DOI:10.13878/j.cnki.jnuist.20230801001

陈学斌[1]　任志强[1]

PFKD:综合考虑数据异构和模型异构的个性化联邦学习框架

摘要

联邦学习是解决机器学习中数据共享和隐私保护两个关键难题的重要方法.然而,联邦学习本身也面临着数据异构和模型异构的挑战.现有研究往往只专注于解决其中一个方面的问题,忽视了两者之间的关联性.为此,本文提出了一个名为 PFKD 的框架,该框架通过知识蒸馏技术解决模型异构问题,通过个性化算法解决数据异构问题,以实现更具个性化的联邦学习.通过实验分析验证了所提出框架的有效性.实验结果显示,该框架能够突破模型的性能瓶颈,提高模型精度约 1 个百分点.此外,在调整适当的超参数后,该框架的性能得到进一步提升.

关键词

联邦学习;数据异构;模型异构

中图分类号 TP18;TP309

文献标志码 A

收稿日期 2023-08-01

资助项目 国家自然科学基金(U20A20179)

作者简介

陈学斌,男,博士,教授,研究方向为数据安全和物联网安全. chxb@qq.com

任志强(通信作者),男,硕士生,研究方向为数据安全和隐私保护. psp1274632466@qq.com

1 华北理工大学 理学院/河北省数据科学与应用重点实验室/唐山市数据科学重点实验室,唐山,063210

0 引言

通过让计算机自动学习和改善,机器学习已经广泛应用于各个领域[1].然而,传统的中心化机器学习方法面临数据孤岛[2]和个人隐私保护要求等挑战.为了应对这些问题,联邦学习[3-4]作为一种分布式机器学习技术被提出,它能确保多个参与方协同训练机器学习模型,同时不泄露参与方的数据.凭借在隐私保护方面的优势,联邦学习被广泛应用于政务、医疗、金融和物流等领域.

联邦学习系统主要由客户端和服务端两部分组成,其中,客户端负责模型的训练,服务端负责模型的聚合.在理想的联邦学习场景下,客户端均使用相同的模型进行训练,服务端对模型执行平均聚合[3].然而,在现实情况中,联邦学习除了要面临安全挑战以外[5],还由于客户端之间的数据量、数据分布、算力资源等存在差异,使得联邦学习面临数据异构和模型异构两大挑战.其中,数据异构是客观存在且不可避免的,主要表现为客户端之间的数据分布不同.而模型异构则指不同客户端的训练模型不一致,是人为可控的.

目前,多数研究仅从数据异构的角度出发,以适应本地预测为目标,提出个性化的联邦学习.然而,他们忽略了一个事实:不同客户端之间的数据量和算力资源是不同的,统一的模型不可能适配所有客户端的本地资源.因此,为了使模型更具个性化、更适用于本地预测,不同的客户端应拥有不同的模型,即:对个性化联邦学习的研究不应局限于只考虑客户端之间的数据异构性,还应当关注客户端之间的模型异构性.

针对目前个性化联邦学习研究中的不足之处,本文综合考虑了客户端之间的数据异构性和模型异构性,提出了更能实际应用的个性化联邦框架(Personalized Federated learning based on Knowledge Distillation,PFKD).首先,分析了联邦学习面临的两大挑战,并强调个性化联邦学习研究需要关注客户端之间的模型异构;然后,对现有解决数据异构和模型异构的研究方案进行比较;接着,针对联邦学习中的数据异构和模型异构,提出了 PFKD 框架,以实现完整的个性化联邦学习;最后,通过实验证明了 PFKD 框架在解决联邦学习中的数据异构和模型异构问题上的有效性,并给出了 PFKD 框架的联邦通信次数

建议和框架中超参数 α 的取值建议.

1 相关工作

目前,对个性化联邦学习的研究大多仅考虑了客户端之间的数据异构性.例如,Briggs 等[6]和 Mansour 等[7]利用聚类技术将同类成员共同训练一个模型,以实现个性化模型训练.也有研究者设计了两个模型,既能为全局模型作出贡献,又能获得适应本地数据的模型[8-9].还有研究者将模型分为基础层和个性层[10],其中,基础层通过联邦学习进行协作训练,个性层则只使用本地数据进行训练,以实现个性化的联邦学习.此外,Fallah 等[11]基于模型无关元学习(Model-Agnostic Meta-Learning,MAML)的思想,试图先协作训练出一个适合所有用户的初始模型,每个客户端再利用本地数据训练此模型,从而实现个性化的联邦学习.然而,以上研究的不足之处在于没有考虑客户端之间的模型异构,从而不能充分利用客户端本地资源,导致个性化效果并不彻底.

在联邦学习中,模型异构是一个十分严峻的挑战.解决这一问题最常见的处理技术是知识蒸馏(Knowledge Distillation,KD)[12].然而,目前将知识蒸馏应用于联邦学习时需要解决两个关键问题:1) 如何选择教师模型;2) 如何解决相同数据集的问题.尤其是第二个问题决定了 KD 方法能否成功应用于联邦学习.本文将沿用文献[13]中的术语,将归一化后的输出预测分布称为"logit vector".已有研究提出了多种方法将 KD 应用于联邦学习,以解决模型异构问题.例如:Jeong 等[13]采用平均 logit vector 的方法来解决相同数据集的问题;Li 等[14]假设存在一个公共数据集来解决相同数据集的问题;Sattler 等[15]和 Gong 等[16]提出了使用辅助数据集的假设.值得一提的是,Shen 等[17]在客户端多设计了一个用于通信的模型,从而避免了相同数据集的问题.具体而言,他们在客户端本地设计了两种模型:私有模型和公有模型.其中,私有模型保留在客户端本地,而公有模型用于实现信息交互.每次更新时,使用 DML[18]同时更新公有模型和私有模型,以完成知识交互.另外,Zhu 等[19]提出了 FEDGEN 方法来解决相同数据集的问题,其核心是学习一个 logit vector 生成器.

综上所述,单方面解决联邦学习中的数据异构或模型异构已经成为可能.因此,本文将综合考虑联邦学习中的数据异构和模型异构,首先利用 KD 技术来解决模型异构,然后根据 KD 的特性改进现有解决数据异构的算法,以实现完整的个性化联邦学习.

2 方法:PFKD

2.1 PFKD 框架概述

PFKD 框架(图 1)通过两个独立的组件来分别解决模型异构和数据异构,从而实现个性化联邦学习.第一个组件的核心是不同功能的模型设计和知识蒸馏技术,该组件将解决联邦学习中的模型异构;第二个组件的核心是个性化算法设计,该组件将解决联邦学习中的数据异构.

在 PFKD 框架中,存在三类模型,分别是个性化模型(P-Model)、知识交流模型(C-Model)和元模型(M-Model).不同类别的模型之间通过知识蒸馏实现知识的传递.个性化算法将为不同的客户端返回合适的 M-Model,用于传递知识并指导本地的 P-Model 更新.

2.2 知识蒸馏设计

本文将知识蒸馏技术中学生模型的目标函数设计为:$L=(1-\alpha)L_{ce}(y,p)+\alpha D_{KL}(q,p)\cdot T^2$.其中:$T$ 表示超参数"温度";α 表示权重系数;L_{ce} 表示交叉熵损失;D_{KL} 表示相对熵(KL 散度);p 表示学生模型的输出(由 0 和 1 组成的向量,也称为 hard logit vector);q 为教师模型的输出(也称为 soft logit vector).

通过上述设计,客户端的模型训练目标可被简要概括为:在最小化学生模型损失、最小化学生模型与教师模型的差距损失之间寻求平衡.这种平衡可以通过超参数 α 的调整来自由控制,从而有效地调节两种损失的权重比例.

在本文提出的 PFKD 框架中,当采用知识蒸馏技术进行训练时,将作为学生模型训练的目标函数修改为上述的目标函数.例如,在利用服务端发送的 M-Model 来更新本地的 P-Model 时,客户端会更改本地模型的训练目标为上述的目标函数.此时,P-Model 充当学生模型的角色,而 M-Model 则充当教师模型的角色,在经过一系列本地数据集上的迭代训练后,P-Model 逐渐在学生模型损失与模型输出差距损失之间找到了一个平衡点,从而实现知识的传递.类似地,当 P-Model 用于指导 C-Model 的训练时,也是经过类似的过程,最终达到相似的平衡状态.

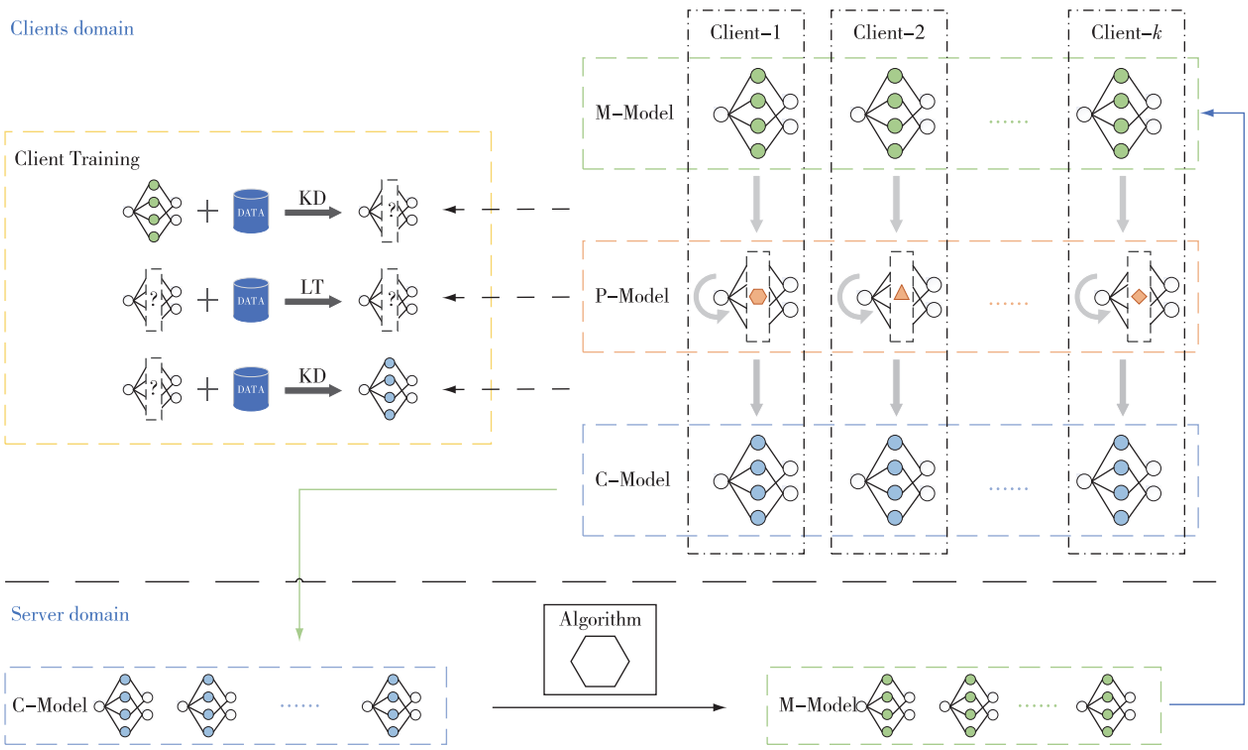

图 1　PFKD 框架,LT 指本地训练

Fig. 1　PFKD framework, in which LT represents local training

2.3　PFKD 模型设计

在 PFKD 框架中,为不同客户端根据本地资源拥有适配的个性化模型(对应 P-Model),也为实现与其他客户端模型的联邦学习,本文引入一个统一的模型(对应 C-Model).此时,服务端可以在相同结构的模型集(对应 C-Model)上执行个性化算法,为每个客户端返回独特的模型(对应 M-Model).因此,客户端将拥有三类模型:P-Model、C-Model 和 M-Model,而服务端将拥有两类模型:C-Model 和 M-Model.以下是三类模型的详细介绍.

1)P-Model:通过客户端本地数据训练得到的模型,适用于客户端的本地数据,为客户端所私有.

2)C-Model:C-Model 是通过使用知识蒸馏(KD)技术,以 P-Model 作为教师模型进行训练得到的模型.C-Model 的设计是为了在不同客户端之间实现知识交流,从而促进模型间的信息传递.所有客户端的 C-Model 结构相同.

3)M-Model:M-Model 是通过个性化算法聚合得到的,由服务端返回给不同客户端的模型.M-Model 的结构与 C-Model 相同,但其内容是根据个性化算法对每个客户端的需求进行调整而得到的.在客户端更新自己的 P-Model 时,M-Model 起着指导和辅助的作用,从而提高本地模型的性能.

2.4　PFKD 个性化算法设计

在 PFKD 框架中,本文利用 C-Model 来有效地解决联邦学习中的数据异构问题.由于 C-Model 在所有客户端中保持一致,并且已有丰富的算法可供借鉴,本文在现有方法的基础上做了一些微创新,以更有效地解决数据异构的挑战.例如,采纳 Briggs 等[6]提出的聚类算法作为基础,将客户端上传的模型分为多个组.然后,将同一组的成员共同训练得到 M-Model,有效地克服了数据异构的问题.在实验部分采用该算法,获得了令人满意的分组结果.

值得强调的是,本文提出的 PFKD 框架利用知识蒸馏技术在模型之间传递知识,而 KD 技术的有效性与教师模型输出的可信度相关.根据实践经验,本文将模型的精确度作为评估教师模型可信度的指标.当服务端所聚合模型的精度越高时,知识传递效果就越明显.因此,在基于上述算法得到的分组基础上,本文引入一种选择策略.具体步骤如下:

1) 对同一组内的所有模型在私有数据集上进行测试,评估其模型精度;

2) 基于精度值,计算前 30%模型精度的平均值,并将其减少 5%(具体减少量可根据经验设定)

作为合格阈值;

3) 将精度高于合格阈值的模型进行平均聚合[3],形成同一组的 M-Model.

除此之外,如果不存在隐私威胁,服务端可以将最终聚合阶段的模型集返回给客户端.客户端可以自主选择哪些模型用于聚合,以生成最终的 M-Model.在确保安全性的前提下,该流程允许客户端参与模型的选择过程,进一步增强了联邦学习的协作性,但仅适合特定场合.

2.5 PFKD 框架流程

PFKD 框架的流程如下:

1) 客户端利用本地数据集训练得到 P-Model;

2) 客户端使用 KD 技术,将 P-Model 作为教师模型,使用本地数据集训练 C-Model;

3) 客户端上传 C-Model 至服务端;

4) 服务端使用个性化算法聚合得到 M-Model;

5) 服务端下发 M-Model 至客户端;

6) 客户端使用 KD 技术,将 M-Model 作为教师模型,使用本地数据集更新 P-Model.

在不同的联邦学习场景[20]中,可以灵活地应用以上流程,从而适应不同的需求.例如,在 Cross-Device 联邦学习场景下,可以重复步骤 2)—6),进行多次联邦通信,加强模型更新和知识传递效果.而在 Cross-Silo 联邦学习场景下,一次联邦通信即可满足实验需求,避免不必要的通信开销.

综上所述,PFKD 框架通过两个分离的组件解决了客户端之间的模型异构和数据异构,从而实现个性化联邦学习.它引入知识蒸馏技术和个性化算法,以实现在联邦学习中的知识传递和模型更新.通过 P-Model、C-Model 和 M-Model 之间的知识交流,客户端能够提高本地模型的精度,而个性化算法则能够根据不同客户端的需求返回合适的 M-Model.PFKD 框架的流程十分简洁,允许在不同联邦学习场景中重复执行以提升 P-Model 的准确性.

3 实验设计和内容

1)硬件和软件:本文的实验在 CentOS Linux 系统上执行,使用的 CPU 配置为 28 个 Intel(R) Xeon(R) CPU E5-2680 v4 @ 2.40 GHz,GPU 配置为 Tesla P40.实验代码采用 Python 3.8 编写,联邦学习的模拟使用 PyTorch 1.13.0 构建.

2)异构模型设置:为了模拟联邦学习中不同客户端的异构模型,本实验选择 4 种不同结构的深度神经网络(DNN)模型.第 1 个模型结构被选为基础结构,第 2 个模型在基础结构上增加深度,第 3 个模型在基础结构上增加宽度,第 4 个模型在基础结构上同时增加深度和宽度.4 个模型的结构分别为:784×360×180×10、784×360×240×180×10、784×500×180×10 和 784×500×360×180×10.除输出层外,中间层都采用 ReLU 激活函数.此外,在本实验中,C-Model 和 M-Model 的结构与第 1 个模型相同.模型使用交叉熵损失函数(Cross Entropy Loss)进行训练,采用随机梯度下降(SGD)进行优化,学习率为 0.1.

3)联邦学习设置:在实验中已经使用文献[6]中的算法成功进行了个性化分组.基于此前提,实验设置了 4 个客户端和 1 个聚合服务端,这 4 个客户端属于同一组成员.除非另有说明,客户端默认在本地数据集上进行 200 轮的训练,包括 P-Model、C-Model 的本地训练以及更新 P-Model 的本地训练.联邦通信的轮数为 1.

4)数据集和划分:本文采用流行的 Fashion-MNIST 数据集(https://github.com/zalandoresearch/fashion-mnist).该数据集包含 70 000 张衣物图像,其中 60 000 张用于训练,10 000 张用于测试.每个图像的大小为 28×28 像素,与 10 个类别的标签相关联.总数据集被划分为 150 个大小为 400 的块,除非另有说明,每个参与方拥有一个包含不同标签的块,数据总量为 4 000.

5)评价指标:本文使用准确率作为评价 PFKD 框架有效性的指标.

6)实验内容:将仅在本地训练的模型的精度变化作为基准;比较在使用 PFKD 框架后模型精度的变化情况;在 PFKD 框架下,将比较一次联邦通信和多次联邦通信对模型精度的影响;比较不同精度的 M-Model 对 P-Model 精度提升的程度,以及 KD 中的超参数 α 对 P-Model 精度的影响.

4 实验结果与分析

4.1 PFKD 框架的有效性分析

PFKD 框架的有效性依赖于模型间知识传递的有效性,而本文采用的核心技术是知识蒸馏.因此,需要验证使用 KD 在本文提出的三类模型间传递知识的有效性.

从图 2 可以观察到,KD 能够有效地将 P-Model 的知识传递给 C-Model,并且相比于 P-Model,C-Model 的精度波动更小.图 3 表示在 200 轮本地迭代

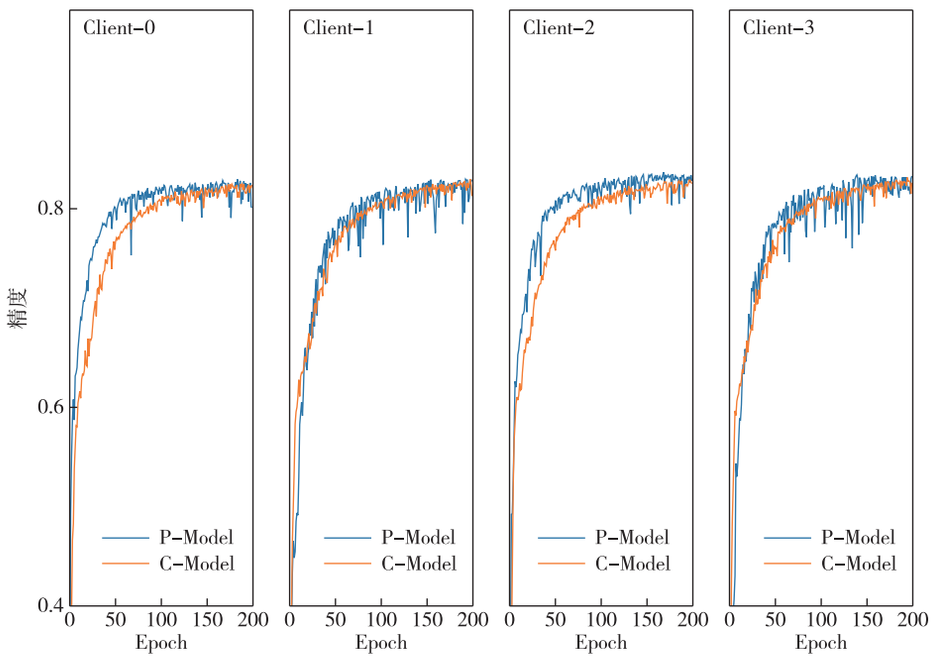

图 2　P-Model 和利用 P-Model 训练的 C-Model 模型精度随本地迭代的变化情况

Fig. 2　Variations of accuracy of P-Model and C-Model trained by P-Model with local iteration

后,使用 PFKD 框架和仅进行本地训练的模型精度随迭代次数的变化情况.其中:LT-1 表示相同的 200 轮本地训练;LT-2 表示在 LT-1 的基础上,保持本地训练的结果;PFKD 表示在 LT-1 的基础上,使用 PFKD 框架的结果;Client-i(i=0,1,2,3)表示第 i 个客户端的模型精度.由图 3 可以看出,在 M-Model 作为教师模型指导 P-Model 进行二次更新后,P-Model 的精度突破了模型精度的瓶颈,将模型精度提升了约 1 个百分点(表 1,其中:Muti-PFKD 表示执行了多次联邦通信;加粗数据表示 P-Model 的平均精度最

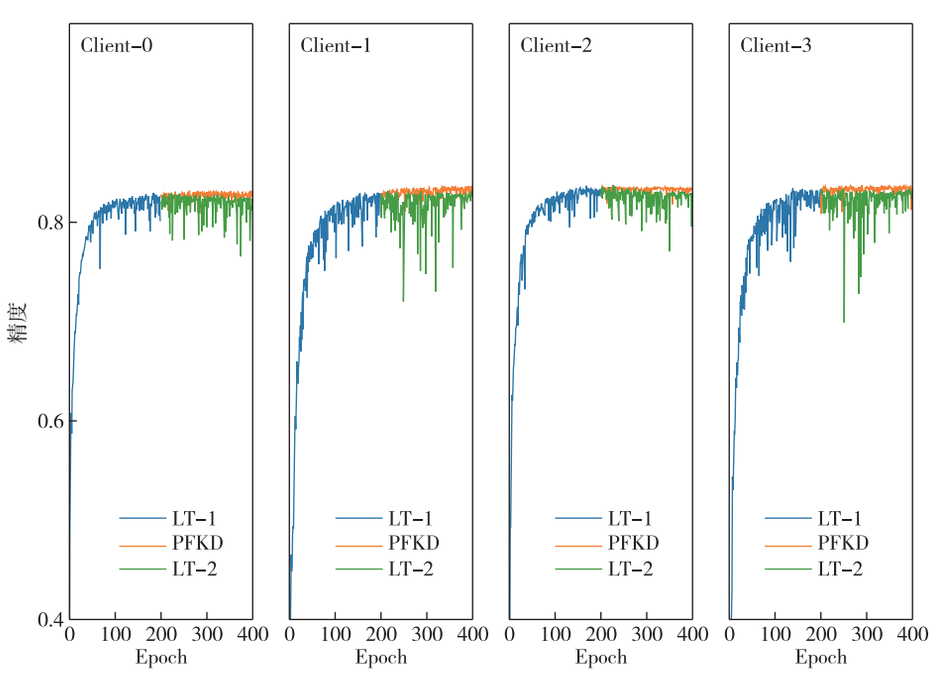

图 3　在 200 轮本地迭代后,使用 PFKD 框架和仅进行本地训练的模型精度随迭代次数的变化情况

Fig. 3　After 200 rounds of local iteration, accuracy of the model using PFKD framework and only local training changing with number of iterations

高值).同时,使用 PFKD 框架还减少了模型性能的波动.另外,通过比较使用 PFKD 框架和仅进行本地训练的模型精度变化,可以观察到该框架能够防止模型的过拟合.

表 1　客户端使用 PFKD 框架或仅进行本地训练下,P-Model 的平均精度(取后 30 轮的平均值)

Table 1　Average accuracy of P-Model (averaged over the last 30 rounds),when clients utilize PFKD framework or only conduct local training　%

训练方式	Client-1	Client-2	Client-3	Client-4
LT-1	81.75	82.03	82.87	82.38
LT-2	81.78	82.18	82.10	82.24
PFKD	82.78	83.19	**83.37**	83.36
Muti-PFKD	**82.86**	**83.58**	83.17	**83.38**

4.2　PFKD 联邦通信次数分析

本节将分析在使用 PFKD 框架的情况下,一次联邦通信和多次联邦通信对模型精度的影响.本文设置 PFKD 框架的多次联邦学习是指在一次完整的联邦通信后,将客户端中 M-Model 指导 P-Model 和 P-Model 指导 C-Model 的训练轮数从 200 轮降低到 20 轮,总共进行 20 次联邦通信.

从表 1 中可以观察到,由于客户端之间的 P-Model 精度差距不大,它们之间的 C-Model 也不会存在较大的精度差异,因此,服务端返回的 M-Model 并不能显著提升 P-Model 的精度.即使用 PFKD 框架进行多次联邦通信对模型精度的提升并不是特别显著.结合第 4.3 节的分析可知,当服务端的个性化算法使客户端组合固定时,PFKD 更适合进行一次联邦通信,只需要在这一次通信中获得更高精度的 M-Model 即可.而当服务端的个性化算法使客户端组合多变时,PFKD 更适合进行多次联邦通信.

4.3　M-Model 精度与超参数的影响分析

本节分析不同精度的 M-Model 以及 KD 中的 α 对 P-Model 精度的影响.为了模拟此场景,首先通过增加客户端 1 和客户端 2 的本地数据量来实现不同的 C-Model 精度,然后通过不同的聚合算法生成不同精度的 M-Model,以分析不同精度的 M-Model 对 P-Model 的影响.接着,使用不同的 α 来分析其对 P-Model 的影响.结果如表 2 所示.其中:M-Acc 表示 M-Model 的模型精度;α 为 KD 中的超参数;加粗数据表示不同精度的 M-Model 与不同 α 下的最高精度.

表 2　不同精度的 M-Model 与不同的 α 对 P-Model 精度的影响

Table 2　Influence of M-Model with different precision and α on P-Model accuracy　%

训练方式	M-Acc	α	Client-1	Client-2	Client-3	Client-4
LT-1		0.5	82.70	87.84	88.34	81.56
LT-2		0.5	82.70	87.98	88.39	80.82
PFKD	84.99	0.5	83.44	88.32	88.26	82.67
	88.28	0.5	84.15	88.99	**89.06**	83.60
	88.28	0.7	**84.76**	**89.07**	89.03	**84.88**

从表 2 中可以观察到:当 M-Model 的精度高于 P-Model 的精度时,可以明显提高 P-Model 的精度;而当 M-Model 的精度低于 P-Model 的精度时,更多地表现为防止模型过拟合的功能.超参数 α 的取值决定了 M-Model 对 P-Model 的影响程度.当 M-Model 的精度高于 P-Model 时,较大的超参数值能更大程度地提高模型精度;当 M-Model 的精度低于 P-Model 时,较小的超参数值能更大程度地提高模型精度.因此,在 PFKD 框架中,根据 M-Model 精度与 P-Model 精度的差异,选择适当的 α 值能更大程度地提升 P-Model 的精度.

综上所述,通过模拟实验,本文证明了 PFKD 的有效性.根据服务端个性化算法得到的结果,本文给出了 PFKD 联邦通信次数的建议.最后,本文分析了不同精度的 M-Model 与不同的超参数 α 对 P-Model 精度的影响.

5　结语

本文在联邦学习领域中提出了 PFKD 框架,该框架综合考虑了数据异构和模型异构的问题,为实现个性化联邦学习提供了有效的解决方案.实验结果验证了该框架的优越性,并为未来的研究提供了指导性的建议.未来可以进一步扩展和改进 PFKD 框架,以适应不断变化的数据异构性和模型异构性挑战,推动联邦学习在更广泛领域的应用和发展.

参考文献

References

[1] Jordan M I,Mitchell T M.Machine learning:trends,perspectives,and prospects[J].Science,2015,349(6245):255-260

[2] Zhang C,Xie Y,Bai H,et al.A survey on federated learning[J].Knowledge-Based Systems,2021,216:106775

[3] McMahan B,Moore E,Ramage D,et al.Communication-efficient learning of deep networks from decentralized

data[C]//Proceedigns of the 20th International Conference on Artificial Intelligence and Statistics. PMLR, 2017:1273-1282
[4] Yang Q,Liu Y,Chen T J,et al.Federated machine learning:concept and applications[J].ACM Transactions on Intelligent Systems and Technology,2019,10(2):1-19
[5] Zhai R,Chen X B,Pei L T,et al.A federated learning framework against data poisoning attacks on the basis of the genetic algorithm[J].Electronics,2023,12(3):560
[6] Briggs C,Fan Z,Andras P.Federated learning with hierarchical clustering of local updates to improve training on non-IID data[C]//2020 International Joint Conference on Neural Networks (IJCNN). July 19 - 24, 2020, Glasgow,UK.IEEE,2020:1-9
[7] Mansour Y,Mohri M,Ro J,et al.Three approaches for personalization with applications to federated learning[J].arXiv e-Print,2020,arXiv:2002.10619
[8] Deng Y Y, Kamani M M, Mahdavi M. Adaptive personalized federated learning[J].arXiv e-Print,2020, arXiv:2003.13461
[9] Pillutla K,Malik K,Mohamed A R,et al.Federated learning with partial model personalization[C]//Proceedings of the 39th International Conference on Machine Learning.PMLR,2022:17716-17758
[10] Arivazhagan M G, Aggarwal V, Singh A K, et al. Federated learning with personalization layers[J].arXiv e-Print,2019,arXiv:1912.00818
[11] Fallah A,Mokhtari A,Ozdaglar A.Personalized federated learning:a meta-learning approach[J].arXiv e-Print, 2020,arXiv:2002.07948
[12] Hinton G,Vinyals O,Dean J.Distilling the knowledge in a neural network[J].arXiv e-Print, 2015, arXiv:1503.02531
[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 e-Print, 2018,arXiv:1811.11479
[14] Li D L, Wang J P. FedMD: heterogenous federated learning via model distillation[J].arXiv e-Print,2049, arXiv:1910.03581
[15] Sattler F, Korjakow T, Rischke R, et al. FedAUX: leveraging unlabeled auxiliary data in federated learning[J].IEEE Transactions on Neural Networks and Learning Systems,2023,34(9):5531-5543
[16] Gong X A, Sharma A, Karanam S, et al. Preserving privacy in federated learning with ensemble cross-domain knowledge distillation[J].Proceedings of the AAAI Conference on Artificial Intelligence, 2022, 36(11): 11891-11899
[17] Shen T,Zhang J,Jia X,et al.Federated mutual learning[J].arXiv e-Print,2020,arXiv:2006.16765
[18] Zhang Y,Xiang T,Hospedales T M,et al.Deep mutual learning[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition.June 18-23,2018,Salt Lake City,UT,USA.IEEE,2018:4320-4328
[19] Zhu Z D,Hong J Y,Zhou J Y.Data-free knowledge distillation for heterogeneous federated learning[J]. Proceedings of Machine Learning Research,2021,139: 12878-12889
[20] 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

PFKD:a personalized federated learning framework that integrates data heterogeneity and model heterogeneity

CHEN Xuebin[1] REN Zhiqiang[1]
1 College of Science/Hebei Key Laboratory of Data Science and Application/Tangshan Key Laboratory of Data Science,North China University of Science and Technology,Tangshan 063210,China

Abstract Federated learning is an important method to address two critical challenges in machine learning:data sharing and privacy protection.However,federated learning itself faces challenges related to data heterogeneity and model heterogeneity.Existing researches often focus on addressing one of these issues while overlook the correlation between them.To address this,this paper introduces a framework named PFKD (Personalized Federated learning based on Knowledge Distillation).This framework utilizes knowledge distillation techniques to address model heterogeneity and personalized algorithms to tackle data heterogeneity,thereby achieving more personalized federated learning.Experimental analysis validates the effectiveness of the proposed framework.The experimental results demonstrate that the framework can overcome model performance bottlenecks and improve model accuracy by approximately one percentage point.Furthermore,with appropriate hyperparameter adjustment,the framework's performance is further enhanced.

Key words federated learning;data heterogeneity;model heterogeneity