中图法分类号:TP309; TP399　　文献标识码: A　　文章编号: 1006-8961(2024)07-1849-12

论文引用格式:Huang W K, Ye M and Du B. 2024. Adaptive heterogeneous federated learning. Journal of Image and Graphics, 29(07):1849-1860
(黄文柯, 叶茫, 杜博. 2024. 自适应异构联邦学习. 中国图象图形学报, 29(07):1849-1860)[DOI:10.11834/jig.230239]

自适应异构联邦学习

黄文柯,叶茫*,杜博

武汉大学计算机学院,武汉　430072

摘　要:目的　模型异构联邦学习由于允许参与者在不损害隐私的情况下独立设计其独特模型而受到越来越多的关注。现有的方法通常依赖于公共共享的相关数据或全局模型进行通信,极大地限制了适用性。且每个参与者的私有数据通常以不同的分布收集,导致数据异构问题。为了同时处理模型异构和数据异构,本文提出了一种新颖的自适应异构联邦学习方法。**方法**　给定一个随机生成的输入信号(例如,随机噪声),自适应异构联邦学习直接通过对齐输出逻辑层分布来实现异构模型之间的通信,实现协作知识共享。主要优势是在不依赖额外相关数据收集或共享模型设计的情况下解决了模型异构问题。为了进一步解决数据异构问题,本文提出了在模型和样本层面上进行自适应权重更新。因此,自适应异构联邦学习(adaptive heteogeneous federated learning,AHF)允许参与者通过模型输出在无关数据上的差异和强调"有意义"的样本来学习丰富多样的知识。**结果**　通过在不同的联邦学习任务上使用随机噪声输入进行通信,进行了广泛的实验,显示出比竞争方法更高的域内精确度和更好的跨域泛化性能。**结论**　本文方法提供了一个简单而有效的基准,为异构联邦学习的未来发展奠定基础。

关键词: 联邦学习(FL);模型异构;数据异构;随机噪声;异构联邦学习

Adaptive heterogeneous federated learning

Huang Wenke, Ye Mang*, Du Bo

School of Computer Science, Wuhan University, Wuhan　430072, China

Abstract: Objective The current development of deep learning has caused significant changes in numerous research fields and has had profound impacts on every aspect of societal and industrial sectors, including computer vision, natural language processing, multi-modal learning, and medical analysis. The success of deep learning heavily relies on large-scale data. However, the public and scientific communities have become increasingly aware of the need for data privacy. In the real world, data are commonly distributed among different entities such as edge devices and companies. With the increasing emphasis on data sensitivity, strict legislation has been proposed to govern data collection and utilization. Thus, the traditional centralized training model, which requires data aggregation, is unusable in the practical setting. In response to such real-world challenges, federated learning (FL) has emerged as a popular research field because it can train a global model for different participants without centralizing data owned by the distributed parties. FL is a privacy-preserving multi-party collaboration model that adheres to privacy protocols without data leakage. Typically, FL requires clients to share a global model architecture for the central server to aggregate parameters from participants and then redistributes the global

收稿日期:2023-05-08;修回日期:2023-12-15;预印本日期:2023-12-22

*通信作者:叶茫　yemang@whu.edu.cn

基金项目:国家重点研发计划项目(2023YFC2705700);国家自然科学基金项目(62361166629, 62176188, 62225113, 623B2080)

Supported by: National Key R&D Program of China (2023YFC2705700); National Natural Science Foundation of China (62361166629, 62176188, 62225113, 623B2080)

model (averaged parameters). However, this prerequisite largely restricts the flexibility of the client model architecture. In recent years, the concept of objective model heterogeneous FL has garnered substantial attention because it allows participants to independently design unique models in FL without compromising privacy. Specifically, participants may need to design special model architecture to ease the communication burden or refuse to share the same architecture due to intellectual property concerns. However, existing methods often rely on publicly shared related data or a global model for communication, limiting their applicability. In addition, FL is proposed to handle privacy concerns in the distributed learning environment. A pioneering FL method trains a global model by aggregating local model parameters. However, its performance is impeded due to decentralized data, which results in non-i.i.d distribution (called data heterogeneity). Each participant optimizes toward the local empirical risk minimum, which is inconsistent with the global direction. Therefore, the average global model has a slow convergence speed and achieves limited performance improvement. **Method** Model heterogeneity largely impedes the local model section flexibility, and data heterogeneity hinders federated performance. To address model and data heterogeneity, this paper introduces a groundbreaking approach called adaptive heterogeneous federated (AHF) learning, which employs a unique strategy by utilizing a randomly generated input signal, such as random noise and public unrelated samples, to facilitate direct communication among heterogeneous model architectures. This task is achieved by aligning the output logit distributions, fostering collaborative knowledge sharing among participants. The primary advantage of AHF is its ability to address model heterogeneity without depending on additional related data collection or shared model design. To further enhance AHF's effectiveness in handling data heterogeneity, the paper proposes adaptive weight updating on both model and sample levels, which enables AHF participants to acquire rich and diverse knowledge by leveraging dissimilarities in model output on unrelated data while emphasizing the importance of meaningful samples. **Result** Empirical validation of the proposed AHF method is conducted through a meticulous series of extensive empirical experiments. Random noise inputs are employed in the context of two distinct federated learning tasks: Digits and Office-Caltech scenarios. Specifically, our solution presents the stable generalization performance on the more challenging scenario, Office-Caltech. Notably, when a larger domain gap exists among private data, AHF achieves higher overall generalization performance on these different unrelated data samples and obtains stable improvements on most unseen private data. By contrast, competing methods achieve limited generalization performance in the Office-Caltech scenario. The empirical findings validate our solution's ability, showcasing a marked improvement in within-domain accuracy and demonstrating superior cross-domain generalization performance compared with existing methodologies. **Conclusion** In summary, the AHF learning method, as extensively examined in this thorough investigation, not only presents a straightforward yet remarkably efficient foundation for future progress in the domain of federated learning but also emerges as a transformative paradigm in comprehensively addressing model and data heterogeneity. AHF not only lays the groundwork for more resilient and adaptable FL models but also serves as a guide for the transformation of collaborative knowledge sharing in the upcoming era of machine learning. Studying AHF is more than an exploration of an innovative FL methodology; it provides numerous opportunities that arise given the complexities of model and data heterogeneity in the development of machine learning models.

Key words: federated learning (FL); model heterogeneity; data heterogeneity; random noise; heterogeneous federal learning

0 引 言

联邦学习(federated learning,FL)已成为重要的机器学习范式,在该范式中,客户端联合参与通过各种通信手段协同学习全局模型(Yang等,2019)。在联邦学习中,参与的客户端通过平均模型参数来共同训练共享模型(McMahan等,2017)。FL一直是一个活跃和具有挑战性的话题,它涉及到安全通信协议和支持隐私敏感应用,例如移动键盘预测(Hard等,2018)、医疗保健(Ju等,2020)和幽默识别(Guo等,2020)。

然而,在现实场景中,FL研究面临着一些关键挑战(Lid等,2020)。流行的方法,例如联邦平均及其后续优化策略(Li和Wang,2019),主要基于聚合本地学习的模型参数。假设参与者具有不同的设计

规范和不同的硬件能力，需要构建定制化模型，从而导致了模型的异构性（Shen 和 Lyu，2020），其中模型平均技术无法解决这一挑战。为了处理模型异构性，现有的策略通常依赖于两种额外的信息：公共共享相关数据（Li 和 Wang，2019；Sun 等，2020）和共享全局模型进行合作训练（Shen 等，2020；Liang 等，2019）。然而，前者需要相关数据，这意味着服务器需要收集具有类似于私有本地参与者的分布的额外公共数据，以便捕获相应的先验知识。这在许多实际场景中很难满足，限制了该方法的适用性。对于共享全局模型的策略，不可避免地增加了计算成本，并需要额外的模型结构设计。这促使本文探索一种可行的解决方案来处理模型异构性，而不依赖相关数据或共享全局模型。在 FL 中，不同参与方之间的数据通常也高度异构。考虑到参与方的数据是以非独立同分布的方式生成和收集的（Kairouz 等，2019），因此参与方的数据之间存在领域差异。尽管已有方法提出了集合蒸馏的模型融合方法，从一定程度上消除了限制，但是在考虑明显的参与方数据领域差异时，它可能只能获得有限的性能提升。目前针对数据异构的联邦方法（Li 等，2018；Jiang 等，2019）直接基于参与方模型参数的平均值，这意味着这些方法不适用于模型异构的联邦学习。因此，不可避免的数据异构进一步增加了异构联邦学习的难度。

本文在图 1 进一步解释了这两个问题：模型异构和数据异构。模型异构性：参与者独立设计模型，导致模型平均策略不可行；数据异构性：私有数据存在领域差异，这导致了联邦学习过程中模型性能的降低。

因此，由参与者设计的独特模型和具有明显域差异的分布式数据同时带来了模型和数据异构的挑战。在本文中，提出了自适应异构联邦学习（adaptive heterogeneous fedearated learning，AHF）来应对上述挑战。

为了解决模型异构性而不依赖相关数据，本文建议通过利用不相关数据（例如随机噪声、网络图像等）来对齐参与者模型的 logits 输出分布。通过在不相关数据上对齐分布，AHF 不需要参与者模型共享相关数据标签或依赖于共享的全局模型。此外，相关数据通常难以收集，而不相关数据易于获取，例如随机噪声和网络图像。虽然通过随机噪声进行学习

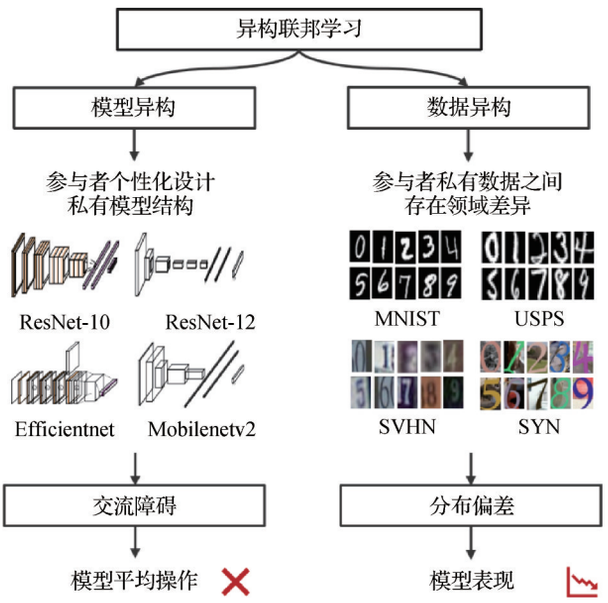

图 1　异构联邦学习

Fig. 1　Heterogeneous federal learning

是一项相当具有挑战性的任务，但 AHF 表现出了卓越的鲁棒性，并在随机噪声上获得了令人满意的性能。具体而言，AHF 通过测量不相关数据（如 Kullback-Leibler 散度）上的 logits 输出分布差异来更新参与者模型。为了处理数据异构性，AHF 希望模型能够更多地从具有更大域差异的私有数据中学习，从而使模型学习到更多多样化的知识并具有更好的泛化性能。因此，AHF 测量参与者之间的差异，并调整每个参与者接收到的知识分布。此外，由于样本的多样性，参与者对不同样本的响应是多样的。通过强调“有意义”的样本并忽略“无意义”的样本，可以使训练过程更加平滑。本文通过全面的实证结果验证了 AHF 的有效性，表明其显著优于以前相关的联邦学习方法。

本文的主要贡献如下：1）提出了一种强健稳定的自适应异构联邦学习方法，同时解决了 FL 中的模型异构性和数据异构性，当与随机噪声输入通信时，表现出了很有潜力的性能。2）制定了一种有效的通信策略，以解决联邦学习中的模型异构性。提出方法通过对不相关数据（如随机噪声和网络图像）的输出分布进行对齐来更新参与者模型。其主要优点是在处理模型异构联邦学习时不依赖于公共共享相关数据和共享全局模型。它为 FL 未来的发展提供了具有启发性的指导。3）探索了异构模型和样本级别的自适应权重更新策略。它自适应地捕获来自不同

模型和样本的各种知识，用于个性化更新。4）借助多种不相关数据，包括随机噪声数据、Tiny-ImageNet数据集（Russakovsky等，2015）和CIFAR-100（Canadian Institute for Advanced Research 100 classes）数据集。本文方法在Digits和Office-Caltech数据集中评估了AHF的性能，结果表明其在域内分类和跨域泛化设定中都表现出卓越的性能。

1 相关工作

1.1 模型异构联邦学习

随着参与者对于独特模型的需求，具有模型异构性的联邦学习已经成为一个活跃的研究领域。Chang等人（2019）和Sattler等人（2020）基于知识蒸馏来聚合来自公共数据的逻辑输出。基于共享预测的思想，Sun等人（2020）在提供无噪音差分隐私保证的同时，提高了效率和泛化能力。另一种策略是基于共享全局模型。例如Shen等人（2020）允许参与者共同训练一个广义模型和很多个性化模型。Liang等人（2020）在每个参与者上协作学习紧凑的本地表示和跨所有参与者的全局模型。此外，已有方法分别从模型结构和数据域的角度对模型进行聚类。当进一步考虑参与者的能力和需求，例如有限的计算能力和更新速度要求时，已有方法允许训练具有不同计算复杂度的异构本地模型，并通过聚合具有类似结构的本地模型生成全局推断模型。Singhal等人（2021）结合了联邦全局模型和本地私有模型，旨在部分本地联邦学习，适用于规模化的训练和推理。这些现有算法中大多需要精心处理的相关数据或共享的全局模型。因此，公共数据的高要求和模型结构的制约将限制模型性能和参与者的独特设计。然而，AHF对于不相关数据具有鲁棒性，并允许参与者独立设计模型。

1.2 数据异构联邦学习

已经有大量关于如何处理数据异构性的联邦学习研究。一种方法是基于参数刚性（刘腊梅 等，2022），例如，在目标函数中增加了一个约束项来提高方法的稳定性，Shamir等人（2014）进一步探索了通过添加修正项来加速收敛，Shoham等人（2019）添加了弹性权重共享以防止灾难性遗忘（肖振久 等，2022），Zhang等人（2021）则基于简单的一阶模型优化来进行联邦更新。此外，一些方法采用了将参与者模型与全局模型相结合，例如Peterson等人（2019）、Arivazhagan等人（2019）和Li等人（2019）的方法。此外，Yoon等人（2021）受到数据增强方法Mixup的启发，使用平均本地数据进行训练。还有一些关于将元学习与联邦组合的研究，例如Jiang等人（2019）和Fallah等人（2020）的方法。然而，现有的方法只能解决模型同构情况，而不考虑模型异构性。在本文中，AHF通过自适应权重策略，在模型异构性的前提下实现了令人满意的性能。

2 本文方法

2.1 问题定义

本文介绍了异构联邦学习问题，其旨在通过所有参与者的协作训练异构模型，从而在私有数据存在领域差异时提高性能。

有K个参与者（由k索引），本文定义θ_k表示其私有模型。此外使用$\boldsymbol{D}_k=\{(\boldsymbol{X}^k,Y^k)\mid\boldsymbol{X}^k\in\mathbf{R}^{N_k\times d},Y^k\in\mathbf{R}^{N_k\times c}\}$表示第$k$个参与者的数据，其中$\boldsymbol{X}^k$，$Y^k$分别表示第$k$个样本的输入图像和对应的标签，$N_k$表示私有数据的数量，$d$表示输入维度，$c$定义为分类的类别数。此外，将参与者的数据分布形式表示为$P_k(x,y)$，并将其重写为$P(\boldsymbol{X}^k\mid Y^k)P(Y^k)$。由于参与者具有相同的任务，对于任意两个参与者$k$和$i$，假设他们共享一致的标签空间：$P(Y^k)=P(Y^i)$。此外，在异构联邦学习中，进一步定义模型异质性和数据异质性如下：

1）模型异质性。$shape(\theta^k)/shape(\theta^i)$，$shape$表示网络结构描述，$\theta$为网络参数。参与者独立设计模型，即网络架构不同。

2）数据异质性。$P(\boldsymbol{X}^k\mid Y^k)\neq P(\boldsymbol{X}^i\mid Y^i)$。私有数据之间存在域差距，比如：私有数据的条件分布$P(\boldsymbol{X}\mid Y)$在参与者之间发生变化。

为了解决上述问题，本文提出利用无关数据$\boldsymbol{D}_0=\{\boldsymbol{X}^0\mid\boldsymbol{X}^0\in\mathbf{R}^{N_0\times d}\}$作为桥梁，使异构本地参与者之间进行通信。这些无关数据通常在实际应用中很容易获得。例如，网络图像或公共数据集。有趣的是，本文证明了即使在FL过程中使用随机生成的噪声作为通信介质，本文的方法仍然具有竞争力。

2.2 AHF算法

本文提出了一种新颖的方法，名为自适应异构联邦学习（AHF），通过同时解决模型异质性和数据

异质性来进行联邦学习。图 2 展示了该方法的概述：在模型和数据异质性的条件下，AHF 通过调整随机噪声的分布来学习知识，并进一步考虑参与者的不相似性和样本的多样性，为不同的参与者做自适应的权重更新。

2.2.1　联合学习

具体而言，AHF 旨在构建异构模型在不相关数据上的通信方法，并提高异构私有数据中的模型性能。与大多数现有方法依赖相关数据来近似平均分类结果不同，AHF 的特点在于展示了在不相关数据上进行协作训练的可行性，包括随机噪声和没有一致标签空间的数据。对于不相关数据，考虑到缺乏特定的含义或标签空间与私有数据的不一致性，计算分类误差相对不太合适。因此，本文获取了无关数据上的 logits 输出 $z_k = f(\theta_k^{t-1}, \boldsymbol{X}^0) \in \mathbf{R}^{N^0 \times c}$，其中，$t$ 代表联邦通信交流周期。由此，测量参与者之间的 logits 输出分布差异为

$$L_{ki} = KL(f(\theta_k^{t-1}, \boldsymbol{X}^0), f(\theta_k^{t-1}, \boldsymbol{X}^0)) \in \mathbf{R}^{N^0} \quad (1)$$

在这里，基于 Kullback-Leiblers 散度（KL）来测量分布差异。通过在 logits 输出上测量分布差异，每个参与者可以获得其他人的知识，而不会泄露私人数据或模型架构细节。同时，与使用平均预测来衡量差异相比，AHF 可以凸显参与者之间的差异。这种差异是由于私有数据之间存在巨大的领域差距，导致在相同的样本上提供了多种预测偏好。自然地，本文计算参与者的不相似度来调整上述分布差异，以便可以从领域差距较小的参与者中学习更相似的知识，即

$$PD_{ki} = \sum_{j=1}^{N^0} \frac{\boldsymbol{z}_k^j \cdot \boldsymbol{z}_k^i}{\left\| \boldsymbol{z}_k^j \right\| \left\| \boldsymbol{z}_i^j \right\|} \in \mathbf{R} \quad (2)$$

式中，z_k^j 代表第 k^t 个参与者在第 j^t 样本上的 logits 输出。本文基于 logits 输出计算余弦相似度来评估参与者之间的不相似度。与其他相似度测量相比，余弦相似度衡量的是 logits 输出的维度之间的相对差异，而不是专注于具体的数值。此外，余弦相似度将结果规范化到[−1, 1]之间，有效地衡量了参与者之间的不相似度，并避免了异常权重的影响。因此，当参与者之间的余弦相似度降低时，对应的参与者不相似度增加。在这种情况下，本文假设他们拥有更多的差异化知识，并增加相应的学习权重。

此外，本文进一步考虑了不相关数据中的样本多样性。虽然由于标签空间不一致，模型无法将它们分类到正确的类别中，但可以通过测量逻辑输出的反应来为样本设置不同的权重，具体为

$$V_k = Var(z_k, dim = 1) \in \mathbf{R}^{N^0} \quad (3)$$

$$SD_k = \left(1 - \frac{V_k - \min(V_k)}{\max(V_k) - \min(V_k)}\right) \times 2 \in \mathbf{R}^{N^0} \quad (4)$$

式中，Var 代表方差，V_k 和 SD_k 分别表示客户端方差和归一化客户端方差。本文使用方差来量化每个样本的重要性，并进一步进行最小—最大归一化操作将值规范化到[0, 2]范围内。然后，对第 k 个参与者进行协同对齐，具体为

$$L_{\mathrm{Col}} = \sum_{i}^{K} \frac{1}{K-1} PD_{ki} \times \frac{\sum_{j}^{N^0} (L_{ki}^j \times SD_k^j)}{N^0} \quad (5)$$

式中，j 表示不相关数据中的第 j 个样本。AHF 通过测量不相关数据上的分布差异，并考虑参与者的不相似性和样本的多样性，让每个模型可以从其他人的逻辑输出中学习知识。因此，有益的知识可以在联邦学习中动态有效地共享。

2.2.2　本地学习

每个参与者都在各自的私有数据上更新模型，以防止从学习私有数据转移到学习通信数据时发生灾难性遗忘。然而，仅仅优化本地数据会带来本地域过度拟合的现象。因此，本文利用通信后的前一个模型进行知识蒸馏，以获得跨领域的知识。在这里，t 代表联邦通信周期索引。然后，本地更新通过以下目标完成，具体为

$$L_{\mathrm{loc}} = CE(f(\theta_k^t, \boldsymbol{X}^k), Y^k) + KL(f(\theta_k^t, \boldsymbol{X}^k), f(\theta_k^{t-1}, \boldsymbol{X}^k)) \quad (6)$$

由于本地更新是基于各自的私有数据，使用交叉熵（CE）来计算第 k 个参与者模型对私有数据的输出与相应标签之间的损失 L_{loc}，并在本地轮次（E）轮内循环此过程。特别地，AHF 不依赖于平均模型参数或添加太多本地计算，相反，基于参与者的对数输出，本文考虑一种新颖的通信模式（分布差异）、参与者的不相似性和样本多样性来进行协作对齐。在这种情况下，成本是可控的，并且不会随着参与者数量或模型复杂度的增加而扩展。

AHF 的概述如图 3 所示，通过测量不相关数据上的对数输出的分布差异，AHF 实现了异构模型之间的通信。自适应权重更新策略为不同参与者提供个性化的更新。此外，本地学习平衡了来自其他参

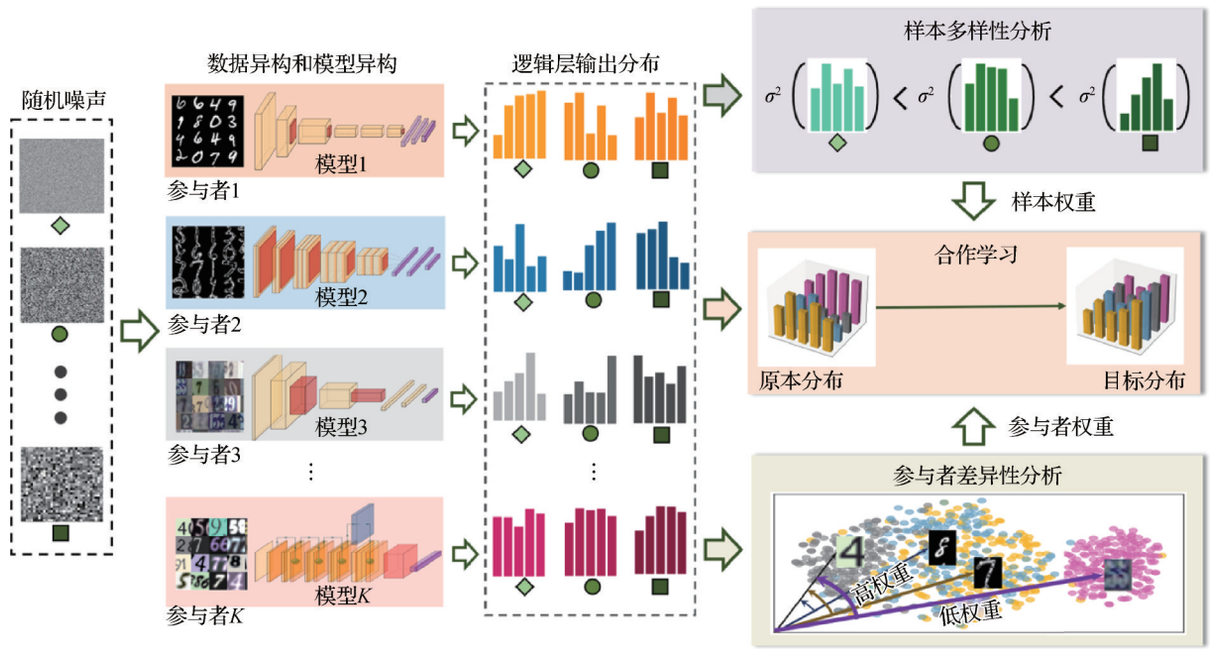

图2　自适应异质联合学习的描述

Fig. 2　Illustration of adaptive heterogeneous federated learning

与者和本身的知识。

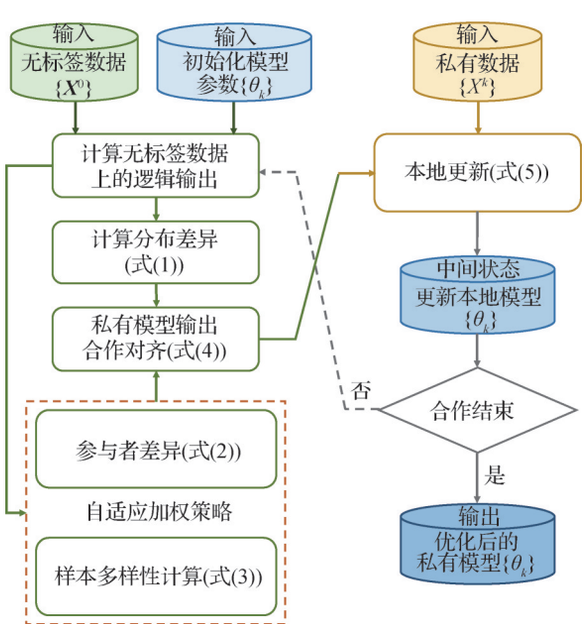

图3　AHF的流程图

Fig. 3　Flow chart of AHF

2.3　理论分析

在这里，本文讨论本地数据分布和无关公开数据之间分布对于泛化性能的影响。本文假定无关数据的分布是D，第i客户端的本地分布和经验分布是D_i和$\hat{D}_i$。假设$h \in \boldsymbol{H}$是在$\hat{D}_i$上学习的，用h_{D_i}表示。K个局部模型在D上的风险上限主要由两部分组成：1）在全局经验分布$\hat{D}$上训练的模型的经验风险；2）依赖于D_i和D之间的分布差异的项，概率为$1-\delta$。

理论分析表明，与全局经验分布的集中式模型相比，性能受到局部分布D_i和全局分布D之间的分布差异影响。因而无关公开数据与本地有更小的分布差异，有利于多方在联邦学习场景下提升合作性能。

2.4　讨论

在AHF中，本文认为协同学习的本质是学习其他参与者的预测模式，而不是通信样本的具体或相似标签。因此，本文测量参与者之间的对数输出的分布差异，并旨在对齐对数输出分布。通过分布差异，不仅具有特定意义的图像，例如Tiny-ImageNet和CIFAR-100，而且随机噪声也可以反映参与者的差异预测模式。因此，AHF适用于通过随机噪声进行通信，并实现了令人满意的性能。然而，现有的相关FL方法，如FEDMD、FEDMD-NFDP和RCFL，依赖于相关数据进行通信。但是当参与者通过包括随机噪声、Tiny-ImageNet和CIFAR-100在内的不相关数据进行通信时，现有方法无法有效地处理。特别是对于随机噪声，不仅存在不一致的标签空间，而且没有具体的含义。由于标签空间的不一致，计算分类的损失是不合理的。因此，现有的方法无效。且由

于没有具体的含义，很难断言随机噪声更类似于某个类。此外，AHF 对于不同的模型结构都可以对数输出进行操作。因此，当参与者共享相同的模型结构时，AHF 仍然有效。但是，本文也注意到任务一致性要求的限制。对于多任务设置，对数输出可能不仅具有不同的维度，而且对于相同的维度也包含不同的含义。这种限制也是相关方法共有的。

3　实　验

3.1　实验设定

1）数据异质性。本文在两种不同的参与者场景（Digits 和 Office-Caltech 分类任务）上进行了实验。在第 1 个场景中，私有数据分别设置为 MNIST（Modified National Institute of Standards and Technology）（LeCun 等，1998）、USPS（U. S. Postal Service）（Hull，1994）、SVHN（Street View House Numbers）（Netzer 等，2011）和 SYN（Synthetic Digits）（Roy 等，2018），它们之间存在领域差距。同样，在第 2 个场景中，它们分别分配给了 Amazon、Caltech、Dslr 和 Webcam。本文还对 3 种不同的不相关数据进行了实验，包括具有与私有数据不一致的标签空间的数据，例如 Tiny-ImageNet 和 CIFAR-100，以及随机噪声数据。值得注意的是，随机噪声数据由在区间[0，1）中的均匀分布中生成的随机数填充。本文在图 4 中进一步说明数据的异质性，在 Digits 和 Office-Caltech 场景中，都有 4 个域差异。

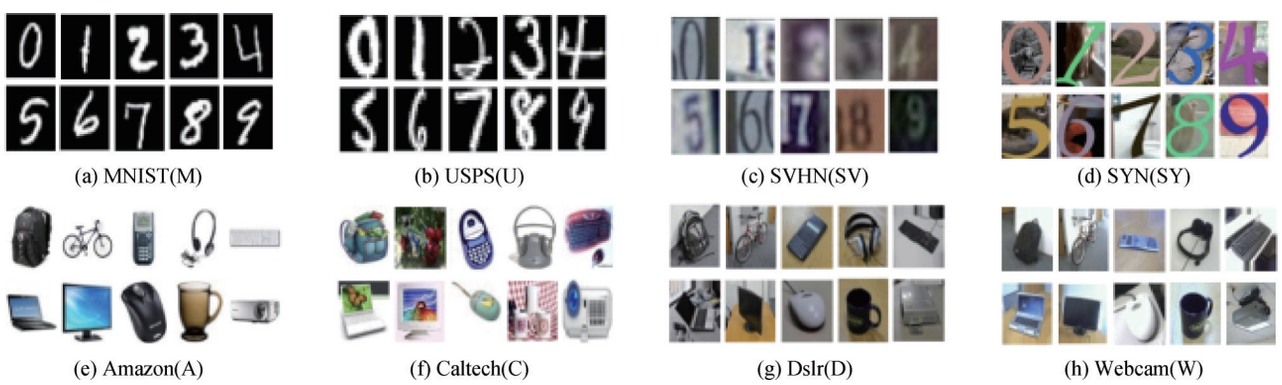

图 4　数据异质性的实验说明

Fig. 4　Experiment illustration of data heterogeneity ((a) MNIST (M); (b) USPS (U); (c) SVHN (SV); (d) SYN (SY); (e) Amazon (A); (f) Caltech (C); (g) Dslr (D); (h) Webcam (W))

2）模型异质性。对于这两个参与者场景，参与者设计了独特的模型结构，可以通过不同的主干网络和不同的分类器进行区分。因此，本文为这两个场景中的每个参与者设置了 ResNet-10（residual neural network-10）（He 等，2016）、ResNet-12（He 等，2016）、Efficientnet（Tan 和 Le，2019）和 Mobilenetv2（Sandler 等，2018）不同的模型架构。

3）实现细节。本文在拥有足够内存的 4 个 TITAN XP GPU 上进行实验，以训练参与者的模型。对于每个参与者，私有数据分为训练数据和测试数据，这些数据用于本地更新和报告测试精确性。具体而言，在 Digits 场景中，MNIST、USPS、SVHN 和 SYN 分配给 4 个参与者。相应的私有数据大小分别设置为 150、100、5 000 和 2 000。至于 Office-Caltech，每个参与者分别被分配为 Amazon、Caltech、Dslr 和 Webcam，相应的私有数据大小为 764、895、121 和 232。考虑到两个场景下私有数据大小的差异，相应的不相关数据大小分别设置为 5 000 和 1 000，batch 大小分别为 512 和 64。对于所有数据，图像分辨率被调整为 32 以提高计算效率。使用 Adam 优化器（Kingma 和 Ba，2014），初始学习率为 $\eta = 10^{-3}$，并为不同方法进行 $T = 41$ 个通信轮次。

为了评估方法的性能，本文记录最后 3 个通信时期每个私有测试数据的平均测试精确度。此外，通过在其他私有数据上测试参与者的模型来评估泛化精确度。

3.2　比较基准

本文比较的相关的联邦方法包括：1）FEDMD（NeurIPS’19）（Li 和 Wang，2019）：依靠相关的公共数据集进行蒸馏操作。2）FEDDF（NeurIPS’19）（Lin

等,2020):通过无标签数据进行集成知识蒸馏,实现鲁棒模型融合。3)RHFL(CVPR'22)(Fang等,2022):用于标签噪声联邦学习下的异构客户端的鲁棒性抗噪声损失。4)FCCL(CVPR'22)(Huang等,2022):使用无标签数据对齐交叉相关性,并蒸馏跨领域知识。由于具体实验设置并不完全一致,本文保留方法的关键特征以在相同条件下进行比较。实际上,参与者的模型已经使用自己的数据进行训练,以完成相应的任务。因此,本文训练模型直至收敛,并在没有联邦学习的情况下获得测试精确性。

3.3 与其他优秀方法比较

与当前先进的方法在两个不相关的数据集CIFAR-100和Tiny-ImageNet的随机噪声上对两个不同参与者场景进行比较。同时,进一步考虑到使用公共数据集存在偏差的情况,在只采样10类的CIFAR-100数据集上进行了额外的公开数据集验证公开数据集存在偏差的性能。

1)域内性能。首先评估当样本没有特定含义时,如随机噪声,AHF的表现如表1所示。表1结果清楚地表明AHF提供了有潜力的性能。特别是在Office-Caltech情景下,由于私有数据样本较少,且私有数据之间的领域差距较大,比Digits情景更具挑战性。AHF具有处理这些挑战的能力,并显示出一致稳定的模型性能。AHF性能仅次于利用了本地预训练的本地模型进行支撑,本文的方法大幅度降低了本地训练开销成本,确保了稳定的域内性能。

表1 自适应异构联邦学习及比较基准的域内泛化表现精度

Table 1 Adaptive heterogeneous federated learning and comparative benchmarks for intra-domain accuracies performance

/%

数据集	方法	Digits					Office-Caltech				
		M←	U←	SV←	SY←	AVG	AM←	CA←	D←	W←	AVG
随机噪声	FEDMD	77.19	62.50	87.96	97.60	81.31	72.41	58.98	57.75	60.34	62.37
	FEDDF	82.17	45.79	88.00	97.98	78.48	72.03	58.53	44.58	51.30	56.61
	RHFL	87.12	62.30	88.63	97.68	83.93	62.18	53.52	19.96	27.57	40.80
	FCCL	86.69	86.23	87.86	93.63	**88.60**	73.66	63.94	68.79	74.47	**70.21**
	本文	84.26	78.99	88.89	97.43	87.39	72.82	62.18	67.31	75.48	69.44
CIFAR-100	FEDMD	77.30	80.85	77.73	87.72	80.70	71.78	57.29	68.37	72.20	67.41
	FEDDF	82.95	78.84	78.46	91.30	83.38	70.01	56.49	56.18	64.86	64.13
	RHFL	86.69	80.39	88.44	97.90	**88.35**	73.35	58.92	70.91	74.47	69.41
	FCCL	88.84	84.42	78.55	91.23	85.69	75.09	60.46	74.95	73.56	**71.01**
	本文	85.44	69.01	86.82	97.60	84.71	73.77	59.66	74.10	75.03	70.64
CIFAR-100 限制10类筛选	FEDMD	83.97	42.12	86.66	97.8	77.63	71.54	60.14	57.54	67.01	64.05
	FEDDF	84.82	50.81	87.75	98.08	80.36	8.56	9.8	13.38	7.12	9.71
	RHFL	83.21	72.06	88.36	98.1	85.43	71.19	56.66	44.37	48.36	55.14
	FCCL	87.74	81.43	88.45	97.87	**88.87**	73.52	59.93	67.94	74.35	**68.93**
	本文	82.97	69.07	87.78	97.83	84.41	73.73	61.83	60.3	72.43	67.07
Tiny-ImageNet	FEDMD	84.64	34.31	87.35	98.10	76.10	72.79	58.89	66.67	69.38	66.93
	RHFL	75.06	55.19	88.49	97.23	78.99	71.61	55.09	64.97	66.10	64.44
	FCCL	87.34	85.03	87.89	93.98	**88.56**	74.32	62.75	73.46	75.25	**71.44**
	本文	86.60	49.98	87.45	97.58	80.40	73.49	61.35	73.04	74.01	70.47

注:"←"表示本域,加粗字体表示平均域内精度的最优结果,下划线字体表示平均域内精度的次优结果。AVG表示在每个域计算出来的平均域内精度。

2)跨域泛化性能。本文在表 2 中报告了泛化性能。对于 Digits 场景，可以看到 AHF 在 3 个无关数据上显示出可竞争的平均泛化精度性能。此外，AHF 在更具挑战性的 Office-Caltech 场景上呈现出稳定的泛化性能。值得注意的是，本文进一步在具有特定含义的不相关数据集上评估了所提出的方法。与 CIFAR-100 相比，由于具有更多类别的噪声网络图像，Tiny-ImageNet 更具挑战性。大多数竞争方法在处理 Tiny-ImageNet 时遇到了很大的麻烦，导致表现更差，而 AHF 显示出显著的鲁棒性来处理 Tiny-ImageNet。此外，本文在 CIFAR-100 上评估了 AHF，结果表明 AHF 超越大多数方法。随着不相关数据（类别、噪声效应）的困难程度增加，竞争方法仍然可以在私有数据之间的相对较小领域差距（即 Digits 情景）下提高有限的表现。但是，在 Office-Caltech 情景下，当领域差距较大时，绝大多数方法表现出负面表现。就 AHF 而言，在各种不相关数据和参与者情况下，几乎所有参与者都获得了有潜力的结果。这意味着 AHF 对具有不同程度领域差距的私有数据以及不同难度的不相关数据具有鲁棒性。本文进一步考虑公开数据存在的偏差，控制公开数据集 CIFAR-100 的可选择类别规模到 10 类，进行有偏采样，本文方法依旧取得了稳定提升。

3.4　消融实验

本文进行了一项消融研究，以评估 AHF 中每个重要组成部分的影响。通过分别将参与者差异度（participant diversity，PD）和样本多样性（sample

表 2　自适应异构联邦学习及比较基准的域外泛化表现精度

Table 2　Adaptive heterogeneous federated learning and comparative benchmarks for inter-domain accuracies performance

/%

数据集	方法	Digits					Office-Caltech				
		M→	U→	SV→	SY→	AVG	AM→	CA→	D→	W→	AVG
随机噪声	FEDMD	16.74	12.49	50.05	47.74	31.75	24.68	34.08	19.58	24.05	25.59
	FEDDF	15.84	12.62	44.29	51.29	31.01	18.46	31.63	15.53	20.37	21.49
	RHFL	22.70	10.93	53.67	54.88	<u>36.54</u>	21.34	21.67	13.97	14.10	17.77
	FCCL	27.86	24.48	49.62	48.03	**37.49**	22.21	34.06	20.44	25.70	<u>25.60</u>
	本文	20.20	15.76	50.26	44.68	32.72	24.78	38.35	16.37	26.77	**26.56**
CIFAR-100	FEDMD	8.97	12.61	40.89	43.03	26.38	21.09	35.13	21.76	30.57	<u>27.13</u>
	FEDDF	13.23.	19.29	45.25	43.95	30.43	23.87	28.29	16.27	22.82	22.81
	RHFL	19.15	16.72	51.74	48.65	**34.06**	19.11	27.50	16.55	23.83	21.74
	FCCL	20.74	20.60	4.68	28.02	<u>33.51</u>	25.16	33.68	17.52	23.81	25.04
	本文	24.73	15.14	44.48	47.34	32.92	28.72	37.77	22.19	33.23	**30.47**
CIFAR-100 限制 10 类筛选	FEDMD	10.62	13.55	45.31	49.98	29.86	21.37	37.19	16.85	20.84	24.06
	FEDDF	16.65	11.31	47.05	52.25	31.81	10.1	9.69	8.49	10.58	9.71
	RHFL	20.14	18.38	50.27	50.46	**34.81**	19.33	24.06	15.11	17.94	19.11
	FCCL	11.26	17.42	49.24	51.39	<u>32.32</u>	20.49	35.1	17.95	23.72	<u>24.31</u>
	本文	10.76	14.69	47.96	46.48	29.97	25.94	36.32	21.39	31.08	**28.68**
Tiny-ImageNet	FEDMD	18.50	13.26	51.23	52.06	<u>33.76</u>	30.33	36.27	22.67	30.61	<u>29.97</u>
	RHFL	16.14	14.05	54.26	47.57	33.00	18.66	30.70	16.02	20.94	21.58
	FCCL	31.84	31.03	46.81	52.75	**40.60**	22.49	43.67	19.54	32.19	29.47
	本文	17.77	13.92	46.11	50.97	32.19	31.15	42.26	27.50	32.87	**33.44**

注："→"表示跨域，加粗字体表示平均跨越精度的最优结果，下划线字体表示平均跨域精度的次优结果。AVG 表示在每个域计算出来的平均跨域精度。

diversity，SD）从本文的方法中删除来评估本文方法的两个关键部分的影响，以观察模型性能。如表3所示，删除其中任一部分都会导致在CIFAR-100和Tiny-ImageNet中的性能下降。

表3　关键模块上的消融实验

Table 3　Comparison results of ablation studies

数据集	PD	SD	Digits					Office-Caltech				
			M→	U→	SV→	SY→	AVG	AM→	CA→	D→	W→	AVG
CIFAR-100	√	–	14.05	14.26	44.23	42.47	28.75	27.68	35.95	19.23	29.08	27.98
	–	√	14.71	14.36	48.16	44.04	30.31	28.35	32.87	19.13	32.66	28.25
	√	√	24.73	15.14	44.48	47.34	**32.92**	28.72	37.77	22.19	33.23	**30.47**
Tiny-ImageNet	√	–	15.30	16.38	47.66	49.85	**32.29**	25.56	42.84	26.12	26.11	30.15
	–	√	11.27	15.81	48.59	48.80	31.11	27.02	39.67	24.58	23.90	28.79
	√	√	17.77	13.92	46.11	50.97	32.19	31.15	42.26	27.50	32.87	**33.44**

数据集	PD	SD	Digits					Office-Caltech				
			M←	U←	SV←	SY←	AVG	AM←	CA←	D←	W←	AVG
CIFAR-100	√	–	84.73	72.06	87.23	96.93	**85.23**	73.49	61.15	63.48	73.79	67.97
	–	√	84.20	66.83	86.37	97.55	83.73	73.83	59.78	70.49	74.69	69.69
	√	√	85.44	69.01	86.82	97.60	84.71	73.77	59.66	74.10	75.03	**70.64**
Tiny-ImageNet	√	–	88.17	61.94	87.76	97.45	**83.83**	73.48	60.52	71.34	73.56	69.42
	–	√	86.83	49.69	88.26	97.58	80.59	73.94	61.30	61.15	65.43	65.45
	√	√	87.30	53.43	87.62	97.20	81.32	73.49	61.35	73.04	74.01	**70.47**

注：“→”表示跨域，“←”表示本域，“√”表示采用，“–”表示未采用，加粗字体表示跨域或域内平均精度的最优结果。

为了探究泛化性能，表3清楚地显示，在Office-Caltech删除任一部分都会导致在这两个参与者场景的不同未见过的私有数据上泛化性能的高概率下降。具体而言，对于每个参与者，AHF都能达到良好的泛化精确性。此外，对于整体泛化性能，AHF获得了更高的平均泛化精确性。当利用丰富的Tiny-ImageNet数据集的时候，AHF表现出了更好的性能，能够从丰富的公开数据集中受益。

4　结　论

为了解决联邦学习中的模型和数据异构性问题，本文提出了自适应异构联邦学习（AHF）。引入无关公开数据，以对齐来自不同私有模型的逻辑输出分布，从而实现多方隐私友好的合作。本文考虑参与者数据分布差异和公开样本多样性的自适应权重聚合调整。从而，AHF可以有效地适应更具挑战性的场景，例如私有数据之间的更大领域差距。在各种参与者情境（Digits和Office-Caltech）和不相关数据（随机噪声、CIFAR-100和Tiny-ImageNet）上的大量实验表明，本文提出的AHF优于现有相关的FL方法。但本文工作还存在一些不足之处，本文仅考虑了输出层面的多方合作，没有充分利用深度神经网络的多维度知识。下一步的改进路线可以考虑在特征层面实现更丰富维度的多方联邦知识交流合作。

参考文献（References）

Arivazhagan M G, Aggarwal V, Singh A K and Choudhary S. 2019. Federated learning with personalization layers [EB/OL]. [2022-05-21]. https://arxiv.org/pdf/1912.00818.pdf

Chang H Y, Shejwalkar V, Shokri R and Houmansadr A. 2019. Cronus: robust and heterogeneous collaborative learning with black-box knowledge transfer [EB/OL]. [2022-05-21]. https://arxiv.org/pdf/1912.11279.pdf

Dinh C T, Tran N H and Nguyen T D. 2020. Personalized federated learning with Moreau envelopes//Proceedings of the 34th Confer-

ence on Neural Information Processing Systems. Vancouver, Canada: Curran Associates Inc.: 21394-21405

Fallah A, Mokhtari A and Ozdaglar A. 2020. Personalized federated learning with theoretical guarantees: a model-agnostic meta-learning approach//Proceedings of the 34th International Conference on Neural Information Processing Systems. Vancouver, Canada: Curran Associates Inc.: 3557-3568 [DOI: 10.5555/3495724.3496024]

Fang X W and Ye M. 2022. Robust federated learning with noisy and heterogeneous clients//Proceedings of 2022 IEEE Conference on Computer Vision and Pattern Recognition. New Orleans, USA: IEEE.

Guo X, Xing P W, Feng S W, Li B A and Miao C Y. 2020. Federated learning with diversified preference for humor recognition//International Workshop on Federated Learning for User Privacy and Data Confidentiality in Conjunction with IJCAI. Yokohama, Japan

Hard A, Rao K, Mathews R, Beaufays F, Augenstein S, Eichner H, Kiddon C and Ramage D. 2018. Federated learning for mobile keyboard prediction [EB/OL]. [2023-05-08]. https://arxiv.org/pdf/1811.03604.pdf

He K M, Zhang X Y, Ren S Q and Sun J. 2016. Deep residual learning for image recognition//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA: IEEE: 770-778 [DOI: 10.1109/CVPR.2016.90]

Howard A G, Zhu M L, Chen B, Kalenichenko D, Wang W J, Weyand T, Andreetto M and Adam H. 2018. MobileNets: efficient convolutional neural networks for mobile vision applications [EB/OL]. [2023-05-08]. https://arxiv.org.pdf.1704.04861.pdf

Huang W K, Ye M and Du B. 2022. Learn from others and be yourself in heterogeneous federated learning//Proceedings of 2022 IEEE Conference on Computer Vision and Pattern Recognition. New Orleans, USA: IEEE.

Hull J J. 1994. A database for handwritten text recognition research. IEEE Transactions on Pattern Analysis and Machine Intelligence: 550-554.

Jiang Y, Konečný J, Rush K and Kannan S. 2019. Improving federated learning personalization via model agnostic meta learning [EB/OL]. [2023-05-08]. https://arxiv.org/pdf/1909.12488pdf

Ju C, Zhao R H, Sun J C, Wei X G, Zhao B, Liu Y, Li H S, Chen T J, Zhang X W and Gao D S. 2020. Privacy-preserving technology to help millions of people: federated prediction model for stroke prevention. [EB/OL]. [2023-05-08]. https://arxiv.org.pdf.2006.10517.pdf

Kairouz P, Brendan McMahan H, Avent B, Bellet A, Bennis M, Bhagoji A N, Bonawit K, Charles Z, Cormode G, Cummings R, D'Oliveira R G L, Eichner H, El Rouayheb S, Evans D, Gardner J, Garrett Z, Gascón A, Ghazi B, Gibbons P B, Gruteser M, Harchaoui Z, He C Y, He L, Huo Z Y, Hutchinson B, Hsu J, Jaggi M, Javidi T, Joshi G, Khodak M, Konecný J, Korolova A, Koushanfar F, Koyejo S, Lepoint T, Liu Y, Mittal P, Mohri M, Nock R, Özgür A, Pagh R, Qi H, Ramage D, Raskar R, Raykova M, Song D, Song W K, Stich S U, Sun Z T, Suresh A T, Tramèr F, Vepakomma P, Wang J Y, Xiong L, Xu Z, Yang Q, Yu F X, Yu H and Zhao S. 2019. Advances and open problems in federated learning [EB/OL]. [2023-05-08]. https://arxiv.org/pdf/1912.04977.pdf

Kairouz P, McMahan H B, Avent B, Bellet A, Bennis M, Bhagoji A N, Bonawitz K, Charles Z, Cormode G and Cummings R. 2019. Advances and open problems in federated learning [EB/OL]. [2023-05-08]. https://arxiv.org.pdf.1912.04977.pdf

Kingma D P and Ba J. 2014. Adam: a method for stochastic optimization [EB/OL]. [2023-05-08]. https://arxiv.org/pdf/1412.6980.pdf

LeCun Y, Bottou L, Bengio Y and Haffner P. 1998. Gradient-based learning applied to document recognition. Proceedings of the IEEE, 86(11): 2278-2324 [DOI: 10.1109/5.726791]

Li T, Sahu A K, Zaheer M, Sanjabi M, Talwalkar A and Smith V. 2018. Federated optimization in heterogeneous networks [EB/OL]. [2023-05-08]. https://arxiv.org/pdf/1812.06127.pdf

Li D L and Wang J P. 2019. FedMD: heterogenous federated learning via model distillation//Proceedings of International Workshop on Federated Learning for User Privacy and Data Confidentiality in Conjunction with NeurIPS. Canada: [s. n.]:#03581

Liang P P, Liu T, Liu Z Y, Allen N B, Auerbach R P, Brent D, Salakhutdinov R and Morency Louis-P. 2019. Think locally, act globally: federated learning with local and global representations//Proceedings of International Workshop on Federated Learning for User Privacy and Data Confidentiality in Conjunction with NeurIPS. Canada: [s. n.]

Lin T, Kong L J, Stich S U and Jaggi M. 2019. Ensemble distillation for robust model fusion in federated learning//Proceedings of Neural Information Processing Systems. Canada: [s. n.]

Liu L M, Zong J X, Xiao Z J, Lan H and Qu H C. 2022. Cross-consistent semantic segmentation algorithm based on manifold regularization. Journal of Image and Graphics, 27(12): 3542-3552 (刘腊梅, 宗佳旭, 肖振久, 兰海, 曲海成. 2022. 流形正则化的交叉一致性语义分割算法. 中国图象图形学报, 27(12): 3542-3552) [DOI: 10.11834/jig.210571]

McMahan B, Moore E, Ramage D, Hampson S and Arcas B A Y. 2017. Communication-efficient learning of deep networks from decentralized data//Proceedings of the 20th International Conference on Artificial Intelligence and Statistic. [s.l.]: PMLR: 1273-1282

Netzer Y, Wang T, Coates A, Bissacco A, Wu B and Ng A Y. 2011. Reading digits in natural images with unsupervised feature learning//Proceedings of Annual Conference on Neural Information Processing Systems. [s. l.]: NeurIPSW

Peterson D, Kanani P and Marathe V J. 2019. Private federated learning with domain adaptation [EB/OL]. [2023-05-08]. https://arxiv.org/pdf/1912.06733.pdf

Roy P, Ghosh S, Bhattacharya S and Pal U. 2018. Effects of degradations on deep neural network architectures [EB/OL]. [2023-05-08]. https://arxiv.org/pdf/1807.10108.pdf

Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S, Huang Z H, Karpathy A, Khosla A, Bernstein M, Alexander C, Berg A C and Li F F. 2015. ImageNet large scale visual recognition challenge. International Journal of Computer Vision, 115(3): 211-252 [DOI: 10.1007/s11263-015-0816-y]

Sattler F, Marban A, Rischke R and Samek W. 2020. Communication-efficient federated distillation [EB/OL]. [2023-05-08]. https://arxiv.org/pdf/2012.00632.pdf

Shen T, Zhang J, Jia X K, Zhang F D, Huang G, Zhou P, Wu F and Wu C. 2020. Federated mutual learning [EB/OL]. [2023-05-08]. https://arxiv. org. pdf. 2006. 16765. pdf

Shoham N, Avidor T, Keren A, Israel N, Benditkis D, Mor-Yosef L and Zeitak I. 2019. Overcoming forgetting in federated learning on non-IID data//Proceedings of International Workshop on Federated Learning for User Privacy and Data Confidentiality in Conjunction with NeurIPS. Canada: [s. n.]

Singhal K, Sidahmed H, Garrett Z, Wu S S, Rush K and Prakash S. 2021. Federated reconstruction: partially local federated learning// Proceedings of International Conference on Machine Learning. Virtual Online: ICML.

Sun L C and Lyu L J. 2020. Federated model distillation with noise-free differential privacy. [EB/OL]. [2023-05-08]. https://arxiv.org.pdf.2009.05537.pdf

Tan M X and Le Q. 2019. EfficientNet: rethinking model scaling for convolutional neural networks//Proceedings of International Conference on Machine Learning. Virtual Online: ICML:6105-6114

Virtual online: ICLR: 1-19

Xiao Z J, Zong J X, Lan H, Wei X and Tang X L. 2022. Image semantic segmentation based on manifold regularization constraint. Journal of Image and Graphics, 27(4): 1204-1215 (肖振久, 宗佳旭, 兰海, 魏宪, 唐晓亮. 2022. 流形正则化约束的图像语义分割. 中国图象图形学报, 27(4): 1204-1215) [DOI: 10.11834/jig.200527]

Yang Q, Liu Y, Chen T J and Tong Y X. 2019. Federated machine learning: concept and applications, ACM Transactions on Intelligent Systems and Technology, 2019: #3298981 [DOI: 10. 1145/3298981]

Yoon T, Shin S M, Hwang S J and Yang E H. 2021. FEDMIX: approximation of mixup under mean augmented federated learning// Proceedings of International Conference on Learning Representations.

Zhang M, Sapra K, Fidler S, Yeung S and Alvarez J M. 2021. Personalized federated learning with first order model optimization//Proceedings of International Conference on Learning Representations. Virtual Online: ICLR.

作者简介

黄文柯,男,博士研究生,主要研究方向为联邦学习。
E-mail:wenkehuang@whu.edu.cn
叶茫,通信作者,男,教授,主要研究方向为计算机视觉、多媒体检索和安防大数据分析。E-mail:yemang@whu.edu.cn
杜博,男,教授,主要研究方向为计算机视觉和高光谱遥感。
E-mail:dubo@whu.edu.cn