



[专家简介] 何昆仑,解放军总医院医学创新研究部某中心主任,主任医师、教授,主要从事医疗大数据、医学人工智能、智能卫勤装备和系统等研究。担任智能医学工程国家重点实验室主任,承担重大项目等27项,制定国家行业标准29项,授权发明专利103项,获国家科技进步二等奖1项,省部级一、二等奖5项。担任中国研究型医院学会副会长、北京医师协会副会长等。

医学数据个性化联邦学习的研究进展

柳青河¹, 姜明明¹, 赵亚威¹, 何昆仑^{1,2}

¹解放军总医院医学创新研究部, 北京 100853; ²医疗大数据应用技术国家工程研究中心, 解放军总医院医学工程实验室, 国家药监局人工智能医疗器械研究与评价重点实验室, 北京 100853

摘要:联邦学习是沟通医学数据和机器学习的桥梁,能以保护隐私的方式对数据进行训练。个性化联邦学习在其结构基础上有着更加优秀的性能,尤其是在不同的用户之间数据分布差异巨大的时候。个性化联邦学习既实现了训练时数据隐私保护,又取得了更加优越的性能,在多中心场景智能诊断和辅助决策等任务中具有广阔的前景。本文就个性化联邦学习在医学领域的发展进行深入讨论。

关键词:医学数据; 个性化联邦学习; 机器学习; 多中心数据; 隐私保护

中图分类号:R197.323; TP18; TP391.41

文献标志码:A

文章编号:2095-5227(2025)01-0113-06

DOI: 10.12435/j.issn.2095-5227.24070103

引用本文: 柳青河, 姜明明, 赵亚威, 等. 医学数据个性化联邦学习的研究进展 [J]. 解放军医学院学报, 2025, 46(1): 113-118.

Research advances in personalized federated learning of medical data

LIU Qinghe¹, JIANG Mingming¹, ZHAO Yawei¹, HE Kunlun^{1,2}

¹Medical Innovation Research Department of PLA General Hospital, Beijing 100853, China; ²National Engineering Research Center for Medical Big Data Application Technology, Medical Engineering Laboratory of PLA General Hospital, Key Laboratory of Artificial Intelligence Medical Device Research and Evaluation of the National Medical Products Administration, Beijing 100853, China

Corresponding author: HE Kunlun. Email: kunlunhe@301hospital.com.cn

Abstract: Federated learning is a bridge between medical data and machine learning, and can train data in a way that protects privacy. Personalized federated learning has better performance on the basis of its structure, especially when there are great differences in data distribution among different users. It not only realizes data privacy protection during training, but also achieves more superior performance, which has broad prospects in tasks such as intelligent diagnosis and auxiliary decision-making of multi center scenes. The development of personalized federated learning in medicine is discussed in this paper.

Keywords: medical data; personalized federated learning; machine learning; multi-central data; privacy protection

Cited as: Liu QH, Jiang MM, Zhao YW, et al. Research advances in personalized federated learning of medical data [J]. Acad J Chin PLA Med Sch, 2025, 46(1): 113-118.

机器学习是一种主要依靠机器算力解决问题的科学方法,包括支持向量机、多层感知机和深度学习等,具备高精度、高效率和高泛化性等优

点,目前广泛应用于医学领域的生产和科研实践中^[1-3]。同时,医院日常工作产生的医疗数据具有维度高、模态多、数量大的特点,符合机器学习对数据的要求^[4-5]。以 ChatGPT 为代表的预训练大模型的成功,从实践的角度证明了海量丰富的高质量数据有利于机器学习模型向更高精度和泛化性的方向优化^[6-7]。但数据隐私保护的实施将限制各大医院之间的数据交流,制约机器学习模型的

收稿日期: 2024-07-01

基金项目: 国家自然科学基金项目(62302522); 省部级课题

第一作者: 柳青河, 在读硕士, 医师。Email: Liuqinghe9638@163.com

通信作者: 何昆仑, 博士, 主任医师, 教授, 博士生导师。Email: kunlunhe@301hospital.com.cn

大规模、多种类、高深度的训练^[8-9]。联邦学习作为一种保护隐私的多中心数据学习方式，获得了广泛的关注^[10-13]。联邦学习一般是在不同医院数据训练的局部模型汇聚到中心服务器进行加权平均聚合，得到一个新模型的更新方法(图1)。这种聚合过程并不涉及私人数据的传输，从原理上杜绝了隐私泄露的风险。随着研究的深入，研究人员发现如果不同中心数据的分布有较大差异，可能会对联邦学习的学习性能产生不利影响。这种数据分布差异巨大的情况，被称为非独立同分布。通过学习具有不同种类、模态和质量等因素的数据，不同的局部模型参数可能差异巨大，经由联邦学习聚合得到的新模型并不能有效解决具有不同分布数据产生的问题^[14-15]。如综合医院的诊断经验并不能完全适用于专科医院，最大的原因在于综合医院的疾病分布与专科医院的差异较大。针对非独立同分布数据对联邦学习聚合出的模型性能造成的不利影响，研究人员提出个性化联邦学习的思路^[16-18]。个性化联邦学习的主要思想是通过提取多中心数据的共同信息，融合到局部信息中，更新得到符合本地局部数据分布的个性化模型^[18]。相比传统联邦学习，个性化联邦学习在解决非独立同分布数据的问题时，更具优越性。在现实的

医疗场景中，个性化联邦学习更具有可行性，原因在于不同中心的医疗数据分布会随着设备参数、病例类型、医师水平等因素的差异而变化^[19-20]。如南方医院因患者的生活习惯、医院的设备参数以及医师的治疗手段不同，与北方医院的某些医疗数据分布差异较大^[20-22]。

1 个性化联邦学习的基本概念及主要算法

1.1 基于联邦学习的个性化延伸

联邦学习从特征角度可以分为纵向联邦学习^[23]和横向联邦学习^[24]，其中纵向联邦学习是指所有参与学习的医院所提供的数据维度不完整，只有所有医院数据集合在一起才是完整的数据。与纵向联邦学习不同，参与横向联邦学习的医院拥有患者的完整数据。目前隐私保护多是针对医院之间的数据交流进行限制，故医学领域现实场景较多且研究较深入的是横向联邦学习^[9]。横向联邦学习最为经典的算法是对参数进行加权平均聚合，即FedAvg，具体的实现方式为：

$$\bar{W} = \frac{\sum_{k=1}^K n_k W_k}{\sum_{k=1}^K n_k}$$

其中 W_k 表示第 k 个用户参与联邦学习时，通

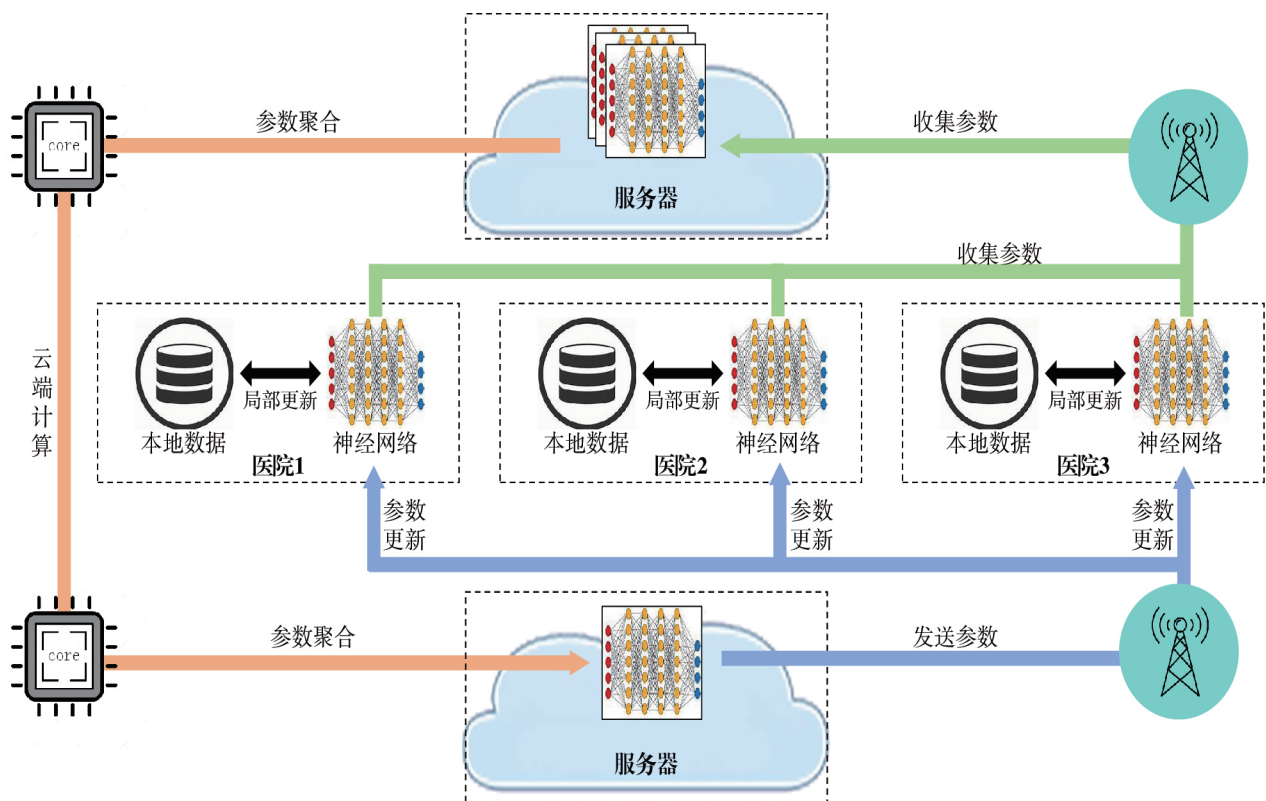


图1 联邦学习过程示意图

Fig. 1 Process of federated learning

过该用户本地数据训练出来的最优模型参数； n_k 表示第 k 个用户的数据量； \bar{W} 表示通过 k 个用户的局部最优模型聚合之后得到的最终模型参数^[25]。这种方式可以有效整合不同中心医院数据的共同信息，但可能会造成各中心医院特有的信息被忽视。因此，当参与联邦学习的中心医院之间的数据差异较大时，单纯地对模型参数进行加权平均聚合，可能忽视它们之间的特有信息。研究人员提出根据数据的分布特点，个性化聚合不同模型的参数，既能整合公共知识，又能根据医院的特点制定个性化聚合模型^[16-18]，实现模型的更高性能，其通用表达式为：

$$W = G_w(W_1^*, \dots, W_n^*, \Theta_1^*, \dots, \Theta_n^*)$$

$$\Theta_k = G_\theta(W_1^*, \dots, W_n^*, \Theta_1^*, \dots, \Theta_n^*; k)$$

其中， W_i^* 和 Θ_i^* 分别是本地模型在第 i 个用户的数据中进行训练得到的最优参数的公共部分和个性化部分； G_w 和 G_θ 分别表示公共部分模型参数的聚合方法和个性化部分模型参数的聚合方法； W 和 Θ_i 分别表示属于第 i 个用户的个性化模型聚合之后的公共部分参数与个性化部分参数，个性化联邦学习过程示意图2。

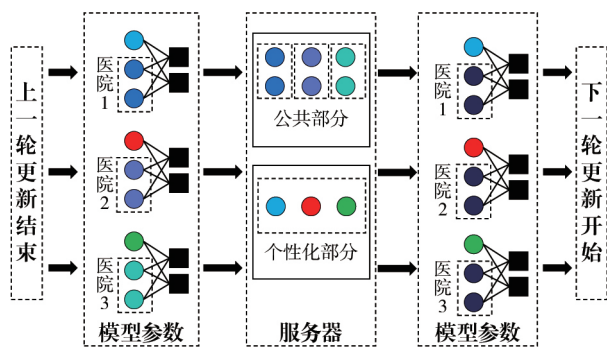


图2 个性化联邦学习过程示意图

Fig. 2 Process of personalized federated learning

通过研究不同的聚合方法 G_w 和 G_θ ，可以改善公共信息和局部信息的提取及聚合能力。目前，研究人员多采用平均、对齐、聚类等方法来改善个性化模型的公共信息提取能力，并且通过保留更新、选择性更新和正则化等方法增强个性化模型的局部信息学习能力。

1.2 个性化联邦学习的主要方法

对于传统联邦学习模型在参与者之间的数据差异较大时表现不佳的问题，Arivazhagan等^[26]针对深度学习模型提出，通过加权平均聚合模型层的表示层参数，保留模型对公共信息的提取能力，并且不聚合分类器层参数，以提高模型对局部信

息的感知能力，这种个性化聚合模型的方式较为简便，但无法调节个性化程度，不能适配数据分布差异不大的情况。Hanzely和Huang等^[27-28]针对可调节的个性化参数提出模型聚合的正则化方法L2GD和FedAMP，其主要的表达式为：

$$\min_W \{G(W): = F(W) + \lambda A(W)\}$$

其中， λ 为正则化项的超参数，用于调节模型参数更新的幅度。上述两方法分别选择不同的函数限制个性化模型之间的参数差异，即允许个性化模型之间根据本地数据差异，发生小范围的波动。在尽可能多地保留公共信息的同时，保留个性化模型对局部信息的感知能力。正则化方法相对简单，同时具有调节能力，但其仅关注个性化部分，削弱了对公共信息的提取能力，并且两者计算复杂度较大，限制了联邦学习的拓展性。Dinh等^[29]将全局信息植入到正则化项中提出pFedMe，使局部模型在本地数据中更新时尽可能不偏离全局信息，其表达式为：

$$F_i(w) = \min_{\theta_i \in R^d} \{f_i(\theta_i) + \frac{\lambda}{2} \|\theta_i - w\|^2\}$$

其中， θ_i 是带有本地信息的个性化模型参数，而 w 是带有全局信息的参数。与L2GD和FedAMP相比，此种方法对算力资源消耗较小，拓展性较强，但该方法将所有参与联邦学习的数据信息纳入正则化选项，个性化模型存在被“误导”的可能，稳定性不够。Wang等^[30]提出rFedAvg，通过不同客户之间的样本在希尔伯特空间映射后的正则化对齐，实现聚合模型更高精度和泛化性。但这种方式可能违反联邦学习的数据保护的原则。不同于正则化的方法，Ghosh等^[31]基于聚类的方式提出IFCA，即对参与联邦学习的每个用户在本数据下的 k 个公共模型中获得的最小损失进行聚类，每一轮更新均以所得最小损失模型参数为起点进行更新，获得 k 个个性化模型。该方式有利于个性化模型学习全局信息，并且局部信息也能在一定范围内保留，但其对资源的消耗较大，非常占用网络传输资源，局部模型的反复更新也会较多地消耗本地算力，不利于实际部署。Lu等^[32]提出OCFL，通过计算不同客户端之间本地模型的余弦距离进行聚类，而后将聚类中心的模型参数发送至对应的客户端，实现高适用性和泛化性的个性化聚合。该方法最大的问题在于聚类中心之间没有交流，部分聚类中心可能会与实际分布出入较大。Morafah等^[33]则是提出FLIS，通过每一客户返

回的本轮局部模型，计算位于服务器的共享数据集结果，而后通过这些结果对本轮局部模型进行聚类，在同一类别下进行加权平均聚合。该类聚合方法克服了参与联邦学习人数较少时聚类偏倚较大的问题，但极大地提高了服务器的计算成本，同时位于服务器的数据集质量直接决定聚类聚合的效果。Long等^[34]则提出FeSEM，即在服务器端直接对模型进行聚类，而后在多轮次聚合过程中，通过发送某一聚类中心的模型，使随机期望最大化，并逐渐收敛，得到最接近该客户局部数据分布的个性化模型。该方法在参与联邦学习的客户较多时具有较大优势，但客户数量较少时难以聚集出有效的聚类中心。Chen和Chao^[35]提出FedRoD，通过定义新的损失函数对齐模型参数，并且在对齐参数的基础上，模型中的判别器进一步学习符合本地数据分布规律的局部信息。该方法在各参与联邦学习的数据类别分布不均时效果较好，并且算力资源消耗较小，利于部署，但是其对各数据的分布规律依赖较强，导致其对占总体样本少数的类别数据忽视较为严重，对于重视模型特异性的领域，如罕见病的诊断、局部影像异常识别以及治疗方案的精确制定等方面不适用。Wang等^[36]提出FedABC，即通过将多分类任务转换成多个二分类任务，之后使用二分类交叉熵尽可能保证二分类类别一致，从而减弱标签偏移带来的对联邦学习聚合模型的影响。在类别数量较少的情况下，该方法有利于实现类别之间的平衡，提高模型精度。但多分类任务转为多个二分类本身会造成极度的类别比例失衡，影响聚合模型性能。McLaughlin和Su^[37]则是提出FedLDA，通过减小表示器输出的局部数据特征分布差异性，对齐不同分布数据训练下的表示器参数，之后通过输出的特征分布，个性化聚合分类器。该方法精度和泛化性能较好，但各局部数据的特征在客户端之间传输，牺牲了数据的隐私安全，表1是对上述几类算法的总结。

表2是以上所述的部分算法在二分类任务中，不同数据分布时，聚合模型最终精度的对比结果。其中， λ 表示数据两个类别在数据中的比例。可以发现，当两个类别平衡时($\lambda=1$)，FedAvg算法表现良好，但数据中的两个类别比例失衡时，个性化联邦学习算法的表现明显优于传统联邦学习。

表1 不同个性化算法的优缺点总结

Tab. 1 Advantages and disadvantages of different personalized federated learning methods

名称	主要方法	优点	缺点
L2GD	正则化	通信成本较低	可拓展性较弱
FedAMP	正则化	通信成本较低	可拓展性较弱
pFedMe	正则化	可拓展性较强	泛化能力较弱
rFedAvg	正则化	泛化性强	数据安全性弱
IFCA	聚类	泛化性较强	计算和通信成本较高
OCFL	聚类	精度较高	抗干扰性弱
FLIS	聚类	抗干扰能力强	计算成本高
FeSEM	聚类	泛化性较强	性能受限于客户多样性
FedRoD	特征对齐	模型精度高	泛化能力弱
FedABC	特征对齐	泛化性较强	适应能力弱
FedLDA	特征对齐	精度高、泛化性好	数据安全性较弱

表2 不同算法在不同数据分布时的聚合精度对比

Tab. 2 Comparison of personalized federated learning methods under different data distributions

名称	数据分布差异程度			
	$\lambda=1$	$\lambda=2$	$\lambda=4$	$\lambda=7$
FedAvg	87.00±0.00	83.93±0.21	84.76±0.31	81.25±0.00
L2GD	77.00±0.20	76.70±0.60	85.44±0.31	86.08±0.12
FedAMP	79.33±0.31	72.28±0.12	84.26±0.60	86.58±0.12
IFCA	85.08±0.12	84.95±0.21	88.72±0.12	88.08±0.62
pFedME	79.58±0.31	78.15±0.12	84.76±0.12	86.67±0.42
FedRoD	87.00±0.00	86.39±0.12	91.75±0.31	89.83±0.12

2 医学数据在联邦学习的应用

2.1 多中心医学数据在联邦学习的应用

在诊断疾病时，医师往往需要结合核磁等影像模态、超声等时序模态以及病例等文本模态的数据进行确诊，体现医学数据的模态多样性。临床上详细的病例资料、高精度的成像检查以及繁多的检查化验项目，表明医学数据的空间维度大。而医学数据模态多、维度大的特点，决定了数据的数量和多样性对训练机器学习模型有重大影响。但当前的医疗水平并不能实现从单个医院收集到足够多的数据，充分训练机器学习模型^[38]。同时，公众对隐私保护的重视，在时间和经济成本上，提高了多中心医疗信息的交流门槛。因此，联邦学习以其降低多中心数据信息交流成本，逐渐走进医疗研究人员的视野。

Sheller等^[39]研究了参与联邦学习的医学中心数量对于生物标记物的诊断和预后模型最终性能的影响，结果支持参与联邦学习用户数量越多，各中心获得的收益越大。Rønn Hansen等^[40]通过联

邦学习模型,在确保多中心数据隐私的前提下,实现了对喉癌患者的分层Cox回归生存分析。Jiménez-Sánchez等^[41]利用多中心数据可以富集罕见病例的优势,通过联邦学习训练性能较好的能够鉴别X线成像的乳腺癌模型。Wang等^[42]利用联邦学习保护隐私的优势,对隐私程度较高的单细胞RNA序列进行分类训练,该模型性能不亚于集中式学习的模型性能。Truhn等^[43]深入优化联邦学习隐私保护的机制,对高度隐私的影像和病理数据中癌症分类模型进行训练,在确保数据安全的同时,获得性能优异的分类模型。Huang等^[44]考虑到各大药理机构的药理数据隐私保护和知识产权问题,使用联邦学习的方式,在多方协作下促进了新型药物的发现过程。Kassem等^[45]通过半监督场景下的联邦学习,在少量标注的情况下,实现更高性能的手术相位识别模型,降低了联邦学习的大规模推广的门槛。这些研究表明,联邦学习拥有不亚于集中式学习的学习能力,并且可以应用到医学行业的多个领域。但在遇到多中心数据分布差异性较大的情况时,传统联邦学习的模型性能将大打折扣。

2.2 非独立同分布数据在个性化联邦学习的应用

传统的联邦学习能够降低多中心医学数据的交流门槛,但多中心数据因其疾病种类分布不均、医院设备参数差异以及医师诊疗水平不同等问题,造成各中心数据的分布具有差异,即非独立同分布数据^[11,17]。这种数据将会对传统的联邦学习模型效能产生不利影响,而当前主要通过个性化联邦学习的方法减轻了这类问题对联邦学习性能造成的影响。

Chen等^[46]通过循环知识蒸馏的方式,实现联邦学习的个性化生理活动监测模型的训练,即使用训练到一定程度的局部模型参数引导其他局部模型更新,从而将该训练到一定程度的模型所包含的局部信息,通过模型引导的方式传递给其他模型,在保证本地信息的安全性前提下,提高了其流动性。相较于传统的联邦学习方法,这一方法保存了更多的局部信息,实现更优的模型性能和更高的保密程度。Peng等^[47]在考虑到不同国家和地区数据异质性较大的可能性下,分别通过对比集中式学习、传统联邦学习(FedAvg)和3种个性化联邦学习(FedAMP、FedProx、FedBN),对42家美国和欧洲医院的新冠肺炎影像数据进行训练,实验证明联邦学习的可行性,FedAvg的优势在于

其泛化能力较强,这可能与其主要聚合全局信息有关^[48]。而在聚合数据异质性较强的数据时,以FedAvg为代表的传统联邦学习方法局限性便凸显出来,这种模型主要聚焦于数量较大的类别,个性化联邦学习则在其内部验证上取得了比较好的结果,在实验上验证了个性化联邦学习优于传统联邦学习方法。Wicaksana等^[48]将模型拆分为特征提取器和分类器,其通过将本地特征提取器的参数与全局模型的特征提取器参数尽可能对齐,在保留本地模型的局部信息的基础上,尽可能保留全局信息,从而增强其泛化能力。同时,作者保留分类器参数不参与联邦聚合,使模型的分类器能够很好地适应本地数据的分布,提高了模型在本地数据上的精度,并在多源的前列腺癌和皮肤病数据上进行了验证。Tarumi等^[49]通过可解释性的协变量调整方法,首先在先验的联邦学习基础上,获得一个代表客户间协变量变化的倾向评分模型,然后根据估计的倾向评分对每个训练样本的个性化模型的贡献进行加权。实现个性化电子病历住院死亡率预测模型,既达到了优秀的准确性,又实现了对预测模型的解释。目前医疗领域个性化联邦学习有待更加深入的研究。

2.3 医学场景下的个性化联邦学习前景展望

当前正值人工智能从实验室走向实际应用的关键时间节点,医疗作为高度程序化的行业,借助人工智能技术实现诊疗向着更高质量、更高效率和更低成本的方向发展是该领域的必然选择^[38]。但人工智能技术面对高维度、多模态和大体积的医疗数据时,离不开充分且丰富的数据进行训练。联邦学习技术便是联通因数据隐私保护而形成的信息孤岛间的桥梁,其中个性化联邦学习技术优化了孤岛之间信息交流的方式,使得联邦学习技术更加贴近真实场景。同时,随着模型精简更新技术的发展^[50]、弱监督和无监督技术的优化^[51]以及大小模型协同训练技术的进步^[52],个性化联邦学习技术将向网络资源、专家资源和计算资源受限的地区覆盖,以减轻国家和社会医疗资源不平衡的问题。随着生物信息技术的发展,医学研究也将向着高维度、大容量和高成本的方向发展^[53],各大医院和研究所对于医学数据的研究将向算力倾斜,彼此交流也将更加密切,个性化联邦学习在降低交流成本的同时,在获得更优研究成果方面有更广阔的前景。

3 结语

本文概述了联邦学习和个性化联邦学习的优势与不足,并在理论层面上对个性化联邦学习的一些主要方法进行比较分析,梳理了这些方法的优缺点,为个性化联邦学习在医学场景中的应用做出理论分析、案例列举以及前景展望。我们对个性化联邦学习在医疗场景中应用持积极态度,不仅在于该方法能够低成本,打破各大医院之间因隐私保护造成的数据孤岛困局,更在于其在高效率收集数据、低成本分析数据和更密切交流数据上有着明显优势。

作者贡献 柳清河: 论文撰写; 姜明明: 审读和修订; 赵亚威: 监督指导, 资金支持; 何昆仑: 监督指导, 资金支持。

利益冲突 所有作者声明无利益冲突。

参考文献

- Kolachalama VB, Garg PS. Machine learning and medical education [J]. NPJ Digit Med, 2018, 1: 54.
- Esteva A, Robicquet A, Ramsundar B, et al. A guide to deep learning in healthcare [J]. Nat Med, 2019, 25 (1): 24-29.
- Khan S. Artificial intelligence and machine learning in clinical medicine [J]. N Engl J Med, 2023, 388 (25): 2398.
- Ristevski B, Savoska S. Healthcare and medical big data analytics [M] // Applications of Big Data in Healthcare. Amsterdam: Elsevier, 2021: 85-112.
- Yang X, Huang KX, Yang DW, et al. Biomedical big data technologies, applications, and challenges for precision medicine: a review [J]. Glob Chall, 2023, 8 (1): 2300163.
- Ray PP. ChatGPT: a comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope [J]. Internet Things Cyber Phys Syst, 2023, 3: 121-154.
- Zhou J, Ke P, Qiu XP, et al. ChatGPT: potential, prospects, and limitations [J]. Front Inf Technol Electron Eng, 2024, 25 (1): 6-11.
- 中华人民共和国国务院新闻办公室. 白皮书: 新时代的中国网络法治建设 [R/OL]. http://www.scio.gov.cn/zfbps/ndhf/49551/202303/t20230320_709279.html.
- 中华人民共和国国务院新闻办公室. 白皮书: 个人信息保护法 [R/OL]. https://www.gov.cn/xinwen/2021-08/20/content_5632486.htm.
- Abhishek VA, Binny S, Johan TR, et al. Federated learning: Collaborative machine learning without centralized training data [J]. International Journal of Engineering Technology and Management Sciences, 2022, 6 (5): 355-359.
- Li T, Sahu AK, Talwalkar A, et al. Federated learning: challenges, methods, and future directions [J]. IEEE Signal Process Mag, 2020, 37 (3): 50-60.
- Hu K, Li YG, Xia M, et al. Federated learning: a distributed shared machine learning method [J]. Complexity, 2021 (1): 8261663.
- Treleaven P, Smietanka M, Pithadia H. Federated learning: the pioneering distributed machine learning and privacy-preserving data technology [J]. Computer, 2022, 55 (4): 20-29.
- Zhu HY, Xu JJ, Liu SQ, et al. Federated learning on non-IID data: a survey [J]. Neurocomputing, 2021, 465 (C): 371-390.
- Wei BJ, Li J, Liu Y, et al. Federated learning for non-IID data: from theory to algorithm [C] // PRICAI 2021: Trends in Artificial Intelligence. Cham: Springer International Publishing, 2021: 33-48.
- Dinh CT, Vu TT, Tran NH. Personalized federated learning: theory and open problems [M] // Federated Learning. Amsterdam: Elsevier, 2024: 125-141.
- Tan AZ, Yu H, Cui LZ, et al. Towards personalized federated learning [J]. IEEE Trans Neural Netw Learn Syst, 2023, 34 (12): 9587-9603.
- Kulkarni V, Kulkarni M, Pant A. Survey of personalization techniques for federated learning [C] // 2020 Fourth World Conference on Smart Trends in Systems, Security and Sustainability (WorldS4). IEEE, 2020: 794-797.
- Azar KMJ, Moreno MR, Wong EC, et al. Accuracy of data entry of patient race/ethnicity/ancestry and preferred spoken language in an ambulatory care setting [J]. Health Serv Res, 2012, 47 (1 Pt 1): 228-240.
- Wagner MM, Hogan WR. The accuracy of medication data in an outpatient electronic medical record [J]. J Am Med Inform Assoc, 1996, 3 (3): 234-244.
- Cowden JD, Flores G, Chow T, et al. Variability in collection and use of race/ethnicity and language data in 93 pediatric hospitals [J]. J Racial Ethn Health Disparities, 2020, 7 (5): 928-936.
- Rees JR, Weiss JE, Gunn CM, et al. Cancer epidemiology in the northeastern United States (2013-2017) [J]. Cancer Res Commun, 2023, 3 (8): 1538-1550.
- Liu Y, Kang Y, Zou TY, et al. Vertical federated learning: concepts, advances, and challenges [J]. IEEE Trans Knowl Data Eng, 2024, 36 (7): 3615-3634.
- Liu DQ, Bai L, Yu TY, et al. Towards method of horizontal federated learning: a survey [C] // 2022 8th International Conference on Big Data and Information Analytics (BigDIA). IEEE, 2022: 259-266.
- McMahan HB, Moore E, Ramage D, et al. Communication-Efficient Learning of Deep Networks from Decentralized Data [C/OL]. <https://doi.org/10.48550/arXiv.1602.05629>.
- Arivazhagan MG, Aggarwal V, Singh AK, et al. Federated learning with personalization layers [EB/OL]. <https://arxiv.org/abs/1912.00818v1>.
- Hanzely F, Richtárik P. Federated learning of a mixture of global and local models [EB/OL]. <https://arxiv.org/abs/2002.05516v3>.
- Huang YT, Chu LY, Zhou ZR, et al. Personalized cross-Silo federated learning on non-IID data [J]. Proc AAAI Conf Artif Intell, 2021, 35 (9): 7865-7873.
- Dinh CT, Tran NH, Nguyen TD. Personalized Federated Learning with Moreau Envelopes [C/OL]. <https://doi.org/10.48550/arXiv.2006.08848>.
- Wang YS, Tong YX, Zhou ZM, et al. Distribution-regularized federated learning on non-IID data [C] // 2023 IEEE 39th International Conference on Data Engineering (ICDE). IEEE, 2023: 2113-2125.
- Ghosh A, Chung J, Yin D, et al. An efficient framework for clustered federated learning [J]. IEEE Trans Inf Theory, 2022, 68 (12): 8076-8091.
- Lu CY, Deng S, Wu YH, et al. Federated learning based on OPTICS clustering optimization [J]. Discrete Dyn Nat Soc,

- 2022 (1): 7151373.
- 33 Morafah M, Vahidian S, Wang WJ, et al. FLIS: clustered federated learning via inference similarity for non-IID data distribution [J]. IEEE Open J Comput Soc, 2023, 4: 109-120.
- 34 Long GD, Xie M, Shen T, et al. Multi-center federated learning: clients clustering for better personalization [J]. World Wide Web, 2023, 26 (1): 481-500.
- 35 Chen H, Chao W. On Bridging Generic and Personalized Federated Learning for Image Classification [C/OL]. <https://doi.org/10.48550/arXiv.2107.00778>.
- 36 Wang D, Shen L, Luo Y, et al. FedABC: targeting fair competition in personalized federated learning [J]. Proc AAAI Conf Artif Intell, 2023, 37 (8): 10095-10103.
- 37 McLaughlin C, Su L. FedLDA: Personalized Federated Learning Through Collaborative Linear Discriminant Analysis [C/OL]. <https://openreview.net/forum?id=1ww9tjEQVL>.
- 38 Garg A, Mago V. Role of machine learning in medical research: a survey [J]. Comput Sci Rev, 2021, 40: 100370.
- 39 Sheller MJ, Edwards B, Anthony Reina G, et al. Federated learning in medicine: facilitating multi-institutional collaborations without sharing patient data [J]. Sci Rep, 2020, 10 (1): 12598.
- 40 Rønn Hansen C, Price G, Field M, et al. Larynx cancer survival model developed through open-source federated learning [J]. Radiother Oncol, 2022, 176: 179-186.
- 41 Jiménez-Sánchez A, Tardy M, González Ballester MA, et al. Memory-aware curriculum federated learning for breast cancer classification [J]. Comput Methods Programs Biomed, 2023, 229: 107318.
- 42 Wang S, Shen BC, Guo LT, et al. scFed: federated learning for cell type classification with scRNA-seq [J]. Brief Bioinform, 2023, 25 (1): bbad507.
- 43 Truhn D, Tayebi Arasteh S, Saldanha OL, et al. Encrypted federated learning for secure decentralized collaboration in cancer image analysis [J]. Med Image Anal, 2024, 92: 103059.
- 44 Huang D, Ye XC, Sakurai T. Multi-party collaborative drug discovery via federated learning [J]. Comput Biol Med, 2024, 171: 108181.
- 45 Kassem H, Alapatt D, Mascagni P, et al. Federated cycling (FedCy): semi-supervised federated learning of surgical phases [J]. IEEE Trans Med Imag, 2023, 42 (7): 1920-1931.
- 46 Chen YQ, Lu W, Qin X, et al. MetaFed: federated learning among federations with cyclic knowledge distillation for personalized healthcare [J]. IEEE Trans Neural Netw Learn Syst, 2024, 35 (11): 16671-16682.
- 47 Peng L, Luo GX, Walker A, et al. Evaluation of federated learning variations for COVID-19 diagnosis using chest radiographs from 42 US and European hospitals [J]. J Am Med Inform Assoc, 2022, 30 (1): 54-63.
- 48 Wicaksana J, Yan ZQ, Yang X, et al. Customized federated learning for multi-source decentralized medical image classification [J]. IEEE J Biomed Health Inform, 2022, 26 (11): 5596-5607.
- 49 Tarumi S, Suzuki M, Yoshida H, et al. Personalized federated learning for institutional prediction model using electronic health records: a covariate adjustment approach [C] //2023 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). IEEE, 2023: 1-4.
- 50 Yu R, Li PC. Toward resource-efficient federated learning in mobile edge computing [J]. IEEE Netw, 2021, 35 (1): 148-155.
- 51 Schmarje L, Santarossa M, Schröder SM, et al. A survey on semi-, self- and unsupervised learning for image classification [J]. IEEE Access, 2021, 9: 82146-82168.
- 52 Min BN, Ross H, Sulem E, et al. Recent advances in natural language processing via large pre-trained language models: a survey [J]. ACM Comput Surv, 2023, 56 (2): 1-40.
- 53 Wouters OJ, McKee M, Luyten J. Estimated research and development investment needed to bring a new medicine to market, 2009-2018 [J]. JAMA, 2020, 323 (9): 844-853.

(责任编辑:孟晓彤)

(上接第103页)

- 10 Huang Q, Ren H, Chen P, et al. Prodigy: Enabling in-context learning over graphs [J/OL]. <https://doi.org/10.48550/arXiv.2305.12600>.
- 11 Dong QX, Li L, Dai DM, et al. A survey on in-context learning [C] //Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing. Stroudsburg: ACL, 2024: 1107-1128.
- 12 Min S, Lewis M, Zettlemoyer L, et al. MetaICL: learning to learn in context [C] //Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Stroudsburg: ACL, 2022: 2791-2809.
- 13 Wang B, Huang HY, Cao YX, et al. QRMeM: unleash the length limitation through question then reflection memory mechanism [C] //Findings of the Association for Computational Linguistics: EMNLP 2024. Stroudsburg: ACL, 2024: 4837-4851.
- 14 Liu R, Li MJ, Zhao S, et al. In-context learning for zero-shot medical report generation [C] //Proceedings of the 32nd ACM International Conference on Multimedia. Melbourne: ACM, 2024: 8721-8730.
- 15 Yao ZW, Wu XX, Li C, et al. Exploring post-training quantization in LLMs from comprehensive study to low rank compensation [J]. Proc AAAI Conf Artif Intell, 2024, 38 (17): 19377-19385.
- 16 Williams M, Aletas N. On the impact of calibration data in post-training quantization and pruning [C] //Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Stroudsburg: ACL, 2024: 10100-10118.
- 17 Zhang Y, Bai H, Lin H, et al. Plug-and-play: An efficient post-training pruning method for large language models [C/OL]. <https://openreview.net/pdf?id=Tr0IPx9woF>.
- 18 Yuan D, Rastogi E, Naik G, et al. A continued pretrained LLM approach for automatic medical note generation [C] //Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 2: Short Papers). Stroudsburg: ACL, 2024: 565-571.

(责任编辑:孙菲)