arXiv:2304.05871v1 [cs.LG] 12 Apr 2023

具有联邦特征和集中式特征的边缘云协同学习

Zexi Li*浙江大学 中国杭州
zexi.li@zju.edu.cn

Qunwei Li*蚂蚁集团 中国杭州
qun
wei.qw@antgroup.com

Zhou Yi 犹他大学 美国盐湖城
yi.zhou@utah.edu

Wenliang Zhong*蚂蚁集团 中国杭州
yice.zwl
@antgroup.com

Guannan Zhang 蚂蚁集团 中国杭州
zgn138592
@antgroup.com

吴超*浙江大学 中国杭州
chao.
wu@zju.edu.cn

摘要

联邦学习（FL）是一种不会损害用户隐私的流行边缘计算方式。当前的联邦学习范式假设数据仅存储在边缘端，而云服务器仅执行模型平均。然而，在推荐系统等现实场景中，云服务器能够存储历史特征和交互特征。在本文中，我们提出的**Edge-CloudCollaborativeKnowledgeTransfer Framework（ECCT）**弥合了边缘端与云端之间的差距，实现二者之间的双向知识迁移，共享特征嵌入和预测逻辑值。ECCT 整合了多项优势，包括增强个性化、支持模型异构性、容忍训练异步化，以及减轻通信负担。在公开数据集和工业数据集上的大量实验表明，ECCT 具有出色的有效性，并具备在学术界和工业界应用的潜力。

CCS 概念

• **计算方法学 → 机器学习**； • **信息系统 → 数据挖掘**。

关键词

边缘-云协同学习、联邦学习、推荐系统

ACM 参考格式：
Zexi Li、Qunwei Li、Yi Zhou、Wenliang Zhong、Guannan Zhang、Chao Wu。2023。具有联邦特征和集中式特征的边缘云协同学习。载于第*46*届国际*ACM*信息检索研究与发展会议论文集（*SIGIR '23*），*2023*年*7*月*23*日至*27*日，台湾台北。ACM，美国纽约州纽约，5页。https://doi.org/10. 1145/3539618.3591976

* 两位作者对本研究贡献相同。⋆ 通讯作者。该工作是在 Zexi Li 于蚂蚁集团实习期间完成的。

特此授权：在不收取费用的情况下，为个人或课堂使用而复制本作品的全部或部分纸质或数字副本，前提是不得为营利或商业利益制作或分发副本，并且副本首页须附有本声明及完整引文。对于本作品中不属于作者所有的组件，必须尊重其版权。允许在注明出处的情况下进行摘要。若要以其他方式复制、重新出版、发布到服务器或重新分发至列表，必须事先获得具体许可和/或支付费用。请通过 permissions@acm.org 申请许可。*SIGIR'23*，*2023*年*7*月*23*日至*27*日，台湾台北。© 2023 版权所有归所有者/作者所有。出版权授权给 ACM。ACMISBN 978-1-4503-9408-6/23/07...$15.00https://doi.org/10.1145/3539618.3591976

1 引言

在传统的集中式训练范式中，数据从多源边缘设备收集，并据此在云服务器上训练大型模型 [16, 21]。近年来，出于对用户隐私和通信效率的考虑，联邦学习（FL）被提出作为集中式训练范式的替代方案 [13–15, 18, 24]。FL假设所有训练数据都保留在边缘设备本地，并通过迭代式边缘训练和云端模型平均来获得全局模型。事实证明，它在隐私保护场景中具有可行性和适用性。例如，键盘输入属于手机用户的敏感信息，而Google已利用FL在保护用户隐私的同时改进下一词预测 [4]。

然而，当前的联邦学习框架忽略了云端数据的存在，也未直接纳入云端训练。在工业应用中，云服务器通常拥有丰富的计算资源，可以利用丰富的历史特征和非私有特征训练大模型。以推荐系统（RS）为例，隐私敏感特征（包括用户的性别、年龄等）以及其他实时特征（例如用户与物品的即时交互）通常无法被云端访问，而是存储在设备本地。我们将这些特征称为**“联邦特征”**。另一方面，集中式历史特征（隐私顾虑较少）包括用户与物品的历史交互、物品类别、物品标记、物品间嵌入等，存储在云服务器上。我们将这些特征称为**“集中式特征”**。我们发现，现有的联邦学习框架未能充分利用这两类特征。主要原因在于，联邦特征和集中式特征具有不同的特征空间、维度和可访问性，因此难以利用联邦特征来提升全局模型的性能。此外，将集中式特征迁移到边缘端并不可行，因为这类特征体量庞大，而边缘设备的存储和计算资源有限。

传统的集中式学习范式和当前的联邦学习范式仅采用单一的云端训练或边缘训练。然而，我们认为，在实践中，边云协同至关重要，既能获得更好的泛化表示，也能提供更个性化的用户服务 [9, 20, 27]。因此，在本文中，我们提出了一种**边**缘-**云**端**协**同知识**迁**移框架（**ECCT**）。在该框架中，我们在服务器上训练一个大模型，并在每台边缘设备上训练一个轻量化模型。 The

特征嵌入和预测逻辑值在边缘端与云端之间共享，以支持知识迁移与协作。随后，边缘端和云端模型采用交替最小化（AM）方法 [2, 3, 10]，通过嵌入融合和知识蒸馏利用迁移的知识。

ECCT 相较于传统的联邦学习框架具有多项优势。首先，它通过将知识从集中式特征迁移到设备端模型，能够增强本地个性化。与联邦学习方法相比，这可以改善个性化效果。其次，ECCT 支持模型异构性，使边缘端模型的架构可以不同，同时仍保持兼容。第三，ECCT 天然支持异步训练，并且在仅选择部分训练设备时具有更强的鲁棒性。最后，ECCT 减轻了通信负担，因为传输的知识大小远小于模型参数。

我们的贡献如下。

• 我们开发了 ECCT 框架，联合利用云端的集中式特征和边缘端的联邦特征。据我们所知，这是首篇利用这两种特征实现边云协同的论文，并具有广阔的工业应用前景。

• 在 ECCT 中，我们通过嵌入融合和知识蒸馏，在边缘端与云端之间实现双向知识迁移。

• 在公共数据集和工业数据集上进行的大量实验均证明了ECCT的有效性，并展现了其在工业领域广泛且切实可行的应用前景。

2 相关工作

边云协同联邦学习。 FedGKT [10] 在联邦学习中融入了分割学习，以实现边云协同。它基于设备的嵌入和逻辑值，在服务器上训练更大的 CNN 模型。然而，它没有利用集中式数据，而且仅通过传输逻辑值，云端向边缘端传递的知识较弱。有些工作将云端数据作为公共数据集，通过知识蒸馏来辅助边缘训练 [7, 19, 28]。我们认为，在边缘设备上存储此类公共数据集并不现实，这阻碍了其在工业中的应用。**边云协同推荐系统**。在 [27], 中，MoMoDistill 被提出用于在边缘端微调云端推荐系统模型的元补丁，从而实现用户个性化。在[26], 中，训练了一个因果元控制器，用于管理设备端推荐与云端推荐之间的权衡。我们注意到， [26, 27] 假设集中式数据是联邦数据的总和。但在实际场景中，由于隐私和实时性问题，集中式数据和联邦数据具有不同的特征。已有方法从重排序 [9] 和物品请求 [20] 的角度解决边缘推荐中的推理问题。

3 方法

在第 3.1 节中，我们给出所提出框架的主要问题表述；在第 3.2 节中，我们提供若干可与该框架结合使用的实用扩展，以进一步提升性能或加强隐私保证。

3.1 主要问题表述

问题设置。网络中有 K 个边缘设备。具体而言，学习任务是在整个数据集中对C 个类别进行分类的监督学习。该数据集包含两部分特征，分别位于服务器（集中式特征，D）和各设备（联邦特征，$\hat{\mathrm{D}}$）。第k个设备拥有自己的私有实时特征数据集 $\hat{\mathrm{D}}^k := \left\{\left(\hat{\mathrm{X}}_i^k, y_i^k\right)\right\}_{i=1}^{N^k}$，其中 $\hat{\mathrm{X}}_i$ 是第 i个训练样本， y_i 是 $\hat{\mathrm{X}}_i$对应的标签， $y_i^k \in \{1, 2, \dots C\}$，而 N^k是数据集 $\hat{}D^k$的样本数量（也记作 D^k）。因此，我们有 $\hat{\mathrm{D}} = \{\hat{\mathrm{D}}1, \hat{\mathrm{D}}^2, \dots, \hat{\mathrm{D}}^K\}$，$N = \sum_{k=1}^K N^k$。服务器拥有所有设备的非私有历史特征，表示为 $\mathrm{D}^k := \left\{\left(\mathrm{X}_i^k, y_i^k\right)\right\}_{i=1}^{N^k}$， $\mathrm{D} = \{\mathrm{D}1, \mathrm{D}^2, \dots, \mathrm{D}^K\}$。

学习目标。对于边缘设备 k，我们部署一个小型特征编码器（提取器）E_d^k和一个小型分类器 C_d^k；对于云服务器，我们部署一个大型编码器 E_s和一个大规模下游模型 C_s（包括分类器）。设 $\mathrm{W}_d^k = \{\mathrm{E}_d^k, \mathrm{C}_d^k\}$ 和 $\mathrm{W}_s = \{\mathrm{E}_s, \mathrm{C}_s\}$。由于云端和边缘端的特征彼此分离，我们将单个全局模型优化重新表述为一个非凸优化问题，需要同时训练服务器模型 W_s和设备模型 W_d^k。服务器和设备 k 的学习目标及其对应的损失函数 ℓ_s 和ℓ_d^k 如下

$$\arg\min_{\mathcal{W}_s} F_s(\mathcal{W}_s) = \arg\min_{\mathcal{W}_s} \sum_{k=1}^{K} \sum_{i=1}^{N^k} \ell_s \left(C_s(h_{d,i}^k \oplus h_{s,i}^k), y_i^k\right), \quad (1)$$

$$\arg\min_{\mathcal{W}_d^k} F_d^k(\mathcal{W}_d^k) = \arg\min_{\mathcal{W}_d^k} \sum_{i=1}^{N^k} \ell_d^k \left(C_d^k(h_{d,i}^k \oplus h_{s,i}^k), y_i^k\right), \quad (2)$$

$$\text{where } h_{d,i}^k = \mathcal{E}_d^k(\hat{\mathrm{X}}_i^k),\ h_{s,i}^k = \mathcal{E}_s(\mathrm{X}_i^k). \quad (3)$$

在公式 1 和 2 中， $h_{d,\,i}^k(h_{s,\,i}^k)$ 表示使用设备（服务器）的编码器 $\mathrm{E}_d^k(\mathrm{E}s)$ 从样本 $\hat{\mathrm{X}}_i^k(\mathrm{X}_i^k)$ 中提取的特征嵌入，而 $\oplus$ 表示将两个嵌入拼接为一个的拼接操作。我们采用交替最小化（AM）方法 [2, 3, 10]优化上述目标。对于服务器的优化，我们固定设备的嵌入 $h_{d,\,i}^k$；对于设备k的优化，我们固定服务器的嵌入 $h_{s,\,i}^k$。

除了嵌入之外，我们还传递预测逻辑值，并采用知识蒸馏（知识蒸馏） [11, 17] 来强化知识学习。共享的嵌入和逻辑值在边云协同训练中实现了双向知识迁移，我们分别采用服务器和边缘设备 k 的损失函数如下

$$\ell_s = \ell_{CE} + \alpha_s \sum_{k=1}^{K} \ell_{KD} \left(z_d^k, z_s^k\right), \quad (4)$$

$$\ell_d^k = \ell_{CE}^k + \alpha_d \ell_{KD}^k \left(z_s^k, z_d^k\right), \quad (5)$$

其中， ℓ_{CE} 是预测值与真实标签之间的交叉熵损失， ℓ_{KD} 是用于知识蒸馏的KullbackLeibler散度函数， z_s^k和 z_d^k是预测逻辑值。此外， α_s 和 α_d 是用于控制知识蒸馏强度的超参数。

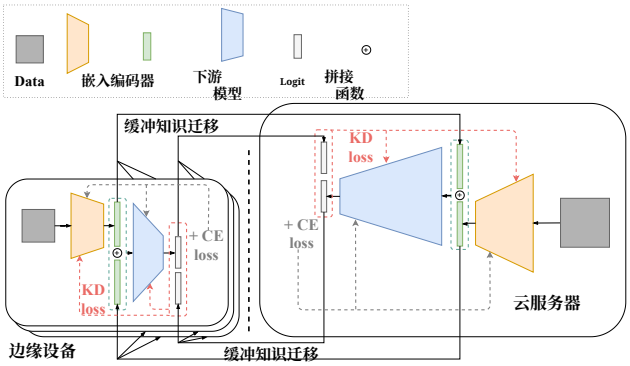


图 1：所提出的边缘—云协同知识迁移框架。

训练流程。所提出的框架演示如图 1 所示。我们的边缘云协同学习框架主要包含三个迭代步骤：边缘训练、云端训练和边云协同。(1) 云端训练。云服务器按照公式 4 进行云端训练，共训练 E_s 个周期。(2) 边缘训练。边缘设备 k 按照公式 5 进行边缘训练，共训练 E_d^k 个周期。(3) 边缘云通信。边缘设备（云服务器）通过编码器在其本地特征上推断嵌入，并通过分类器推断逻辑值，然后将嵌入和逻辑值发送至云服务器（边缘设备）。

步骤 1 和 2 可以在云端与边缘端并行执行，而基于 AM 的方法使我们的框架能够更好地适应异步更新和通信。实际应用中，设备模型会将其嵌入累积到缓冲区中；缓冲区满时，设备会与服务器通信并传输这些嵌入，我们将其称为缓冲知识迁移。缓冲的每一部分知识不必由同一个边缘模型生成。这放宽了联邦学习的同步要求：设备无需使用相同的模型版本（即相同的边缘云通信频率）和相同的本地训练轮次（$\forall i,\ j \in [K]$, $E_d^i = E_d^j = E_d$）。此外，得益于云端强大的计算资源，我们可以设置更多的云端训练轮次（$E_s \gg E_d$），从而利用丰富的云端特征生成更具知识代表性的嵌入，帮助设备模型实现快速训练和更好的个性化。

推理策略。推理时有两种策略可供选择，即基于云端和基于边缘端的策略。云端（边缘端）模型使用从边缘端（云端）传递的嵌入以及自身的特征，对给定样本进行预测。通常，基于边缘端的推理实时性更强、个性化程度更高，而基于云端的推理则具有更强的泛化能力和鲁棒性。可以根据具体的应用场景灵活选择推理策略。

3.2 实用扩展

两阶段策略。我们注意到，在模型泛化能力较弱的早期训练阶段，逻辑值中的知识在表征能力方面较为不足。因此，知识蒸馏可能会导致训练失真。于是，在最初的几个训练周期中，我们只传输嵌入，并使用交叉熵损失（即 $\alpha_s = \alpha_d = 0$）训练轮次，并在后续训练阶段额外传输逻辑值，同时使用知识蒸馏和交叉熵损失。

过滤知识迁移。我们发现，如果模型对某些样本的预测标签错误，那么它在这些样本上的泛化能力很弱；因此，将这些样本的逻辑值作为知识会降低蒸馏效果 [7]。为此，我们采用过滤知识迁移，仅使用正确的逻辑值进行知识蒸馏。

隐私保证。与直接传输设备的原始隐私特征相比，传输嵌入带来的隐私风险较低 [5, 8]。在我们的框架中，还可以通过对来自设备的嵌入应用差分隐私方法，进一步加强隐私保证 [1, 25]。

4 实验

在实验中，我们在三个数据集上评估我们的框架：CIFAR10、Avazu 和 IndustryData。我们在第 4.1 节介绍实验设置，在第 4.2 节展示 CIFAR10 合成特征划分的结果，并在第 4.3 节说明推荐系统数据集（Avazu 和 IndustryData）上的结果。

4.1 实验设置

为进行公平比较，我们考虑三种特征设置："F"仅使用联邦特征，"C2F"将centralized 特征**to** 移至边缘端作为**f**ederated 特征，而"C&F"将云端的集中式特征与边缘端的联邦特征相结合。我们注意到，由于存储限制和通信开销，"C2F"在现实场景中并不实用。不过，为了进行全面评估，我们仍将其纳入实验。我们的实验使用三个数据集，在边缘设备的本地测试集上评估个性化结果。在训练的前 50 轮中，α_s 和 α_d 均设为 0，随后 50 轮均设为 1。传递的嵌入大小为 128，我们采用具有两层的多层感知机作为下游分类器。

数据集。**CIFAR10** [12] 是一个用于分类的图像数据集，我们通过合成特征划分生成联邦特征和集中式特征。具体而言，对于给定的 32×32 图像，我们将其划分为边缘端的10×32 图像和云端的 22×32 图像。我们分别在边缘端和云端部署 ResNet-8 和 ResNet-56 作为编码器。设备数量为 20。**Avazu** [23] 是一个用于真实世界推荐系统（RS）点击率（CTR）预测的公开数据集。我们随机抽取 200,000 条数据样本，并根据设备_ip 将其分配给各个设备，共有 1769 台设备。我们将主要与物品详情（例如应用_类别）相关的稀疏特征设为云端的集中式特征，并将稠密特征视为联邦特征。我们使用 AutoInt [22]作为骨干编码器。**IndustryData** 是一个真实世界的工业数据集，用于推荐系统（RS），从支付宝应用的系统日志中提取而来。我们随机抽取 500,000 条数据样本，并将其分配给 2000 台设备。我们将用户敏感特征（如 u_性别和 u_职业）分配在边缘端，并将历史交互特征作为云端的集中式特征。我们使用 AutoInt 作为编码器，并在点击率（CTR）和转化率（CVR）预测任务上开展实验。

基线。我们将我们的框架与联邦学习基线进行比较，主要包括FedAvg [18]（通用联邦学习框架）和 FedGKT [10]（最先进的边缘-云协同联邦学习框架）。对于 CIFAR10 实验，由于分割特征具有相同的特征空间，

表1：不同设备数据异质性和 E_d 条件下 CIFAR10 的 Top-1 测试准确率（%）。「homo」/「hetero」：设备模型同质性/异质性。

	数据异质性	IID		非独立同分布	
特征	方法/E_d	1	3	1	3
F	FedAvg	57.36	59.75	52.36	56.96
	FedGKT（同构）	66.78	67.90	38.88	44.59
	FedGKT（异构）	51.57	54.34	23.36	26.20
C2F	FedAvg	61.81	66.32	57.09	62.69
	FedGKT（同构）	78.44	79.05	30.91	57.70
	FedGKT（异构）	66.09	66.04	30.38	49.30
C&F	FedDF	31.9	49.14	28.86	41.36
	FedBE	38.28	51.55	31.73	43.09
	ECCT（同构）	**81.89**	**82.46**	**63.06**	**62.61**
	ECCT（异构）	**77.31**	**77.18**	**60.86**	**64.06**

表 2：异步化条件下 CIFAR10 上的 Top-1 测试准确率（%）结果。

方法	FedAvg		ECCT	
选择比例	1.0	0.5	1.0	0.5
Syn.	62.69 -	55.78 11.0% ↓	67.81 -	69.28 2.1% ↑
异步版本	60.51 3.5% ↓	53.29 15.0% ↓	67.56 0.4% ↓	64.42 5.3% ↓
异步轮次	59.55 5.0% ↓	50.93 18.8% ↓	66.92 1。3% ↓	64.65 4.9% ↓
异步两者	58.47 6。7% ↓	51.49 17.9% ↓	67.31 0.7% ↓	64.81 4.6% ↓

我们还测试了基于蒸馏的联邦学习方法（FedDF [17] 和 FedBE [6]），这些方法可以通过蒸馏利用集中式特征。

4.2 合成特征划分

在本节中，我们在采用合成特征划分的CIFAR10上进行实验。表1展示了在不同设备数据异质性和边缘训练轮次数量（E_d）下，使用狄利克雷采样 [17] 为设备分配非独立同分布样本时的结果。我们的方法在独立同分布和非独立同分布设置下均优于基线。此外，当存在设备模型异质性时，与FedGKT 相比，我们的方法表现出更小的性能下降。在集中式特征上进行蒸馏（FedDF 和 FedBE）会导致模型性能更差，这表明传统联邦学习方法无法有效利用集中式特征。

表 2 展示了异步设置下的结果。与 ECCT 相比，FedAvg 的性能下降更明显，尤其是在每个训练轮次中结合均匀随机的部分设备选择时。

表 3 展示了不同设备数量和选择比例下各种方法的性能。我们的 ECCT 在所有设置下均优于 FedAvg 和 FedGKT，并且对边缘设备参与情况的差异有更强的适应性。

4.3 推荐系统

本节在推荐系统中实现这些方法，并使用公开数据集（Avazu）和工业数据集（IndustryData）进行实验。

表 3：不同设备数量和设备选择比例下 CIFAR10 上的 Top-1 测试准确率（%）结果。

设备数量	50			100		
选择比例	0.6	0.3	0.1	0.6	0.3	0.1
FedAvg	47.64	33.99	14.12	44.04	31.55	16.19
FedGKT	31.06	44.84	45.16	59.68	57.82	52.53
ECCT	**68.43**	**68.68**	**45.33**	**78.36**	**75.91**	**67.01**

表 4：Avazu 点击率预测。

特征	F		C2F		C&F
方法	FedAvg	FedGKT	FedAvg	FedGKT	**ECCT**
AUC	0.5783	0.5293	0.6595	0.5246	**0.6694**
MSE	0.1370	0.1488	0.1320	0.1559	**0.1373**

表 5：IndustryData 的点击率和转化率预测。

Task	CTR		CVR	
方法/指标	AUC	MSE	AUC	MSE
FedAvg	0.6829	0.0879	0.8145	0.0199
FedGKT	0.5409	0.1023	0.6068	0.0262
ECCT	**0.6885**	**0.0875**	**0.8754**	**0.0184**

首先，我们在 Avazu 上开展点击率预测任务实验，结果如表 4 所示。即使将所有集中式特征都移至边缘端（“C2F”），我们的方法在曲线下面积方面仍优于 FedAvg 和 FedGKT。其次，我们在 IndustryData 上测试这些方法，用于点击率和转化率预测任务。表 5 的结果表明，我们的边缘—云协同框架优于联邦学习方法，尤其是在转化率预测任务中。这说明该框架在工业领域，特别是在线推荐方面，具有很大的潜力。

我们注意到 FedGKT 在推荐系统中的表现不佳，这在我们看来是合理的。在 FedGKT 中，从云端到边缘端的知识传递完全基于逻辑值。对于点击率和转化率预测这类二分类任务，逻辑值所包含的信息并不比标签更多，因此 FedGKT 无法实现从云端到边缘端的有效传递。然而，我们的框架在传递的嵌入中融入了更多知识。

5 结论

本文提出了一种边缘—云协同知识迁移框架，用于联合利用集中式特征和联邦特征。与以往的联邦学习方法相比，所提出的方法具有多项优势：提升边缘个性化能力、支持模型异构性、容忍异步化，并减轻边缘端与云端之间的通信负担。

致谢

本工作得到中国国家重点研发计划（2021ZD0110400）、国家自然科学基金（U19B2042）、浙江省科技计划项目（2022C01044）、安徽省高校协同创新项目（GXXT-2021-004）、浙江大学社会治理研究院以及中央高校基本科研业务费专项资金（226-2022-00064）的支持。

参考文献

[1] Martin Abadi、Andy Chu、Ian Goodfellow、H Brendan McMahan、Ilya Mironov、Kunal Talwar 和 Li Zhang。2016年。基于差分隐私的深度学习。载于*2016*年*ACM SIGSAC*计算机与通信安全大会论文集。308–318。[2] Hédy Attouch、Jérôme Bolte、Patrick Redont 和 Antoine Soubeyran。2010年。用于非凸问题的近端交替最小化和投影方法：一种基于Kurdyka-Łojasiewicz不等式的方法。《运筹学数学》35卷2期（2010年），438–457。[3] Jérôme Bolte、Shoham Sabach 和 Marc Teboulle。2014年。用于非凸和非光滑问题的近端交替线性化最小化。《数学规划》146卷1期（2014年），459–494。[4] Keith Bonawitz、Hubert Eichner、Wolfgang Grieskamp、Dzmitry Huba、Alex Ingerman、Vladimir Ivanov、Chloe Kiddon、Jakub Konečn`y、Stefano Mazzocchi、Brendan McMahan 等。2019年。迈向大规模联邦学习：系统设计。《机器学习与系统会议论文集》1（2019年），374–388。[5] Fahao Chen、Peng Li、Toshiaki Miyazaki 和 Celimuge Wu。2021年。Fedgraph：采用智能采样的联邦图学习。《*IEEE*并行与分布式系统汇刊》33卷8期（2021年），1775–1786。[6] Hong-You Chen 和 Wei-Lun Chao。2020年。FedBE：使贝叶斯模型集成适用于联邦学习。载于国际学习表征会议。[7] Sijie Cheng、Jingwen Wu、Yanghua Xiao 和 Yang Liu。2021年。Fedgems：通过选择性知识融合实现更大型服务器模型的联邦学习。*arXiv*预印本 *arXiv:2110.11027*（2021年）。[8] Xingbo Fu、Binchi Zhang、Yushun Dong、Chen Chen 和 Jundong Li。2022年。联邦图机器学习：概念、技术和应用综述。*arXiv*预印本 *arXiv:2207.11812*（2022年）。[9] Yu Gong、Ziwen Jiang、Yufei Feng、Binbin Hu、Kaiqi Zhao、Qingwen Liu 和 Wenwu Ou。2020年。EdgeRec：手机淘宝上的边缘端推荐系统。载于第*29*届*ACM*信息与知识管理国际会议论文集。2477–2484。[10] Chaoyang He、Murali Annavaram 和 Salman Avestimehr。2020年。群组知识迁移：边缘端大型卷积神经网络的联邦学习。《神经信息处理系统进展》33（2020年），14068–14080。[11] Geoffrey Hinton、Oriol Vinyals、Jeff Dean 等。2015年。提炼神经网络中的知识。*arXiv*预印本 *arXiv:1503.025312*卷7期（2015年）。[12] Alex Krizhevsky、Geoffrey Hinton 等。2009年。从微型图像中学习多层特征。（2009年）。[13] Tian Li、Anit Kumar Sahu、Ameet Talwalkar 和 Virginia Smith。2020年。联邦学习：挑战、方法与未来方向。《*IEEE*信号处理杂志》37卷3期（2020年），50–60。https://doi.org/10.1109/MSP.2020.2975749[14] Zexi Li、Jiaxun Lu、Shuang Luo、Didi Zhu、Yunfeng Shao、Yinchuan Li、Zhimeng Zhang、Yongheng Wang 和 Chao Wu。2022年。迈向有效的聚类联邦学习：采用自适应邻居匹配的点对点框架。《*IEEE*大数据汇刊》（2022年）。[15] Zexi Li、Feng Mao 和 Chao Wu。2022年。如果无法共享数据，我们能否共享模型？《*Patterns*》3卷11期（2022年），100603。[16] Shih-wei Liao、Tzu-Han Hung、Donald Nguyen、Chinyen Chou、Chiaheng Tu 和 Hucheng Zhou。2009年。面向数据中心应用的基于机器学习的预取优化。载于高性能计算、网络、存储与分析会议论文集。1–10。[17] Tao Lin、Lingjing Kong、Sebastian U Stich 和 Martin Jaggi。2020年。联邦学习中的鲁棒模型融合集成蒸馏。《神经信息处理系统进展》33（2020年），2351–2363。[18] Brendan McMahan、Eider Moore、Daniel Ramage、Seth Hampson 和 Blaise Aguera y Arcas。2017年。基于去中心化数据的深度网络通信高效学习。载于人工智能与统计学。PMLR，1273–1282。[19] Minh NH Nguyen、Huy Q Le、Shashi Raj Pandey 和 Choong Seon Hong。2022年。CDKT-FL：在联邦学习中使用代理数据集进行跨设备知识迁移。*arXiv*预印本 *arXiv:2204.01542*（2022年）。[20] Xufeng Qian、Yue Xu、Fuyu Lv、Shengyu Zhang、Ziwen Jiang、Qingwen Liu、Xiaoyi Zeng、Tat-Seng Chua 和 Fei Wu。2022年。推荐系统中的智能请求策略设计。载于第*28*届 *ACM SIGKDD*知识发现与数据挖掘会议论文集。3772–3782。[21] Saim Salman、Christopher Streiffer、Huan Chen、Theophilus Benson 和 Asim Kadav。2018年。DeepConf：利用机器学习自动管理数据中心网络拓扑。载于*2018*年网络遇见人工智能与机器学习研讨会论文集。8–14。[22] Weiping Song、Chence Shi、Zhiping Xiao、Zhijian Duan、Yewen Xu、Ming Zhang 和 Jian Tang。2019年。AutoInt：通过自注意力神经网络自动进行特征交互学习。载于第*28*届 *ACM*信息与知识管理国际会议论文集。1161–1170。[23] Will Cukierski Steve Wang。2014年。点击率预测。https://kaggle.com/competitions/avazu-ctr-prediction [24] Jianyu Wang、Zachary Charles、Zheng Xu、Gauri Joshi、H Brendan McMahan、Maruan Al-Shedivat、Galen Andrew、Salman Avestimehr、Katharine Daly、Deepesh Data 等。2021年。联邦优化实用指南。*arXiv*预印本 *arXiv:2107.06917*（2021年）。[25] Kang Wei、Jun Li、Ming Ding、Chuan Ma、Howard H Yang、Farhad Farokhi、Shi Jin、Tony QS Quek 和 H Vincent Poor。2020年。差分隐私联邦学习：算法与性能分析。*IEEE*信息取证与安全汇刊15（2020年），3454–3469。[26] Jiangchao Yao、Feng Wang、Xichen Ding、Shaohu Chen、Bo Han、Jingren Zhou 和 Hongxia Yang。2022年。通过元控制器实现端云协同推荐。载于第*28*届 *ACM SIGKDD*知识发现与数据挖掘会议论文集。4353–4362。[27] Jiangchao Yao、Feng Wang、Kunyang Jia、Bo Han、Jingren Zhou 和 Hongxia Yang。2021年。面向推荐的端云协同学习。载于第*27*届 *ACM SIGKDD*知识发现与数据挖掘会议论文集。3865–3874。[28] Hong Zhang、Ji Liu、Juncheng Jia、Yang Zhou、Huaiyu Dai 和 Dejing Dou。2022年。FedDUAP：利用服务器上的共享数据进行动态更新和自适应剪枝的联邦学习。*arXiv*预印本 *arXiv:2204.11536*（2022年）。