

解决联邦学习Non-IID问题的基础模型方法综述

王鑫, 陈坤, 孙凌云

引用本文

王鑫, 陈坤, 孙凌云. [解决联邦学习Non-IID问题的基础模型方法综述](#)[J]. 计算机科学, 2025, 52(12): 302-313.

WANG Xin, CHEN Kun, SUN Lingyun. [Research on Foundation Model Methods for Addressing Non-IID Issues in Federated Learning](#) [J]. Computer Science, 2025, 52(12): 302-313.

相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

[面向图垂直联邦学习的对抗攻击方法](#)

Adversarial Attack on Vertical Graph Federated Learning

计算机科学, 2025, 52(11A): 241200220-10. <https://doi.org/10.11896/jsjcx.241200220>

[基于知识蒸馏的联邦学习后门攻击方法](#)

Backdoor Attack Method for Federated Learning Based on Knowledge Distillation

计算机科学, 2025, 52(11): 434-443. <https://doi.org/10.11896/jsjcx.250100146>

[基于国密算法SM9的加法同态加密方案](#)

Additively Homomorphic Encryption Scheme Based on Domestic Cryptographic Algorithm SM9

计算机科学, 2025, 52(11): 408-414. <https://doi.org/10.11896/jsjcx.241100188>

[基于多策略改进的电鳗觅食优化算法](#)

Multi-strategy Improved Electric Eel Foraging Optimization Algorithm

计算机科学, 2025, 52(11): 245-254. <https://doi.org/10.11896/jsjcx.241100106>

[SAM-Retina:基于SAM的双模态视网膜图像动静脉分割](#)

SAM-Retina:Arteriovenous Segmentation in Dual-modal Retinal Image Based on SAM

计算机科学, 2025, 52(10): 123-133. <https://doi.org/10.11896/jsjcx.240800013>

解决联邦学习 Non-IID 问题的基础模型方法综述

王鑫^{1,2} 陈坤¹ 孙凌云²

1 浙江工业大学计算机科学与技术学院 杭州 310023

2 浙江大学计算机科学与技术学院 杭州 310058

摘要 联邦学习因具有隐私保护的天然特性,已经逐渐成为一个被广泛认可的分布式机器学习框架。但由于参与方数据分布的差异性,特别是呈现非独立同分布(Non-Independent and Identically Distributed, Non-IID)时,其面临着泛化性能不足、收敛性能下降、数据倾斜等严峻挑战。用预训练基础模型缓解 Non-IID 问题作为一种新颖的方法,演变出了各种各样的解决方案。对此,从预训练基础模型的角度,对现有工作进行了综述。首先介绍了基础模型方法,对典型的基础模型编码结构进行对比分析。其次从修改输入、基础模型部分结构再训练,以及参数高效微调 3 个角度,提出了一种新的分类方法。最后探讨了该类工作的核心难题和未来研究方向。

关键词: 联邦学习; 分布式系统; 隐私计算; 非独立同分布数据问题; 基础模型

中图分类号 TP181; TP309

Research on Foundation Model Methods for Addressing Non-IID Issues in Federated Learning

WANG Xin^{1,2}, CHEN Kun¹ and SUN Lingyun²

1 College of Computer Science and Technology, Zhejiang University of Technology, Hangzhou 310023, China

2 College of Computer Science and Technology, Zhejiang University, Hangzhou 310058, China

Abstract Federated learning, due to its inherent privacy-preserving nature, has gradually become a widely recognized framework for distributed machine learning. However, it faces significant challenges such as insufficient generalization performance, degraded convergence efficiency, and data skew, particularly in the presence of Non-IID data. Using pre-trained foundation models to mitigate Non-IID issues has emerged as a novel approach, leading to the development of various solutions. This review examines existing works from the perspective of pre-trained foundation models. Firstly, it introduces foundation model methodologies and provides a comparative analysis of typical foundation model architectures. Secondly, a new classification framework is proposed from three perspectives: modifying inputs, retraining parts of the foundation model, and parameter-efficient fine-tuning. Finally, it explores the core challenges of this type of work and outlines future research directions.

Keywords Federated learning, Distributed system, Privacy computing, Non-IID, Foundation models

1 引言

随着机器学习和大数据的迅猛发展,以及智能设备在日常生活中的普及,国民数字化程度不断加深,数字经济正逐渐成为国民经济的支柱。传统的集中式机器学习已无法满足数据共享带来的隐私保护要求,以联邦学习为代表的分布式机器学习可以有效解决此问题,其主要特点是允许多个客户端共同训练模型而无需共享其拥有的私有数据。联邦学习因在大规模应用中的潜在效应而受到学术界和工业界的广泛关注。然而,在实际生产中,参与者拥有的数据往往是差异巨大的,即数据是非独立同分布(Non-IID)的。参与者之间的分布差异,具体可以分为特征分布偏移、标签分布偏移、相同标签不同特征、相同特征不同标签和数量偏差五大类^[1]。

有学者已经证明 Non-IID 会对联邦学习的收敛性能产生影响,出现“客户端漂移”等问题^[2],而这个问题的出现主要是因参与方的数据分布不同(特别是特征分布和标签分布的差异)、计算能力不同而引起的模型参数发散,又被称为权重发散(Weight Divergence)^[3]。特别是在图联邦学习中,由于图结构、节点特征以及连接方式的异质性,Non-IID 问题的影响尤为突出。

目前,针对联邦学习的 Non-IID 问题,已经提出了较多方法,较流行的有共享数据、加入正则化项、多任务学习和元学习等^[4]。近年来,预处理基础模型方法作为一种新颖的方法,有望解决(缓解)Non-IID 问题带来的“客户端漂移”以及通信开销增加等问题。基础模型的联邦学习过程如图 1 所示,首先全局基础模型下发至各个参与方,参与方根据自己的本地

到稿日期:2024-12-09 返修日期:2025-01-20

基金项目:浙江工业大学科技项目(KYY-HX-20220288, KYY-HX-20180649)

This work was supported by the Zhejiang University of Technology Science and Technology Project(KYY-HX-20220288, KYY-HX-20180649).

通信作者:王鑫(xinw@zjut.edu.cn)

数据对基础模型进行训练,然后将更新的参数上传至全局

基础模型,全局对这些更新进行聚合收敛,以此重复。

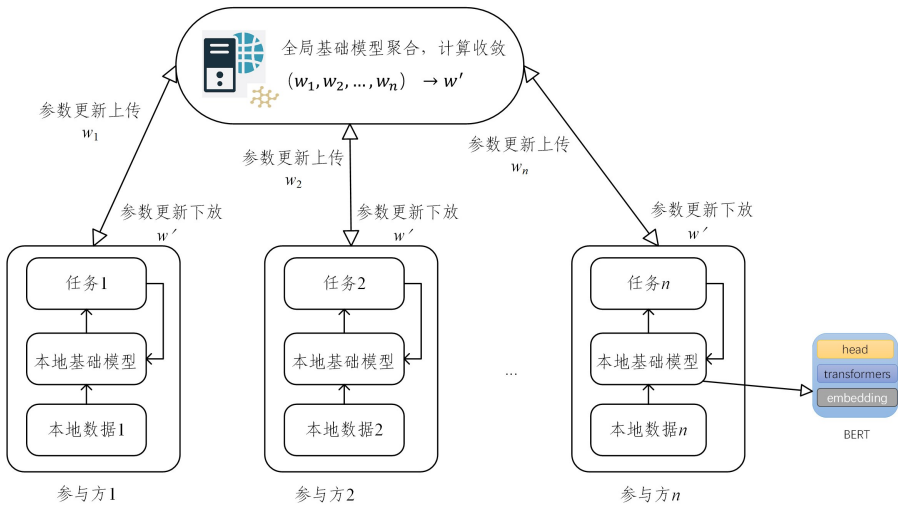


图 1 基础模型的联邦学习训练方法过程

Fig. 1 Federated learning training method process of the foundation model

尽管基础模型的多种应用取得了重大进展,但是仍然面临着诸多挑战,如计算成本过高,通信负载较大等。目前,用预处理基础模型方法去解决联邦学习的 Non-IID 问题是一个新兴的研究领域,随着研究的不断深入,该领域受到越来越多的关注。因此,有必要回顾和分析基础模型预处理方法和分析其中的优化策略。基础模型方法使用了多种优化算法,研究调查可以全面地阐述这些算法的优越性、局限性以及在不同场景中的实用性。研究这些优缺点和适用性,有助于研究人员未来提出新技术和新方向。

目前很少有使用基础模型解决联邦学习 Non-IID 方法的相关调查,对该方法的概念、技术和应用缺乏全面的了解。本文旨在填补这一空白,全面阐述了联邦学习基础模型使用的现有策略,对不同领域使用的预处理方法进行了概括和总结,使研究人员和从业者能够全面了解该领域。本文对基础模型提出了新的分类,按照过程的调整将其分成修改输入、基础模型部分结构再训练和参数高效微调三大类。同时,还对现有经典的基础模型特征提取结构进行了比较和分析,对现有联邦学习与预处理方法的核心难题进行了总结,对未来研究方向进行了展望。

近期,文献[5]已经综述了联邦学习和基础模型结合的

相关研究,但本文的研究方法和侧重点与其不同。文献[5]中强调了基础模型在联邦学习上解决隐私、数据去中心化和计算效率问题的能力,并对模型训练、聚合、可信度和激励方法进行分类;而本文侧重于从预训练基础模型的角度出发,综述现有基础模型方法缓解 Non-IID 问题的方法与技术,并以此提出了新的分类方法。

2 相关背景知识

2.1 个性化联邦学习

个性化联邦学习是为了应对联邦学习面对高度异构数据的收敛性差和缺乏个性化解决方案而提出的。有学者将个性化联邦学习分为解决异构数据训练全局模型的性能问题和解决方案个性化问题两大类^[6]。

个性化联邦学习能有效解决数据异质性问题,各参与方都在本地训练个性化模型,可以很好地根据本地数据的分布情况和设备性能,训练个性化的参与方模型,提高模型的泛化能力。基于此,提出众多方法框架,最近提出的预训练基础模型方法和个性化联邦学习之间的关系值得探析一番;首先值得强调的是,两者都是参与方基于本地数据微调,以适应本地数据的特性来抑制异质性;而两者的不同,如表 1 所列。

表 1 个性化联邦学习和预训练基础模型方法对比

Table 1 Comparison of personalized federated learning and pre-trained foundation model methods

对比项	个性化联邦学习	预训练基础模型方法
目的	为参与者定制模型,以适应本地数据的分布和需求,针对参与者的异质数据提供个性化解决方案	利用预训练的基础模型应对联邦学习 Non-IID 问题,缩短收敛时间,并增强性能
方法	全局模型引入个性化更新策略,客户端在全局模型上进行个性化微调	利用基础模型初始权重,缩短收敛时间,转移基础模型通用特征到客户端,并进行局部优化
性能	参与方个性化差异较大,会使全局模型收敛能力变差,需要在本地数据差异和全局模型收敛之间保持平衡	基础模型经过大量公用数据集预处理,性能强
隐私	个性化训练会牺牲用户隐私,差分隐私等技术在小数据集上效果较差 ^[7]	差分隐私等技术可在基础模型的公有数据集上训练 ^[8]

在基础模型方法中,没有将本地参与方的个性化放在一个较高的位置,但是,用基础模型方法实现个性化仍是一类可行方法。例如,大模型能生成高质量的合成数据,为每个参与

方创建个性化的数据集,并在此基础上进行个性化^[9]。

2.2 基础模型

预训练模型是指通过大量数据和计算资源训练出来的

大模型,可以直接使用或者经过后续微调后,迁移应用到新应用场景中。预训练模型一般包括图像大模型、语言大模型。目前,人工智能领域的预训练基础模型主要有 BERT, GPT, RoBERTa, T5 等,这些拥有超大规模参数的预训练模型也被称为基础模型^[10],即基础模型都是预训练模型,而预训练模型不一定是基础模型。

预训练基础模型在众多领域取得了优异成效,是如今的研究热点。BERT 和 GPT 均为基于 Transformer 的预训练语言模型,不同的是,前者使用了 Transformer 的编码部分,而后者使用了 Transformer 的解码部分。近年来,Transformer 的出现引发了很多架构创新,并且在大量数据集上预训练这些模型呈明显增长的趋势,因此,为客户配备大型的预训练 Transformer 成为一个简单可行的策略,将基础模型进行少量的本地适应性训练,就能较好地适应本地客户端,为联邦学习起到“热启动”的效果。此外,移动硬件设备、知识蒸馏技术的不断发展为客户端设备部署规模较大或者规模较小但效果相近的模型提供了可能。

最近,一些学者试图从基础模型角度去看待联邦学习的 Non-IID 问题,通过对基础模型的初始化和 Transformer 模型结构的调整,降低 Non-IID 带来的负面增益。Nguyen 等^[11]通过实验表明,从基础模型的初始化开始联邦学习可以减少数据和系统异构性的影响,同时缩短收敛时间和提高目标准确率。Chen 等^[12]指出基础模型能够很好地兼容 FEDPROX, FEDDYN 等联邦学习算法,改进这些算法能带来比集中学习更多的好处,从而适应非独立同分布或者客户端数据少等更具挑战的环境。

基础模型的出现带来了智能涌现和同质化。涌现是指系统的行为是隐式归纳,不是显示构造,可解释性差;同质化是指目前先进的 NPL 模型都源于少数几个基础模型之一,基础模型的任何一点改动就可以迅速普及,但也会继承基础模型的缺陷。基础模型已经初步展现出巨大潜力,应该尽快建立专业规范,使可靠的基础模型研究和部署成为可能。

2.3 基础模型编码结构比较

特征提取结构在基础模型中是不可或缺的组成部分,其负责从原始数据中提取出具有意义的特征,并将其特征转化、编码成模型可理解的形式。通过预训练,模型可以学习到通用的特征表示,并在不同领域和任务中都展现出不错的适应能力。这使得模型在面对新任务时,能有效应对非独立同分布问题带来的挑战,同时提高模型的泛化能力和鲁棒性。

目前,卷积神经网络(Convolutional Neural Network, CNN)、循环神经网络(Recurrent Neural Network, RNN)和 Transformer 是流行的预训练编码结构。在自然语言处理中,基于 RNN 的预训练模型擅长处理序列数据和长距离特征提取,能有效建模上下文信息,常用于机器翻译、语音识别。基于 CNN 的基础模型可并行计算,提升效率,但长距离特征提取能力较弱,多用于文本分类、情感分析。基于 Transformer 的基础模型能捕获全局依赖,并行计算,在机器翻译、文本生成中表现出色。基于 CNN 的基础模型是图像分类、目标检测的主流。基于 RNN 的基础模型应用较少,仅在视频分析

等图像序列处理中有所探索。此外,基于 Transformer 的基础模型(如 Vision Transformer, ViT^[13])也开始兴起。

不少研究支撑了这 3 种基础模型的对比。Qing^[14]的研究表明,基于 CNN 的基础模型在图像特征提取上优于 RNN。Lu 等^[15]发现,ViT 模型的准确率对大规模数据集更敏感。Merx 等^[16]指出,基于 Transformer 的基础模型在人类句子处理上优于 RNN。从整体性能而言,Transformer 具有明显优势,并被越来越多人接受,这点从引用的文献数量可以看出。这是否表明 CNN 和 RNN 没有继续研究的必要呢?

随着研究人员的不断创新,他们对各种网络变体进行了不同的性能对比。Qu 等^[17]在联邦学习环境中比较 Transformer 和 CNN 的优势,发现由于 Transformer 对异构数据的鲁棒性,两者之间的差异随着异构数据的增大而增大。但 Xu 等^[18]的研究结果表明,通过策略性的架构修改,纯 CNN 在处理 FL 中的异构数据客户端时可以达到与 ViT 相当,甚至超过 ViT 的鲁棒性水平。Tay 等^[19]在大量的数据集实验中对预训练后的 CNN 架构和 Transformer 进行比较,发现 CNN 的基础模型仍具有竞争力,并在某些情况下优于 Transformer。此外,目前基础模型的发展呈现出融合的趋势,即集多家之长。SwinUNet^[20]和 CMT^[21]架构交替使用 CNN 和 Transformer 层,前者提取局部特征,后者建模长距离依赖关系;RWKV^[22],Transformer-XL^[23]等模型结合了 RNN 与 Transformer 的双重优点。

3 基础模型方法分类

3.1 修改输入

基础模型的输入部分在很大程度上决定了最终模型的质量和泛化能力,合适的输入数据质量和规模可以减轻模型的过拟合和欠拟合风险,帮助模型学到通用的特征和模式。基于此,本节通过特定的手段提高模型输入的数据质量并扩大其规模,从而提高基础模型解决联邦学习的 Non-IID 问题的效率。修改输入的方式有共同学习和自监督学习两种。

3.1.1 共同学习

共同学习旨在通过合作和共享改进模型的性能,每个参与方独立训练模型,然后通过合作参数共同训练模型。这种方式有助于解决各参与方的数据量不足或者数据偏差等问题。Tan 等^[24]设计了一种联合原型对比学习(FedPCL)方法,该方法通过在参与方之间共享类别原型来传递知识,并采用原型对比的方式构建客户特定的表示。与共享可学习的模型参数不同,共享原型允许每个客户以个性化的方式整合表示,同时以紧凑的形式保留共享知识,以便进行有效的交流。

知识蒸馏也是共同学习的一种,其将联邦知识蒸馏的每个设备视为学生,并将其他设备的平均模型输出视为教师模型的输出,通过蒸馏损失计算来指导学生设备的训练。不同客户端的数据分布不一致,可能会导致模型在全局的泛化能力下降,联邦知识蒸馏技术可以在一定程度上缓解 Non-IID 问题对模型性能的影响。它通过在每个客户端上训练局部学生模型,并将全局教师模型的知识传输到每个客户端,使得模型在考虑到每个客户端数据分布特点的同时进行训练。即使

客户端之间的数据分布不一致,也可以通过知识蒸馏来提升模型的泛化能力和性能。Ma 等^[25]提出的 FedID 允许每次只处理和预测一小批未标记的公共数据集,从而增强知识转移过程中中心模型和本地模型之间的交互。此外,无数据知识蒸馏也是一个可行的方向,其从预训练的教师模型生成伪数据,并将教师模型上的知识转移到另一个学生模型。FedFTG^[26]设计了具有硬样本挖掘的无数据知识蒸馏方法,利用从本地模型提取的知识在服务器中微调该初步全局模型,从而隐式地缓解客户之间的分布差异问题。

3.1.2 自监督学习

联邦自监督学习是指从未标记的去中心化数据中进行自监督学习。自监督技术大致可以分为基于对比的方法、基于时间的方法和基于上下文的方法。在联邦学习中,当本地客户端的标记数据有限时,主要将对对比学习视为自监督学习^[27-29],又称为联邦对比学习。此外,由于联邦对比学习在缓解 Non-IID 问题上表现出不错的能力,因此受到广泛关注。Yan 等^[30]引入了基于 Transformer 的自监督学习范例,在高度异构且数量有限的分区中,开发了使用掩蔽图像建模

作为自监督任务的联邦预训练框架,显著提升了在高度异构数据上的联邦学习性能。Selfed^[31]使用基于 Swim Transformer 的编码器执行增强建模,引入对比网络和新颖的聚合策略,以分散的方式对目标标记有限的目标数据进行训练,对 Non-IID 数据有不错的效果。CLIP^[32]是一种结合了语言和图像的对比训练方法,能有效地借助自然语言的监督来学习视觉的概念。Ramesh 等^[33]使用带有 FedSGD 的对比预测编码框架预训练 LSTM 编码器,并在未标记语音数据集 Libri-Light 模拟的 Non-IID 数据上进行实验,展现出不错的效果。

3.1.3 小结和对比

修改输入的方法通过特定的手段提高模型输入的数据质量并扩大其规模,从而提高基础模型解决联邦学习的 Non-IID 问题的效率,如图 2 所示。共同学习通过合作和共享来改进模型的性能,而自监督学习则从未标记的去中心化数据中学习表示,如图 2 所示。这两种方法都不同程度地缓解了 Non-IID 问题,但它们的核心贡献在于通过数据增强和知识共享来提高模型的泛化能力。

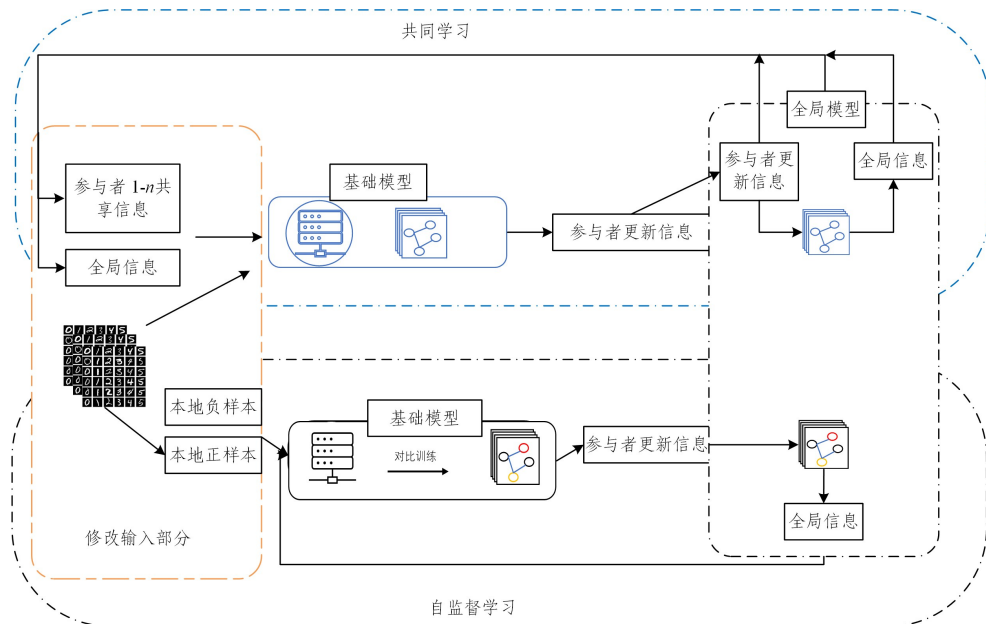


图 2 修改输入

Fig.2 Modify the input

联邦对比适用于未标记数据较多的场景,在高度异构的数据上表现较好,能够通过自监督学习提取通用特征,这在包含多种模态数据的物联网应用上有较大的优势,例如 Le 等^[34]提出了 DC-MMFed,通过对比学习不同模态之间的判别性特征,该模型适用于物联网应用中的多模态数据处理。而联邦知识蒸馏适用于需要提升模型精度和处理数据不平衡的场景,例如提供个性化服务的推荐系统和医学影像分析等医疗领域,但在处理多模态数据时不如对比学习灵活。Ni 等^[35]利用联邦知识蒸馏,在保护数据隐私的同时提升各医院训练目标多样性导致的参与度下降的问题,在异构环境中实现个性化医疗联邦学习。自监督学习适用于未标记数据多,或者数据稀缺、标注成本高昂的场景,能有效提升模型的泛化能力,这在疾病诊断以及医学影像分析中具有重要意义,如用

于胸部 CT 图像中 COVID 区域的分割^[36]、检测^[37],处理高分辨率的全切片图像^[38]等。但其缺点是需要消耗大量的计算资源进行预训练,预训练效果依赖于输入数据的质量。

3.2 基础模型部分结构再训练

基础模型通常具有较强的泛化能力,但在个性化能力和适应 Non-IID 数据方面的能力不佳。为了应对多样化复杂应用场景中的 Non-IID 数据分布问题和用户个性化需求,提取基础模型的部分结构并针对特定任务或场景进行再训练成为一种高效的解决方案。这种方式不仅能够保留基础模型的核心知识,还可以通过精简和微调使模型适应特定数据分布或用户需求。混合专家系统(Mixture of Experts, MoE)与云侧端侧大小模型协同训练这两类方法均通过提取部分结构再训练的方式,提高了计算效率和资源动态分配效率。

3.2.1 混合专家系统

MoE 模型将网络分解为多个专家和一个门控网络,每个专家负责处理特定类型的输入数据并生成预测结果。门控网络则根据输入数据的特定特征,动态地选择最合适的专家进行预测。这种设计允许模型在保持总参数数量的同时,通过仅激活相关的专家参数来降低推理时的计算成本,这对于联邦学习中的参数服务器和客户端设备来说是一个巨大的优势。在联邦学习的场景中,由于客户端存在 Non-IID 问题,路由器会动态选择与数据分布最合适的专家模型来进行局部训练,从而充分发挥该专家模型对当前客户端独特数据特征分布的适配优势,有效提升模型在该客户端本地数据上的任务性能;然后在服务器上进行专家模型的聚合更新;在全局更新后,可以对基础模型的参数进行再训练,尤其是部分专家结构的共享参数,以提高模型在不同任务之间的泛化能力。在采用离散路由策略时,除了 Top-K 激活的专家外,其他专家的参数无需进行计算。这进一步降低了模型在联邦学习中的通信和计算开销。

Fedmix^[39] 模型由多个专家组成,每个专家负责处理输入空间的一部分,门控机制负责根据输入数据的特征选择合适的专家。其通过在每个客户端上训练一组专家模型,并使用门控机制来确定哪些专家应该处理特定的数据点,有效解决了 Non-IID 问题。FedMoE^[40] 通过为每个客户端建立最优的子 MoE,并将知识带回全局 MoE,有效解决了数据异质性问题。pFedMoE^[41] 是个性化联邦学习和专家混合的方法,其为本地客户端的本地模型分配一个共享的同构特征提取器和本地门控网络,实现了数据层面的个性化学习,同时支持模型异构性。为解决微调大模型时的灾难性遗忘问题,Dou 等^[42] 将同一位置的 LoRA 专家分为两组,分别负责保存预训练权重中的世界知识和微调时学习的新任务,并为此目标设计了新的负载均衡 loss。Zadouri^[43] 通过将 MoE 架构与轻量级专家结合,提出了参数极其高效的 MoE,在性能上优于标准的参数高效微调方法。

3.2.2 云侧端侧大小模型协同技术

端侧设备具有天然靠近用户和数据源的优势,让端侧小模型与云侧大模型协同完成学习任务。充分利用云侧泛化能力与端侧个性化能力,是云端协同机器学习的新范式。一般流程是:云侧模型在聚合客户端的更新后进行再训练,以整合和适应客户端 Non-IID 数据分布问题,之后通过剪枝方法生成部分结构小模型,供端侧使用;端侧收到云侧更新后,利用本地数据进行基础模型部分结构的个性化再训练处理。如何利用端侧提升云侧大模型训练和推理效率,同时在提供个性化服务时保护隐私,解决端侧资源受限的问题,是当前研究的重点。

FedCoLLM^[44] 自适应地将服务器端 LLM 知识转移到客户的 SLM,同时利用客户的领域知识丰富 LLM,将轻量级适配器与 SLM 结合使用,协同调优 LLM 和 SLM。MocoSFL^[45] 将大骨干模型拆分为小客户端模型和大服务器端模型,通过向量连接、特征共享和频繁同步 3 个组件解决跨客户端应用中的计算量大和本地数据需求大的问题,在 Non-IID

下表现出较好的性能。LLMaaS^[46] 通过 API 微调,提供文本输入和输出的服务,但存在多轮通信成本过高,以及在大模型上缺乏探索等问题。CRaSh^[47] 通过聚类、删除和共享来增加场外微调,无需完整的云侧大模型,展现出在模型协调系统中实现“不可能三角”的潜力。Walle^[48] 是一个端到端、通用的端云协同系统,包括部署平台、数据管道、计算容器等组件,并在阿里巴巴大规模生产使用,产生了广泛的影响。

3.2.3 小结和对比

基础模型部分结构再训练的方法通过提取基础模型的部分结构并针对特定任务或场景进行再训练,提高了计算效率和资源动态分配效率。MoE 通过门控网络为客户端 Non-IID 数据动态匹配专家模型进行局部训练,随后服务器聚合更新并对基础模型进行再训练;云侧端侧大小模型协同训练,通过端云协同,提高端侧对本地异构数据的个性化能力,两者均有效提升了模型在 Non-IID 数据下的性能与适应性。

MoE 适用于客户端数据分布差距较大的场景,尤其是需要动态分配计算机资源的任务,从而通过门控网络选择最合适的专家处理数据,适应客户端的数据分布。例如,在大语言模型中,MoE 作为突出的架构可以很好地平衡模型性能和计算效率^[49]。云侧端侧大小模型协同适用于端侧资源受限、需要个性化服务的场景,其能降低端侧的计算资源成本,提高端侧模型的推理能力。例如,CE-CoLLM^[50] 框架通过早期退出机制、云上下文管理器和量化等技术,实现了低延迟的独立边缘推理;通过云侧端侧协同,提升了工业设备在不同工况下的故障诊断性能^[51]。

3.3 参数高效微调

在联邦学习的研究领域,通常将预训练模型的核心部分称为“主干网络”(Backbone Network),它在模型结构中起着至关重要的作用,不仅构成了模型的主体框架,还负责从数据中提取关键特征。在将这些基础模型适配到特定的下游任务时,研究者常常会对主干网络进行调整或优化,以提升模型在特定数据集上的表现。

随着基础模型参数规模的增长,对基础模型进行适配的计算代价也变得十分昂贵。近年的研究表明,大模型的适配并不需要更新全部参数,仅优化其中的极小部分参数并保持绝大部分参数不变,即可高效地将模型适配到特定的场景。尤其是在资源受限的联邦学习环境中,这一方法尤为有效。将这类参数高效适配的方法称为参数高效微调(Parameter-Efficient Fine Tuning, PEFT^[52]),其核心思想在于对大型模型的主干网络参数进行冻结,同时引入少量可训练的参数进行微调。

3.3.1 指定训练

指定训练的核心思想在于对大型模型的主干网络参数进行冻结,同时指定少量可训练的参数进行微调,如图 3 所示。这种方法不仅保持了模型原有的性能,而且大幅降低了显存和存储需求,为资源受限的联邦学习环境提供了一种有效的解决方案。Cai 等^[53] 提出的 TinyTL 方法通过在训练过程中冻结模型权重,学习小型的残差特征图,优化特征提取器,实现了在设备上的高效学习。Bitfit^[54] 专注于修改模型特定的

偏置参数部分,冻结大部分 Transformers 编码器参数,从而简化了模型的部署和训练。

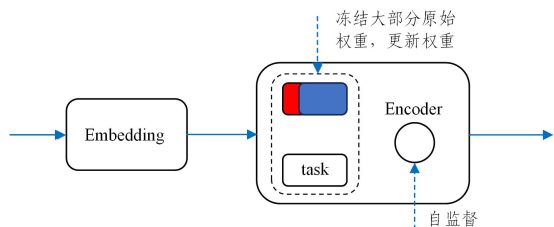


图3 冻结枝干

Fig. 3 Freeze the branches

3.3.2 添加模块

参数高效的调整方法在实现 Transformer 架构的轻量级适应方面显示出巨大的前景,在网络中集成小模块可以避免对后端架构的修改,降低成本。模块化指通过预训练参数高效的模块来适应各种下游任务,小的可调参数可以是基础模型层之间的轻量级神经适配器,或者是附加到输入示例的软提示^[55],如图4所示。Zhang等^[56]证明了模块化不仅可以大大减少通信成本,还可以提高局部适应方法的泛化能力和基础模型的鲁棒性。FedCK^[57]是一种带预训练模块的基于联邦卡尔曼滤波置信度的联邦学习算法,通过加入基于生成对抗网络的预训练模块来抑制 Non-IID,利用基于置信度的联邦卡尔曼滤波来增加鲁棒性。

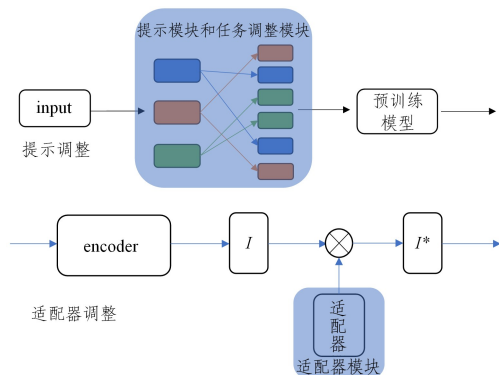


图4 添加模块

Fig. 4 Add modules

1) 提示学习

提示(prompt)是一种模块化的方法,指在基础模型中,为了使模型能够从接收的输入中学到有用的表示或者知识,需要给模型提供一些指导或者提示,以指导模型更关注哪些方面的信息。提示是基础模型技术中常见的方法,被广泛运用在自然语言处理和计算机视觉等领域。

提示学习(Prompt Learning)与传统的监督学习不同,其基于语言模型,直接对文本的概率进行建模^[58]。提示学习的优势主要是:通过一系列恰当的提示,一个以完全无监督方式训练的单一语言模型可以用于处理许多任务。提示学习得到广泛应用的主要原因是其语言模型在海量的原始文本上进行预训练,在定义一个新的提示函数的情况下,该模型能执行少样本,甚至零样本学习任务。

提示学习需要根据提示形式(填空提示、前缀提示),决

定采用手动还是自动方法来创建所需形式的提示。Jiang等^[59]提出了一种基于挖掘的自动寻找模板的 Mine 方法,给定一组训练输入 x 和输出 y ,该方法能从大量文本语料库(如维基百科)中抓取包含 x 和 y 的字符串,并找到输入和输出之间的中间词或依存路径。Wallace等^[60]在实际标记上应用基于梯度的搜索,以找到可以触发底层预训练语言模型生成所需目标预测的短序列。这种搜索以迭代方式进行,逐步通过提示中的标记。Tsimpoukelli等^[61]训练了一个视觉编码器,该编码器将图像编码成一系列嵌入向量,这些向量可以用来提示一个冻结的自回归语言模型生成适当的标题,所得到的模型可以执行视觉语言任务(如视觉问答等)的少样本学习。

pFedPT^[62]首先将 prompt 学习引入联邦学习,在原始数据和提示组成的输入中训练本地大模型,学习提示包含的数据分布信息,使模型具有根据自适应“微调”的能力,从而提高其解决客户端异质性等问题的综合能力。Yang等^[63]提出了一种新颖的客户端特定提示生成(pFed PG)的个性化 FL 框架,它学习在服务器上部署个性化提示生成器以生成客户端特定的视觉提示,可以有效地使冻结的主干适应本地数据分布,从而优化了本地个性化提示适应,实现了在各类数据异构性下的高效性能。Zhao等^[64]提出了 FedPrompt,以拆合聚合的方式冻结大量 PLM 的参数,仅调整和聚合软提示。相较于全参数微调,该方法极大减少了通信成本,提升了 Non-IID 下的准确性。Guo等^[65]提出了 PromptFL,用联邦提示训练代替联邦模型训练。具体地,客户端根据本地数据训练更新 PromptFL 提供的基础模型,共享软提示,PromptFL 只需要更新提示而不需要更新整个模型,显著增加了局部训练和全局聚合的能力。

2) 适配器调整

适配器调整是一种模块化的方法,用于微调基础模型以适应特定任务或领域的需求,而无需重新训练整个模型。该方法通过在基础模型的各个层之间添加适配器,来实现对模型的局部调整,从而提高模型在特定任务上的性能和泛化能力。Lu等^[66]提出了 FedCLIP,利用一个基于注意力的适配器的更新来代替整个模型的更新,以实现联邦学习中 CLIP 的快速泛化和个性化。FedDAT^[67]使用双适配器教师模块,一个是保持冻结状态的全局适配器,另一个在参与方本地优化,使本地适配器能够捕获特定于客户端的知识,有助于规范全局适配器并解决数据异构性问题。ASR^[68]基于 Conformer 的 RNN-T 主干进行预训练,通过联邦学习微调主干,在本地训练并存储,并结合个性化适配器来增强主干。CEFHRI^[69]在预先训练的模型中引入可训练的时空适配器,用于人机交互的视频理解任务,以解决数据异构性问题。

3) 小结和对比

指定训练冻结大型模型主干网络参数,微调少量指定参数,防止模型在 Non-IID 数据下过拟合,例如 TinyTL 和 Bitfit 分别通过特定方式在设备端训练中降低计算成本、简化训练,

提升模型在各客户端的泛化性能。添加模块通过引入可训练模块,如提示学习中的 pFedPT 利用提示让模型适应客户端异质性;适配器调整中的 FedCLIP 和 FedDAT 借助适配器实现快速泛化、捕捉特定知识,解决数据异构性问题,增强模型对 Non-IID 数据的适应能力与鲁棒性。

指定训练适用于内存资源受限的边缘设备,能够显著减少训练时的内存占用,提高训练速度和能量效率,但在复杂任务上的性能提升可能较为有限;提示训练和添加模块适用于资源受限、多任务学习,以及大规模的模型微调场景,在指令

调优、大模型领域的适配应用较为普遍,其微调成本低。X-PEFT^[70]通过微调极小的紧张张量集,自适应地选择适配器,相比传统适配微调,内存降低了 10 000 倍,成本效益高,但是其依赖于特定的模型架构,微调灵活性有限。

3.4 总结

三者从不同层面优化基础模型,对三者联系和差异的对比说明如表 2 所列。表 3 统计了代表性方法用于解决 Non-IID 问题的对比实验结果。为了更好地突出各个方法的性能特点,只选取了其报告的实验中部分数据集的对比结果。

表 2 分类方法的关系与差异对比

Table 2 Comparison of relationships and differences in classification methods

对比维度\方法	修改输入	基础模型部分结构再训练	参数高效微调
目的	通过提高输入数据质量并扩大其规模,增强基础模型解决联邦学习 Non-IID 问题的能力	针对特定任务或场景,提取基础模型部分结构再训练,提升模型对 Non-IID 数据的适应性和个性化能力	在资源受限的联邦学习环境中,通过高效微调少量参数,将基础模型与特定场景进行适配
对象	基础模型的输入数据	基础模型的部分结构,如混合专家系统中的专家模块、云侧端侧协同中的部分模型结构	基础模型的参数,冻结大部分参数,仅微调少量特定参数或添加少量可训练模块
优势	有助于解决参与方数据量不足或数据偏差问题,在数据有限时,利用自监督学习提升性能	提高计算效率和资源动态分配效率,MoE 可降低推理计算成本,云侧端侧协同可利用双方优势	大幅降低显存和存储需求,减少通信成本,提高局部适应方法的泛化能力和基础模型的鲁棒性
局限性	共同学习中知识蒸馏可能因数据分布不一致,影响模型泛化能力;自监督学习在复杂任务中的效果有待提升	MoE 模型结构复杂,训练难度增加;云侧端侧协同面临通信开销大、隐私保护和端侧资源受限等问题	当前方法不够灵活,对不同任务和数据集的优化能力有限,部分方法依赖特定模型架构
应用领域	物联网应用、个性化推荐、医疗影像分析等	大语言模型应用、自动驾驶、医疗影像分析等	低资源或专业领域语言翻译、个性化文本生成等

表 3 分类方法解决 Non-IID 问题的效果对比

Table 3 Comparison of the effects of classification methods in solving Non-IID problems

所属分类	方法	基线方法	数据集	异构设置	效果对比				
					方法\数据集	Digit-5	Office-10	DomainNet	—
修改输入	FedPCL	FedAvg, Solo	Digit-5, Office-10, DomainNet	特征异构, 标签异构	FedAvg	32.98	33.84	19.25	—
					Solo	38.35	36.38	27.15	—
					FedPCL	45.22	41.4	32.23	—
					方法\数据集	WNLI	CoLA	MNLI	—
	FedID	FedAvg	glue	迪利克雷分布 Dir($\alpha = \rho$)	FedAvg	48.3	48.9	81.4	—
					FedID	50.9	49.6	81.9	—
	FedFTG	FedAvg	CIFAR10, CIFAR100	迪利克雷分布 Dir($\alpha = 0.3$)	方法\数据集	IID	CIFAR10	IID	CIFAR100
					FedAvg	83.78	79.59	50.29	50.17
					FedFTG	87.34	84.38	56.94	55.96
	SelfFed	FedAvg	COVID-FL	迪利克雷分布 Dir($\alpha = 0.5$)	方法\数据集	COVID-FL	—	—	—
FedAvg					74.6	—	—	—	
SelfFed					82.1	—	—	—	
基础模型部分结构再训练	fedmix	FedAvg	CIFAR10, CIFAR100	迪利克雷分布 Dir($\alpha = 1$)	方法\数据集	CIFAR10	CIFAR100	—	—
					FedAvg	85.98	65.67	—	—
					FedMix K=4	88.54	67.54	—	—
	pFedMoE	FedProto	CIFAR10, CIFAR100	迪利克雷分布 Dir($\alpha = 0.3$)	方法\数据集	CIFAR10	CIFAR100	—	—
					pFedMoE	42.23	9.61	—	—
					FedProto	53.64	11.23	—	—
	MocoSFL	FL-BYOL	CIFAR10, CIFAR100	—	方法\数据集	CIFAR10	CIFAR100	—	—
FL-BYOL					83.34	61.78	—	—	
MocoSFL-1					87.81	58.78	—	—	
参数高效微调	FedCK	FedAvg	CIFAR-10, SVHN	—	方法\数据集	IID	CIFAR10	IID	SVHN
					FedAvg	47.1	30.72	79.69	60.57
					FedCK	54.6	43.16	98.65	76.25
	pFedPT	FedAvg	CIFAR-10, CIFAR100	迪利克雷分布 Dir($\alpha = 0.3$)	方法\数据集	IID	CIFAR10	IID	CIFAR10
					FedAvg	60.5	53.01	29.6	25.93
					pFedPT	60.01	74.92	31.66	36.8
	pFedPG	FedAvg	CIFAR-10, CIFAR100	迪利克雷分布 Disjoint	方法\数据集	CIFAR10		CIFAR100	
						Disjoint	Dir(0.1)	Disjoint	Dir(0.1)
					FedAvg	88.04	79.79	63.33	51.37
					pFedPG	90.08	87.57	70.96	55.91

4 联邦基础模型的核心难题

4.1 通信开销

在联邦学习中,客户端和服务端之间需要不断地传输模型参数和更新信息。因此,通信开销增加是一个不可避免的问题。在网络上传输模型的参数量(特别是大模型的参数量)巨大,在网络带宽有限或不可靠的情况下,可能会消耗大量的资源。大量边缘节点的存在会增加计算成本以及所需的计算能力和存储容量,网络带宽的差异可能会导致本地模型从客户端发送到服务器时出现延迟甚至丢失。此外,频繁的通信还会增加数据在传输过程中被截获的风险,带来隐私问题。因此,提高通信效率和有效性是联邦学习研究的重点。

目前,降低联邦学习的通信开销的主要方法分为:基于降低模型更新频率、基于模型压缩、基于客户端选择,以及其他方法。降低模型更新频率通过增强参与方的本地计算量(Federated Proximal)和提高并行度(重叠联邦平均)来实现^[71]。模型压缩主要是联邦蒸馏和模型量化,联邦蒸馏只交换局部模型输出而非交换传统模型参数,这些输出的尺寸通常比模型尺寸小得多,因此可以减少通信消耗。COMET^[72]使用集群协同蒸馏的方式设计了一个实用的 PFL 框架,通信成本降低了几个数量级。此外还有 FedCodl^[73],Gan^[74]。模型量化通过本地梯度计算转化为低精度值,而不是直接上传原始梯度值,从而降低了通信过程中的数据量和比特数,能有效减少通信成本,提高通信效率,但也会带来精度损失。通过合理的量化策略,可以在保持模型精度的同时降低通信成本,例如 AQG^[75]根据本地梯度更新自适应调整量化级别,充分利用本地数据分布的异质性以减少不必要的传输,减少了 25%~50% 的额外传输。

联邦蒸馏和模型量化都可以降低通信开销,但两者对模型性能的影响是不同的。通常来说,蒸馏都可以提高模型性能,而量化会在一定程度上降低模型准确性。在带宽影响上,量化可以直接减少每轮通信开销,但是精度的降低可能会需要更多的通信轮次,而蒸馏对带宽的适应性相对较好。

客户端选择是选取一部分参与方参与训练过程,以适应带宽受限。Fed-PLT^[76]允许部分参与方参与本地训练,从而解决通信问题,减少了中央协调器和计算代理之间的通信次数;Ayache 等^[77]提出一种基于随机游走的自适应训练算法,通过 Sleeping MAB 框架设计随机游走策略以优先选择高信息量节点,用其本地数据更新模型,其收敛速度显著优于基准算法,且在数据异质性较强的场景下表现更加突出。

4.2 复杂异构性

异构性一直是联邦学习不可绕过的核心难题,异构联邦学习可分为统计异构性、设备异构性、模型异构性和通信异构性^[78]。先前讨论的预训练处理联邦学习的 Non-IID 问题就属于统计异构性问题。在真实的联邦学习应用场景中,异构性是常见的,因此仅仅讨论统计异构性是不够的。例如,在设备异构性方面,除了设备本身的性能之外,设备在负载能力和承受负载能力上也存在着较大差异,这对联邦学习的整体性能有很大影响。从算法设计上看,掉队者效应^[79]和拜占庭设备等问题很难通过算法方面的改进完全解决。在实际工程应

用中,成百上千的设备型号、各种机器学习框架以及操作系统给用户端的开发带来巨大的挑战。其次,在统计异构性方面,除了数据分布带来的 Non-IID 问题外,在实际生产关联的数据中(如工业数据)还存在着标签缺失、样本质量不齐、概念漂移等^[80]问题,使得在数据集中提取共性知识训练模型变得更加困难,因此需要在隐私保护的基础上,探索在异构联邦学习中提高全局模型泛化能力和性能的方法。

自适应学习旨在让模型根据环境或数据的变化自动调整自身的自调整梯度平衡器模块组成,其能根据全局长尾先验的反馈,以闭环方式重新加权客户端的梯度,可以有效缓解模型训练过程中数据异质性导致的分布漂移。

4.3 本地计算资源限制

联邦学习是一种分布式机器学习技术,涉及多个设备和节点的同时训练,每个设备都需要执行一些计算任务,与其他设备进行数据和参数的交换。由于参与方的计算资源有限,因此需要设计有效的算法和策略,以最大限度地利用这些资源。计算卸载是边缘计算的一项关键技术,其通过无线网络将任务按照既定策略分发到资源充足的服务器上执行,并将执行结果回传至终端设备,可以有效利用各参与方的计算资源,提高计算效率。FedAdapt^[82]是一种自适应卸载 FL 框架,它通过深度神经网络到服务器的层卸载来加速计算受限设备的本地训练,同时用基于强化学习的优化和聚类,先将计算与带宽相似的设备聚类分组,RL 智能体为每组自适应地确定最优卸载点,后处理模块再将组策略映射至组内设备,以应对计算异构性和网络带宽变化的挑战。FedFC^[83]利用模型拆分和特征连接,将部分训练负载从客户端卸载到聚合服务器,其中每个客户端都可以协作训练具有不同切割层的模型,从而减少资源受限的参与方的计算负载,加快收敛速度。

除了提高参与方对本地计算资源的利用率,提高本地计算资源的能力也是可行的解决方案。具体地,提高参与方的硬件能力(GPU,TPU 等),利用缓存技术将频繁使用的数据或计算结果存储在本地,以减少重复计算的开销。例如,将模型参数存储在闪存中,按需将它们传到 DRAM,利用“窗口化”,通过重用先前激活的神经元来战略性地减少数据传输和针对闪存的顺序数据访问优势而定制的“行列捆绑”,以及增加从闪存读取的数据块的大小等方法,使得其可支持运行可用 DRAM 两倍大小的模型。与 CPU 和 GPU 中的简单加载方法相比,该方法的推理速度分别提高了 4~5 倍和 20~25 倍^[84]。

5 未来研究方向

5.1 隐私加强

隐私安全是联邦学习提出的初衷,客户端不直接共享原始数据,而仅通过共享更新数据来保护基本的隐私。然而事实上,客户端仍可能向服务器泄露私人信息,例如 Zhu 等^[85]发现可以从梯度中反推出用户的个人信息,导致隐私泄露。Pan 等^[86]提出了针对嵌入式表示的攻击算法来重构目标文本的关键词;算法假定攻击者知道并利用已知的基础模型,将公开数据集及其标签输入基础模型生成嵌入式表示,并以此

作为训练集来训练攻击模型。此外,联邦学习在预训练中还可能容易受到模型投毒、后门攻击等,进一步揭示了隐私保护面临的严峻挑战^[87]。

隐私加强是联邦学习未来的一个重要研究方向,研究人员引入了密码学和可信硬件等技术手段,以增强联邦学习系统的隐私保护能力。目前,针对联邦学习的隐私加强技术主要分为加密、扰动和可信硬件 3 类^[88]。差分隐私是主流的隐私加强技术,其通过向模型的输出加入噪声,使得外部无法辨别相差一个样本的相邻数据集,以保护原始私人信息。但是,差分隐私仅在相对较大的数据集上得到了效果证明,而在个性化联邦学习这种数据集较小的情况下,效果尚未明确^[89]。差分隐私模型通常在大型数据集上进行预训练,然后在相对较大且分布与预训练数据相似的私有下游数据集上进行微调。因此,将差分隐私框架集成到基础模型中可能是未来的方向。此外,将差分隐私和密码学结合也是一种可行的策略,Hybird-Alpha^[90]的联邦计算框架通过融合差分隐私和密码学技术,提供了健壮且安全的隐私保护方案。

此外,还可以考虑结合硬件级别的安全机制,如可信执行环境(TEE)和安全多方计算(MPC),来增强隐私保护能力;研究在基础模型的预训练阶段融入隐私保护机制,如通过隐私增强的自监督学习或多方预训练,来确保模型在联邦学习中的隐私性。

5.2 参数高效微调的提升

事实上,对基础模型进行微调已经成为一种学习范式,然而,传统的微调方法微调了基础模型的所有参数,随着基础模型的规模和任务数量的增长,训练成本越来越高。参数高效微调虽然仅微调少量的参数就能获得强大的性能,但也存在不足之处。首先,当前的参数高效微调方法不够灵活,无法针对所有类型的任务和数据集进行优化,对参数高效微调方法技术背后成功的重要因素没有足够的认识,这些限制了参数高效微调技术在多样化任务上的应用。其次,参数高效微调技术在数据量和参数较少的应用中没有得到很好的效果证明,这限制了其在计算资源限制等场景下的应用。不仅如此,目前许多参数高效微调方法依赖于特定模型架构,需要对模型架构进行特定的调整,限制了其在不同模型上的应用,例如 Adapter 等方法需要插入层和模块,依赖于 Transformer 结构,致使在其他模型结构上的使用受限。

基础模型在 FL 客户端进行适配时广泛使用有代表性的几种 PEFT 方法,但在解决 Non-IID 问题时到底采用哪类方法最好并无理论证明,而是采用了一些逐类尝试的经验性验证方法。目前,已经有工作^[91]尝试使用统一框架来对 Adapter, Prefix Tuning, LoRA 这 3 种代表性方法建立一个统一视角。Ding 等^[92]也指出,参数高效微调方法的本质是对“增量参数”(Delta Parameters)进行调整,因此将此类方法命名为“增量微调”(Delta Tuning)。他们还进一步从参数优化和最优控制两个角度,提出了增量微调的理论框架,为探索和解释增量微调的内在机理提供了可行方案。

微调混合策略是未来研究中可行且有前景的方向,它通过在基础模型中综合应用不同层次和程度的参数调整,实现了灵活性与稳定性的平衡。目前,较为流行的混合微调方法

包括部分层微调与冻结、混合权重微调,以及逐层调整等。未来的发展可能集中于自动化调节策略和自适应增强,借助自动化工具选择最优的混合微调配置,以降低手动调参的复杂性并提高模型的适应性与效率。

5.3 预训练模型架构改进

现有的基础模型架构主要有 CNN, Transformer 等, Transformer 已经被证明是一个有效的预训练架构,得到了广泛使用。然而,Transformer 的计算复杂度较高,容易受到训练 GPU 设备的限制,当前大多数的基础模型无法处理包含 512 个标记的序列^[93]。CNN 架构虽然日常中无法达到 Transformer 的水平,但是有研究表明,对 CNN 进行架构修改,可以使其在异构数据上的处理能力超越 Transformer。因此,改进架构或寻求更有效的模型架构至关重要。

目前,神经架构搜索的自动方法是一个可能的未来研究方向。例如,开发基于强化学习或进化算法的神经架构搜索算法,以联邦学习的性能指标(如准确率、收敛速度等)为优化目标,在搜索空间中自动探索最优架构。此外,不同的下游任务适合不同的架构,需要根据下游任务的类型设计特定于任务的架构。例如,在图像领域,针对不同分辨率、不同场景的图像数据,需设计专门的 CNN 与 Transformer 融合架构,充分发挥两者优势,提升模型在特定任务上的性能;此外,还需探索新型模型架构,结合新的计算原理和技术,突破现有架构的局限,提高模型处理 Non-IID 数据的能力和效率。

5.4 多模态预训练

随着多模态数据量的快速增长和多模态模型训练的普及,未来的研究及开发可以在联邦学习环境中使用多模态基础模型。多模态基础模型可以通过学习不同模态数据之间的关联性,一定程度上缓解 Non-IID 数据带来的问题,使得模型能够更好地适应不同参与者的数据分布。但是与单模态相比,多模态联邦学习因为跨客户端的数据分布和模态分布的异质性,其性能会显著下降。Chen 等^[94]提出了层次梯度混合方法,并在大量的 Non-IID 多模态数据上进行实验,结果表明该方法有着很好的表现。多模态联邦学习的发展还处于初期, Lin 等^[95]指出,多模态联邦学习的主要挑战在于多模态对齐,而现有大多数多模态模型都是需要客户端对齐,很少对数据集对齐,因此需要提前对多模态数据进行预处理,限制了算法的性能。

未来可研究更高效的多模态对齐算法,例如,基于注意力机制或图神经网络的对齐方法,提高不同模态数据的对齐精度和效率。探索对数据集进行统一对齐的方法,在数据预处理阶段对多模态数据进行深度融合和对齐,减轻客户端的处理负担,提升算法性能。将多模态基础模型与联邦学习更紧密结合,设计适合多模态联邦学习的训练算法和模型结构,兼顾处理多模态数据的性能和收敛能力。例如,开发能够同时处理多种模态数据的联邦学习算法,在模型训练过程中充分利用不同模态数据的互补信息,提高模型的性能和泛化能力。

结束语 作为一种新兴的机器学习范式,联邦学习的出现在保障数据隐私安全的基础上,有效解决了数据孤岛的问题,为人工智能应用赋能。然而,在实际应用中,数据异构性,特别是当数据以 Non-IID 进行分布时,联邦学习的性能和收

敛能力会受到严重影响。用基础模型处理缓解 Non-IID 问题被证明是可行的方案。对此,本文对现有的基础模型方法进行了合理评估和分类总结,以推动了联邦学习的长效发展。

本文综述了联邦学习的基础模型的方法技术,对现有的基础模型编码结构进行了深入总结和对比,描述了基础模型解决 Non-IID 问题的方法,并对其进行归纳和分类。此外,探讨了现有联邦学习需要解决的核心难题和未来研究的方向。

参 考 文 献

- [1] KAIROUZ P, MCMAHAN H B, AVENT B, et al. Advances and Open Problems in Federated Learning [J]. Foundations and Trends © in Machine Learning, 2021, 14(1/2): 1-210.
- [2] KARIMIREDDY S P, KALE S, MOHRI M, et al. SCAFFOLD: Stochastic Controlled Averaging for Federated Learning [C] // Proceedings of the 37th International Conference on Machine Learning. PMLR, 2020: 5132-5143.
- [3] ZHAO Y, LI M, LAI L, et al. Federated Learning with Non-IID Data [J]. arXiv:1806.00582, 2018.
- [4] GAO D, YAO X, YANG Q. A Survey on Heterogeneous Federated Learning [J]. arXiv:2210.04505, 2022.
- [5] REN C, YU H, PENG H, et al. Advances and open challenges in federated learning with foundation models [J]. arXiv: 2024.15381, 2020.
- [6] TAN A Z, YU H, CUI L, et al. Towards Personalized Federated Learning [J]. IEEE Transactions on Neural Networks and Learning Systems, 2023, 34(12): 9587-9603.
- [7] LI S, LIU Y, FENG F, et al. HierFedPDP: Hierarchical federated learning with personalized differential privacy [J]. Journal of Information Security Applications, 2024, 86: 103890.
- [8] FU J, YE Q, HU H, et al. DPSUR: accelerating differentially private stochastic gradient descent using selective update and release [J]. arXiv:2311.14056, 2023.
- [9] CHEN J, LIU Z, HUANG X, et al. When Large Language Models Meet Personalization: Perspectives of Challenges and Opportunities [J]. arXiv:2307.16376, 2023.
- [10] BOMMASANI R, HUDSON D A, ADELI E, et al. On the opportunities and risks of foundation models [J]. arXiv: 2108.07258, 2021.
- [11] NGUYEN J, WANG J, MALIK K, et al. Where to Begin? On the Impact of Pre-Training and Initialization in Federated Learning [J]. arXiv:2206.15387, 2022.
- [12] CHEN H Y, TU C H, LI Z, et al. On the Importance and Applicability of Pre-Training for Federated Learning [J]. arXiv: 2206.11488, 2022.
- [13] ALEXEY D. An image is worth 16x16 words: Transformers for image recognition at scale [J]. arXiv:2010.11929, 2020.
- [14] QING X. A Comparison Study of Convolutional Neural Network and Recurrent Neural Network on Image Classification [C] // Proceedings of the 2022 10th International Conference on Information Technology: IoT and Smart City. ACM, 2022: 112-117.
- [15] LU K, XU Y, YANG Y. Comparison of the potential between transformer and CNN in image classification [C] // ICMLCA 2021; 2nd International Conference on Machine Learning and Computer Application. 2021: 1-6.
- [16] MERKX D, FRANK S L. Comparing Transformers and RNNs on predicting human sentence processing data [J]. arXiv:2005.09471, 2020.
- [17] QU L, ZHOU Y, LIANG P P, et al. Rethinking Architecture Design for Tackling Data Heterogeneity in Federated Learning [J]. arXiv:2106.06047, 2021.
- [18] XU P, WANG Z, MEI J, et al. FedConv: Enhancing Convolutional Neural Networks for Handling Data Heterogeneity in Federated Learning [J]. arXiv:2310.04412, 2023.
- [19] TAY Y, DEGHANI M, GUPTA J, et al. Are Pre-trained Convolutions Better than Pre-trained Transformers? [J]. arXiv: 2105.03322, 2021.
- [20] CAO H, WANG Y, CHEN J, et al. Swin-Unet: Unet-Like Pure Transformer for Medical Image Segmentation [C] // Computer Vision—ECCV 2022 Workshops. Springer, 2023: 205-218.
- [21] GUO J, HAN K, WU H, et al. CMT: Convolutional Neural Networks Meet Vision Transformers [J]. arXiv:2107.06263, 2021.
- [22] PENG B, ALCAIDE E, ANTHONY Q, et al. RWKV: Reinventing RNNs for the Transformer Era [J]. arXiv: 2305.13048, 2023.
- [23] DAI Z, YANG Z, YANG Y, et al. Transformer-XL: Attentive Language Models Beyond a Fixed-Length Context [J]. arXiv: 1901.02860, 2019.
- [24] TAN Y, LONG G, MA J, et al. Federated learning from pre-trained models: A contrastive learning approach [J]. Advances in Neural Information Processing Systems, 2022, 35: 19332-19344.
- [25] MA X, LIU J, WANG J, et al. FedID: Federated Interactive Distillation for Large-Scale Pretraining Language Models [C] // Conference Association for Computational Linguistics. 2023: 8566-8577.
- [26] ZHANG L, SHEN L, DING L, et al. Fine-tuning Global Model via Data-Free Knowledge Distillation for Non-IID Federated Learning [J]. arXiv:2203.09249, 2022.
- [27] ZHANG F, KUANG K, CHEN L, et al. Federated unsupervised representation learning [J]. Frontiers of Information Technology & Electronic Engineering, 2023, 24(8): 1181-1193.
- [28] ZHUANG W, GAN X, WEN Y, et al. Collaborative Unsupervised Visual Representation Learning from Decentralized Data [J]. arXiv:2108.06492, 2021.
- [29] ZHUANG W, WEN Y, ZHANG S. Divergence-aware Federated Self-Supervised Learning [J]. arXiv:2204.04385, 2022.
- [30] YAN R, QU L, WEI Q, et al. Label-Efficient Self-Supervised Federated Learning for Tackling Data Heterogeneity in Medical Imaging [J]. IEEE Transactions on Medical Imaging, 2023, 42(7): 1932-1943.
- [31] KHOWAJA S A, DEV K, ANWAR S M, et al. SelfFed: Self-supervised Federated Learning for Data Heterogeneity and Label Scarcity in IoMT [J]. arXiv:2307.01514, 2023.
- [32] RADFORD A, KIM J W, HALLACY C, et al. Learning Transferable Visual Models From Natural Language Supervision [C] // Proceedings of the 38th International Conference on Machine Learning. PMLR, 2021: 8748-8763.

- [33] RAMESH G V, CHENNUPATI G, RAO M, et al. Federated Representation Learning for Automatic Speech Recognition [J]. arXiv:2308.02013, 2023.
- [34] LE H Q, QIAO Y, NGUYEN L X, et al. Federated multimodal learning for iot applications: A contrastive learning approach [C]//2023 24st Asia-Pacific Network Operations and Management Symposium(APNOMS). IEEE, 2023;201-206.
- [35] NI L, SONG C, ZHAO H, et al. Personalized Medical Federated Learning Based on Mutual Knowledge Distillation in Object Heterogeneous Environment[C]//Blockchain and Web3 Technology Innovation and Application Exchange Conference. Springer, 2024;362-374.
- [36] YANG D, XU Z, LI W, et al. Federated semi-supervised learning for COVID region segmentation in chest CT using multi-national data from China, Italy, Japan [J]. Medical Image Analysis, 2021,70;101992.
- [37] NGUYEN D C, DING M, PATHIRANA P N, et al. Federated learning for COVID-19 detection with generative adversarial networks in edge cloud computing [J]. IEEE Internet of Things Journal, 2021,9(12);10257-10271.
- [38] LU M Y, CHEN R J, KONG D, et al. Federated learning for computational pathology on gigapixel whole slide images [J]. Medical Image Analysis, 2022,76;102298.
- [39] REISSER M, LOUZOS C, GAVVES E, et al. Federated Mixture of Experts [J]. arXiv:2107.06724, 2021.
- [40] MEI H, CAI D, ZHOU A, et al. FedMoE: Personalized Federated Learning via Heterogeneous Mixture of Experts [J]. arXiv:2408.11304, 2024.
- [41] YI L, YU H, REN C, et al. pFedMoE: Data-Level Personalization with Mixture of Experts for Model-Heterogeneous Personalized Federated Learning [J]. arXiv:2402.01350, 2024.
- [42] DOU S, ZHOU E, LIU Y, et al. LoRAMoE: Alleviate World Knowledge Forgetting in Large Language Models via MoE-Style Plugin [J]. arXiv:2312.09979, 2023.
- [43] ZADOURI T, ÜSTÜN A, AHMADIAN A, et al. Pushing Mixture of Experts to the Limit: Extremely Parameter Efficient MoE for Instruction Tuning [J]. arXiv:2309.05444, 2023.
- [44] FAN T, KANG Y, MA G, et al. FedCoLLM: A Parameter-Efficient Federated Cotuning Framework for Large and Small Language Models [J]. arXiv:2411.11707, 2024.
- [45] LI J, LYU L, ISO D, et al. MocoSFL: enabling cross-client collaborative self-supervised learning[C]//The Eleventh International Conference on Learning Representations, 2022;1-13.
- [46] YIN W, XU M, LI Y, et al. LLM as a system service on mobile devices [J]. arXiv:2403.11805, 2024.
- [47] ZHANG K, DING N, QI B, et al. CRaSh: Clustering, Removing, and Sharing Enhance Fine-tuning without Full Large Language Model [C]//Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing, 2023;9612-9637.
- [48] LYU C, NIU C, GU R, et al. Walle: An {End-to-End}, {General-Purpose}, and {Large-Scale} Production System for {Device-Cloud} Collaborative Machine Learning [C]//16th USENIX Symposium on Operating Systems Design and Implementation (OSDI 22). 2022;249-265.
- [49] QIAN Y, LI F, JI X, et al. EPS-MoE: Expert Pipeline Scheduler for Cost-Efficient MoE Inference [J]. arXiv:2410.12247, 2024.
- [50] JIN H, WU Y. CE-CoLLM: Efficient and Adaptive Large Language Models Through Cloud-Edge Collaboration [J]. arXiv:2411.02829, 2024.
- [51] WANG Q, LI Q, WANG K, et al. Efficient federated learning for fault diagnosis in industrial cloud-edge computing [J]. Computing, 2021,103(10);2319-2337.
- [52] LIALIN V, DESHPANDE V, RUMSHISKY A. Scaling Down to Scale Up: A Guide to Parameter-Efficient Fine-Tuning [J]. arXiv:2303.15647, 2023.
- [53] CAI H, GAN C, ZHU L, et al. Tinytl: Reduce memory, not parameters for efficient on-device learning [J]. Advances in Neural Information Processing Systems, 2020,33;11285-11297.
- [54] ZAKEN E B, RAVFOGEL S, GOLDBERG Y. Bitfit: Simple parameter-efficient fine-tuning for transformer-based masked language-models [J]. arXiv:2106.10199, 2021.
- [55] SUN T, HE Z, ZHU Q, et al. Multitask Pre-training of Modular Prompt for Chinese Few-Shot Learning [J]. arXiv:2210.07565, 2022.
- [56] ZHANG X, LI M, CHANG X, et al. FedYolo: Augmenting Federated Learning with Pretrained Transformers [J]. arXiv:2307.04905, 2023.
- [57] HU K, WU J, WENG L, et al. A novel federated learning approach based on the confidence of federated Kalman filters [J]. International Journal of Machine Learning and Cybernetics, 2021,12(12);3607-3627.
- [58] LIU P, YUAN W, FU J, et al. Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing [J]. ACM Computing Surveys, 2023,55(9);1-35.
- [59] JIANG Z, XU F F, ARAKI J, et al. How can we know what language models know? [J]. Transactions of the Association for Computational Linguistics, 2020,8;423-438.
- [60] WALLACE E, FENG S, KANDPAL N, et al. Universal adversarial triggers for attacking and analyzing NLP [J]. arXiv:1908.07125, 2019.
- [61] TSIMPOUKELLI M, MENICK J L, CABI S, et al. Multimodal few-shot learning with frozen language models [J]. Advances in Neural Information Processing Systems, 2021,34;200-212.
- [62] LI G, WU W, SUN Y, et al. Visual Prompt Based Personalized Federated Learning [J]. arXiv:2303.08678, 2023.
- [63] YANG F E, WANG C Y, WANG Y C F. Efficient Model Personalization in Federated Learning via Client-Specific Prompt Generation [J]. arXiv:2308.15367, 2023.
- [64] ZHAO H, DU W, LI F, et al. FedPrompt: Communication-Efficient and Privacy-Preserving Prompt Tuning in Federated Learning[C]//2023 IEEE International Conference on Acoustics, Speech and Signal Processing(ICASSP 2023). 2023;1-5.
- [65] GUO T, GUO S, WANG J, et al. PromptFL: Let Federated Participants Cooperatively Learn Prompts Instead of Models-Federated Learning in Age of Foundation Model [J]. IEEE Transactions on Mobile Computing, 2024,23(5);5179-5194.
- [66] LU W, HU X, WANG J, et al. FedCLIP: Fast Generalization and Personalization for CLIP in Federated Learning [J]. arXiv:

- 2302.13485,2023.
- [67] CHEN H,ZHANG Y,KROMPASS D, et al. FedDAT: An Approach for Foundation Model Finetuning in Multi-Modal Heterogeneous Federated Learning [C] // Proceedings of the AAAI Conference on Artificial Intelligence. 2024;11285-11293.
- [68] JIA J,LI K,MALEK M, et al. Joint Federated Learning and Personalization for on-Device ASR[C] // 2023 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU). 2023;1-8.
- [69] KHALID U,IQBAL H,VAHIDIAN S, et al. CEFHRI: A Communication Efficient Federated Learning Framework for Recognizing Industrial Human-Robot Interaction[C] // IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). 2023;10141-10148.
- [70] KWAK N,KIM T. X-PEFT: eXtremely Parameter-Efficient Fine-Tuning for Extreme Multi-Profile Scenarios [J]. arXiv: 2401.16137,2024.
- [71] CUI X, QIU X, YE Z, et al. Survey of communication overhead of federated learning [J]. Journal of Computer Applications, 2022,42(2):333-342.
- [72] CHO Y J,WANG J,CHIRVOLU T, et al. Communication-Efficient and Model-Heterogeneous Personalized Federated Learning via Clustered Knowledge Transfer [J]. IEEE Journal of Selected Topics in Signal Processing,2023,17(1):234-247.
- [73] NI X,SHEN X,ZHAO H. Federated optimization via knowledge codistillation [J]. Expert Systems with Applications,2022,191:116310.
- [74] MICIELI U,TOLDO M,OZAY M. Federated Learning via Attentive Margin of Semantic Feature Representations [J]. IEEE Internet of Things Journal,2023,10(2):1517-1535.
- [75] MAO Y,ZHAO Z,YAN G, et al. Communication-efficient federated learning with adaptive quantization [J]. ACM Transactions on Intelligent Systems and technology,2022,13(4):1-26.
- [76] BASTIANELLO N,LIU C,JOHANSSON K H. Enhancing Privacy in Federated Learning through Local Training [J]. arXiv: 2403.17572,2024.
- [77] AYACHE G,DASSARI V,ROUAYHEB S E. Walk for Learning: A Random Walk Approach for Federated Learning From Heterogeneous Data [J]. IEEE Journal on Selected Areas in Communications,2023,41(4):929-940.
- [78] YE M,FANG X,DU B, et al. Heterogeneous Federated Learning: State-of-the-art and Research Challenges [J]. ACM Computing Surveys,2023,56(3):79.
- [79] CHAI Z,FAYYAZ H,FAYYAZ Z, et al. Towards taming the resource and data heterogeneity in federated learning[C] // 2019 USENIX Conference on Operational Machine Learning (OpML 19). 2019;19-21.
- [80] WU W T,WU Y L,LIN W W, et al. Horizontal Federated Learning: Research Status, System Applications and Open Challenges [J]. Chinese Journal of Computers,2025,48(1):35-67.
- [81] XIAO Z,CHEN Z,LIU S, et al. Fed-GraB: Federated long-tailed learning with self-adjusting gradient balancer [J]. arXiv: 2310.07587,2023.
- [82] WU D,ULLAH R,HARVEY P, et al. FedAdapt: Adaptive Offloading for IoT Devices in Federated Learning [J]. IEEE Internet of Things Journal,2022,9(21):20889-20901.
- [83] CHUNG W C,CHANG Y C,HSU C H, et al. Federated feature concatenate method for heterogeneous computing in Federated Learning [J]. Computers, Materials Continua,2023,75(1):351-370.
- [84] ALIZADEH K,MIRZADEH I,BELENKO D, et al. LLM in a flash: Efficient Large Language Model Inference with Limited Memory [J]. arXiv:2312.11514,2023.
- [85] ZHU L,LIU Z,HAN S. Deep leakage from gradients [C] // Proceedings of the 33rd International Conference on Neural Information Processing Systems. 2019;14774-14784.
- [86] PAN X,ZHANG M,JI S, et al. Privacy Risks of General-Purpose Language Models[C] // IEEE Symposium on Security and Privacy(SP). 2020;1314-1331.
- [87] GU Y H,BAI Y B. Survey on Security and Privacy of Federated Learning Models [J]. Ruan Jian Xue Bao/Journal of Software, 2023,34(6):2833-2864.
- [88] TANG L T,CHEN Z N,ZHANG L F, et al. Research Progress of Privacy Issues in Federated Learning [J]. Ruan Jian Xue Bao/Journal of Software,2023,34(1):197-229.
- [89] TOBABEN M,SHYSHEYA A,BRONSKILL J, et al. On the Efficacy of Differentially Private Few-shot Image Classification [J]. arXiv:2302.01190,2023.
- [90] XU R,BARACALDO N,ZHOU Y, et al. HybridAlpha: An Efficient Approach for Privacy-Preserving Federated Learning[C] // Proceedings of the 12th ACM Workshop on Artificial Intelligence and Security. ACM,2019;13-23.
- [91] HE J,ZHOU C,MA X, et al. Towards a unified view of parameter-efficient transfer learning [J]. arXiv:2110.04366,2021.
- [92] DING N,QIN Y,YANG G, et al. Parameter-efficient fine-tuning of large-scale pre-trained language models [J]. Nature Machine Intelligence,2023,5(3):220-235.
- [93] HAN X,ZHANG Z,DING N, et al. Pre-trained models: Past, present and future [J]. AI Open,2021,2:225-250.
- [94] CHEN S,LI B. Towards Optimal Multi-Modal Federated Learning on Non-IID Data with Hierarchical Gradient Blending [C] // IEEE INFOCOM 2022 - IEEE Conference on Computer Communications. 2022;1469-1478.
- [95] LIN Y M,GAO Y,GONG M G, et al. Federated Learning on Multimodal Data: A Comprehensive Survey [J]. Machine Intelligence Research,2023,20(4):539-553.



WANG Xin, born in 1984, Ph.D, associate professor, master supervisor, is a member of CCF (No. 11687M). His main research interests include machine learning, big data analysis and federated learning.