一共包含 468 名受试者,包括 218 名自闭症患者(autism spectrum disorder,ASD)和 250 名年龄相匹配的对照组. 5 个精神分裂症数据集分别使用的是南京脑科医院(NBH)数据集、生物医学卓越研究中心(COBRE)数据集、诺丁汉(Nottingham)数据集、复旦(Fudan)数据集和湘雅(Xiangya)数据集. 这些数据集的受试者均满足以下要求:(1)无其他精神类疾病;(2)无吸毒史;(3)无临床明显的头部创伤. 表 1 展示了受试者的人口统计学特征.

表 1　受试者的人口统计学特征

Table 1　Demographic characteristics of the participants

站点	患者		对照组		站点	患者		对照组	
	男	女	男	女		男	女	男	女
Leuven	21	4	24	8	NBH	6	15	10	14
USM	66	5	79	14	COBRE	42	11	46	21
UCLA	28	8	31	7	Nottingham	27	5	95	85
UM	43	5	56	9	Fudan	35	34	25	37
NYU	30	8	21	1	Xiangya	49	34	35	25

3.2　实验设置及对比方法

本文的实验环境为 PyTorch = 1.10.1,设置随机数种子 seed = 0,使用随机梯度下降(stochastic gradient descent,SGD)算法进行迭代优化,初始学习率 learning rate = 0.000 5,实验设备为英伟达 RTX 3070. 实验将本文所提出的方法与 FedAvg[2]、FedBN[18]、FedProx[19] 3 种联邦学习方法进行了比较.

3.3　实验结果

首先,将本文所提出的联邦知识蒸馏算法在自闭症数据集上进行消融实验,以验证每个模块的具体作用. 其后,将所得出的实验结果在两种疾病的数据集上分别与联邦学习方法进行对比实验,以验证所提出算法的性能. 所有实验均采用十折交叉方法进行验证.

3.3.1　消融实验

本文通过在 ABIDE 数据集上进行消融实验展示了所提方法每个模块的效果. 通过单独提出批标准化联邦加权算法模块及基于特征的知识蒸馏模块,分别对各模块的作用进行了验证. 在不添加任何模块时,使用 FedAvg 算法进行比较. 表 2 展示了本文消融实验的结果. 从准确率来看,无论哪个模块添加进联邦平均算法中,均提升了模型的准确性. 其中,在站点 Leuven、NYU 及 UM 中,两个模块对准确率的影响较为接近;而在站点 UCLA 和 USM 中,基于特征的知识蒸馏模块相较于批标准化联邦加权算法模块在提升模型准确率上起着更大的作用. 同时,由于采用的是十折交叉验证算法,实验结果中的标准差可以反映出数

据集每一折准确率的波动大小. 因此,本文通过标准差的大小来评估模型的稳定性,即拥有较小标准差的模型稳定性更强. 可以看出,基于特征的知识蒸馏对模型的稳定性帮助更大,这也验证了前文的猜想. 综合起来看,本文方法可达到最大的性能提升.

表 2　在 ABIDE 数据集上的消融实验

Table 2　Ablation of the proposed method on ABIDE dataset

批标准化联邦加权算法	基于特征的知识蒸馏	准确率				
		Leuven	NYU	UCLA	UM	USM
×	×	57.67±15.57	66.99±8.95	61.07±21.23	57.52±13.22	66.67±23.33
√	×	61.67±13.67	71.21±8.93	66.07±13.97	63.71±12.93	73.33±20.00
×	√	61.00±11.86	71.21±7.56	76.50±10.72	63.56±12.43	76.67±17.00
√	√	62.67±13.23	71.84±6.11	75.54±10.72	65.38±8.93	78.33±15.00

3.3.2　对比实验

FedAvg 算法是最经典的联邦学习算法,但不针对异构数据进行特殊处理. 相对的,FedBN 算法及 FedProx 算法分别从服务器端聚合策略及客户端训练两个角度来解决 Non-IID 问题. ABIDE 数据集上的实验结果如表 3 所示. 与前文猜想一致,FedAvg 算法在数据异构的场景下表现不佳,所有结果均低于其余针对解决数据 Non-IID 问题的联邦学习算法. FedBN 算法在 FedAvg 算法的基础上通过保留神经网络中的批标准化层在本地更新的策略来处理数据异构问题,从实验结果可以看出,FedBN 在准确率上相较于 FedAvg 有明显的提升,但在模型稳定性上所起作用不大. 在站点 Leuven 和 NYU 上,FedBN 算法的模型稳定性较 FedAvg 均有所下降. 相对的,尽管 FedProx 算法在模型准确性上只有在站点 Leuven 上略高于 FedBN,但其模型的稳定性较为突出.

表 4 所示为精神分裂症数据集上的结果. 可以看出,本文方法在保证最高准确率的同时,标准差也最小,且在 COBRE 站点及 Nottingham 站点中,模型的稳定性相对于其他方法有更明显的提升.

本文算法通过批标准化联邦加权算法缓解多站点数据之间的 Non-IID 问题,在客户端构建教师-学生框架加强模型稳定性. 从实验结果可以看出,本文算法在模型准确性和稳定性上均有明显提升.

表 3　4 种不同的疾病诊断方法在 ABIDE 自闭症数据集上的性能

Table 3　Performance of four different methods in ASD identification on the multi-site ABIDE dataset

方法	准确率				
	Leuven	NYU	UCLA	UM	USM
FedAvg	57.67±15.57	66.99±8.95	61.07±21.23	57.52±13.22	66.67±23.33
FedBN	59.33±15.76	68.24±9.82	72.86±16.71	64.77±12.81	73.33±19.44
FedProx	61.00±13.91	67.61±6.42	70.18±10.13	62.95±11.78	70.00±17.95
FKDA	62.67±13.23	71.84±6.11	75.54±10.72	65.38±8.93	78.33±15.00

表 4　3 种不同的疾病诊断方法在精神分裂症数据集上的性能

Table 4　Performance of three different methods in schizophrenia identification on the multi-site dataset

方法	准确率				
	NBH	COBRE	Nottingham	Fudan	Xiangya
FedAvg	76.50±22.47	73.33±14.43	63.21±19.50	79.45±9.56	66.52±10.91
FedBN	77.78±20.30	70.83±12.05	65.09±20.52	77.86±9.81	67.33±9.65
FKDA	80.50±20.01	75.83±10.17	67.86±14.45	80.16±8.03	71.43±9.12

4　结论

本文基于在多中心脑疾病诊断中的两个问题:不同医疗机构的数据分布不同及无法保护病人的隐私,设计了一种联邦知识蒸馏算法 FKDA,通过将联邦学习算法应用到医学图像分类任务上,解决了传统数据中心化问题带来的隐私泄露问题;通过重新设计联邦学习在服务器端的聚合策略,缓解了不同医疗机构数据分布不同所带来的 Non-IID 问题,帮助模型学习更鲁棒的特征;同时,在客户端部署了教师-学生模型,通过结合基于特征的知识蒸馏损失和分类损失,保证模型的稳定性和准确性. 实验结果表明,本文的框架

在医学图像分类任务上比其他联邦模型的准确率更高,性能表现也更稳定.

[参考文献](References)

[1] 周海榆,张道强. 面向多中心数据的超图卷积神经网络及应用[J]. 计算机科学,2022,49(3):129-133.

[2] MCMAHAN H B,MOORE E,RAMAGE D,et al. Communication-efficient learning of deep networks from decentralized data[C]// Proceedings of the 20th International Conference on Artificial Intelligence and Statistics(AISTATS). Fort Lauderdale,USA: JMLR,2017.

[3] YANG Q,LIU Y,CHEN T,et al. Federated machine learning:concept and applications[J]. ACM Transactions on Intelligent Systems and Technology,2019,10(2):1-19.

[4] YANG Q,LIU Y,CHENG Y,et al. Federated Learning[M]. San Rafael,USA:Morgan & Claypool Publishers,2019.

[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] HINTON G,VINYALS O,DEAN J. Distilling the knowledge in a neural network[J]. Computer Science,2015,14(7):38-39.

[7] VIELZEUF V,LECHERVY A,PATEUX S,et al. Towards a general model of knowledge for facial analysis by multi-source transfer learning[J]. arXiv Preprint arXiv:1911.03222,2019.

[8] WANG J,BAO W D,SUN L C,et al. Private model compression via knowledge distillation[J]. Proceedings of the AAAI Conference on Artificial Intelligence,2019,33(1):1190-1197.

[9] VONGKULBHISAL J,VINAYAVEKHIN P,VISENTINI-SCARZANELLA M. Unifying heterogeneous classifiers with distillation [C]//Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR). Long Beach, USA:IEEE,2019.

[10] SHELLER M J,REINA G A,EDWARDS B,et al. Multi-institutional deep learning modeling without sharing patient data:a feasibility study on brain tumor segmentation[C]//Proceedings of the 4th International MICCAI Brainlesion Workshop. Granada,Spain:Springer,2018.

[11] ZHANG W S,ZHOU T,LU Q H,et al. Dynamic fusion-based federated learning for COVID-19 detection[J]. IEEE Internet of Things Journal,2021,8(21):15884-15891.

[12] MA X,ZHU J,LIN Z,et al. A state-of-the-art survey on solving Non-IID data in federated learning[J]. Future Generation Computer Systems,2022,135:244-258.

[13] JEONG E,OH S,KIM H,et al. Communication-efficient on-device machine learning:federated distillation and augmentation under Non-IID private data[J]. arXiv Preprint arXiv:1811.11479,2018.

[14] JIANG D L,SHAN C,ZHANG Z H. Federated learning algorithm based on knowledge distillation[C]//Proceedings of the 2020 International Conference on Artificial Intelligence and Computer Engineering(ICAICE). Beijing,China:IEEE,2020.

[15] CHA H,PARK J,KIM H,et al. Proxy experience replay:federated distillation for distributed reinforcement learning[J]. IEEE Intelligent Systems,2020,35(4):94-101.

[16] ITAHARA S,NISHIO T,KODA Y,et al. Distillation-based semi-supervised federated learning for communication-efficient collaborative training with Non-IID private data[J]. arXiv Preprint arXiv:2008.06180,2020.

[17] MARTINO A D,YAN C G,LI Q,et al. The autism brain imaging data exchange:towards a large-scale evaluation of the intrinsic brain architecture in autism[J]. Molecular Psychiatry,2014,19(6):659-667.

[18] LI X X,JIANG M R,ZHANG X F,et al. FedBN:federated learning on Non-IID features via local batch normalization[J]. arXiv Preprint arXiv:2102.07623,2021.

[19] LI T,SAHU A K,ZAHEER M,et al. Federated optimization in heterogeneous networks[J]. avXiv Preprint arXiv:1812.06127,2020.

[责任编辑:严海琳]