资源受限场景中的联邦学习技术综述

王鹏举[1,2], 卢江虎[1,2], 刘博超[1,2], 葛仕明[1]

[1] 中国科学院信息工程研究所, 北京 中国 100093

[2] 中国科学院大学网络空间安全学院, 北京 中国 100049

摘要 随着数字经济的快速发展，数据安全威胁日益严峻，数据安全已成为数字经济时代最紧迫的安全问题。联邦学习作为一种新兴的分布式机器学习框架，在实现数据安全和隐私保护的前提下，聚合多方数据资源，协同构建联合模型，为打破数据孤岛现象和实现数据安全融合提供了一种行之有效的方案，为数字经济发展夯实了安全基础，受到了学术界和工业界的广泛关注。然而，在实际的应用部署中，联邦学习面临着资源受限场景所带来的严峻挑战，资源受限场景中存在计算设备、通信网络和建模数据等一系列资源受限问题，这些问题严重地限制了联邦学习的应用和发展。因此，有必要从资源受限的角度来研究联邦学习，着力解决资源受限场景中的突出问题，以实现高效地部署联邦学习。本文主要探讨了资源受限场景中部署联邦学习的实际解决方案。首先，介绍了数字经济的发展现状和联邦学习的背景知识；其次，讨论了资源受限场景中的联邦学习所面临的问题与挑战；然后，对资源受限场景中的联邦学习的研究现状展开了系统深入地调研，分别从架构高效、通信高效、计算高效和异构融合等四个方面对比分析了典型的联邦学习；最后，对资源受限场景中的联邦学习进行了总结与展望。

关键词 联邦学习；资源受限；架构高效；通信高效；计算高效；异构融合

中图法分类号 TP309. 2

Federated Learning in Resource Constrained Scenarios: A Comprehensive Survey

WANG Pengju[1,2], LU Jianghu[1,2], LIU Bochao[1,2], GE Shiming[1]

[1] Institute of Information Engineering, Chinese Academy of Sciences, Beijing 100093, China

[2] School of Cyber Security, University of Chinese Academy of Sciences, Beijing 100049, China

Abstract With the rapid development of the digital economy, the threats to data security are becoming serious by degrees, so data security has become the most urgent security issue in the current digital economy era. As an emerging distributed machine learning framework, federated learning aggregates data resources from multiple parties and realizes the collaborative construction of a federated model under the premise of data security and privacy protection. It provides an effective solution to break the phenomenon of isolated data islands and realize data security fusion, in addition to consolidating the security foundation for the development of the digital economy. As a result, it has received extensive attention from both academia and industry. However, in practical application deployment, it shows that federated learning faces severe challenges brought by resource constrained scenarios. In resource constrained scenarios, there are a series of resource constrained problems such as computing device, communication network and modeling data, which seriously restrict the application and development of federated learning. Therefore, it is necessary to explore federated learning from the perspective of resource constrained scenarios, and it is urgent to solve the outstanding problems in these scenarios, in order to achieve efficient deployment of federated learning. In this survey paper, we focus on practical solutions for deploying federated learning in resource constrained scenarios. Firstly, we introduce the development status of the digital economy and the background knowledge of federated learning. Secondly, we discuss the problems and challenges of federated learning in resource constrained scenarios. Thirdly, we conduct a systematic and in-depth investigation into the current research status of federated learning, and analyze a number of typical federated learning technologies in terms of architecture efficient, communication efficient, computation efficient, and heterogeneous fusion. Finally, we make a summary and outlook for the development trend of federated learning in resource constrained scenarios.

Key words federated learning; resource constrained; architecture efficient; communication efficient; computation efficient; heterogeneous fusion

---

通讯作者：葛仕明，博士，副研究员，Email：geshiming@iie.ac.cn。

本课题得到北京市自然科学基金（No.L192040）和国家重点研发计划（No.2020AAA0140001）资助。

收稿日期：202X-X-X；修改日期：202X-X-X；定稿日期：202X-X-X

1 引言

随着万物互联时代的到来，越来越多的移动设备和物联网设备等智能终端设备融入社会生活的各个领域，这些智能终端设备持续生成的海量数据呈现出一种爆炸式增长趋势，人类已经进入了大数据时代[1-2]，海量集聚的数据蕴藏了巨大的价值，为智能化发展带来了新的机遇。数字经济是继农业经济、工业经济之后的一种极具发展潜力的经济形态[3]，作为一种新的经济形态，数字经济发展速度之快、覆盖范围之广、影响程度之深前所未有，以数据资源为关键要素的数字经济即将迈向全面扩展期。

与此同时，数字经济发展面临着一些严峻问题和重大挑战。第一是用户隐私保护，我国颁布的《网络安全法》《数据安全法》和《个人信息保护法》等法律法规中，对用户数据的收集和处理提出了严格的隐私保护规范，要求企业或机构对用户数据的收集必须公开和透明，并且在用户未授权时不能交换用户数据；第二是数据孤岛现象，由于不同群体、不同行业、不同区域之间存在数字化竞争壁垒，数据资源呈现出一种割裂和垄断的现状，导致数据资源难以开放和共享，这严重影响了依赖大数据的人工智能技术的快速发展；第三是数据自身特性，数字经济中海量数据资源虽然规模庞大，但是存在形式多样化、流转速度快和价值密度低等缺点，导致数据资源的价值潜力还没有得到充分释放。

如何在满足法律法规、隐私保护和数据安全的前提下，打破数据孤岛现象和实现数据安全融合是数字经济发展的重要战略难题，联邦学习成为了解决这一难题的关键技术。联邦学习（Federated Learning，FL）是由谷歌公司提出的一种新兴的隐私计算技术[4-5]，在保障数据安全和实现隐私保护的前提下，它是一种聚合多方数据资源和协同构建联合模型的分布式机器学习框架。作为隐私计算领域的研究热点，联邦学习与大数据、区块链和边缘计算等前沿技术实现了深度交叉融合，在学术界和工业界有着广泛应用前景，目前已经应用于金融风控[6]、智慧医疗[7]、城市安防[8]和自动驾驶[9]等领域。

由于人工智能技术的高速发展，神经网络结构变得越来越复杂，模型参数量也越来越庞大。联邦学习依托于人工智能技术进行协同训练，需要消耗大量的计算资源和网络资源才能训练出一个性能强大的模型。然而联邦学习所面临的实际应用场景中，移动设备或物联网设备等智能终端设备存在计算资源受限、网络资源受限和部署场景复杂等诸多局限性[10]。因此，如何在资源受限场景中高效地部署联邦学习成为当前亟需解决的重要问题。

为了解决如何高效地部署联邦学习这一难题，国内外学者从不同角度开展了一系列卓有成效的研究工作，涌现出大量联邦学习的综述文章。Yang 等人[11]讨论了联邦学习的定义、架构和应用，Li 等人[12]概述了联邦学习的挑战、现状和未来，Chen 等人[13]分析了联邦学习的安全威胁和防御方案，Li 等人[14]调研了联邦学习和区块链相结合的应用需求和发展方向，Lim 等人[15]总结了联邦学习在移动边缘网络中的研究现状和应用场景，Niknam 等人[16]探讨了联邦学习在无线移动通信中的应用案例和潜在挑战，Nguyen 等人[17]指出了联邦学习在物联网中的应用领域和关键技术，Imteaj 等人[18]研究了联邦学习在物联网中的资源受限挑战和实际解决方案。鉴于已有的综述文章对资源受限场景中的联邦学习尚未展开调研[11-14]，以及调研范围存在局限性[15-18]，本文的主要贡献包括：1）根据资源受限场景所面临的挑战对联邦学习的研究思路实现了分类；2）针对资源受限场景中的联邦学习进行了充分地调研和分析；3）面向资源受限场景展望了联邦学习的研究方向。

本文的章节安排如下：第 1 节介绍了数字经济的发展现状和联邦学习的背景知识；第 2 节讨论了资源受限场景中的联邦学习所面临的问题与挑战；第 3 节到第 6 节对资源受限场景中的联邦学习展开了系统调研，分别从架构高效、通信高效、计算高效和异构融合等四个方面，对典型的联邦学习做出了综合对比分析；第 7 节对资源受限场景中的联邦学习进行了总结与展望。

2 问题与挑战

目前联邦学习的研究大多面向资源丰富场景，其中服务器和客户端具备充足的计算资源和稳定的网络资源。然而实际应用场景往往是资源受限场景，其中移动设备和物联网设备等智能终端设备面临着资源匮乏和环境复杂的难题。除此之外，随着联邦学习中深度神经网络模型的规模性与复杂度与日俱增，在资源受限场景中联邦学习的部署效率极其低下，甚至无法正常运行。因此，亟需研究资源受限场景中的联邦学习，以实现联邦学习的高效部署。

2.1 联邦学习

传统的联邦学习由一个中心服务器和多个客户端构成，其中每个客户端拥有本地数据，在中心服务器的协调下，多个客户端联合训练完成联邦学习。如图 1 所示，联邦学习的一轮训练过程如下：

1）模型分发，服务器将全局模型分发给客户端；
2）本地训练，客户端用本地数据训练本地模型；
3）模型上传，客户端将本地模型上传到服务器；
4）模型聚合，服务器聚合模型并更新全局模型。
重复以上步骤，执行多轮迭代，直至训练结束。

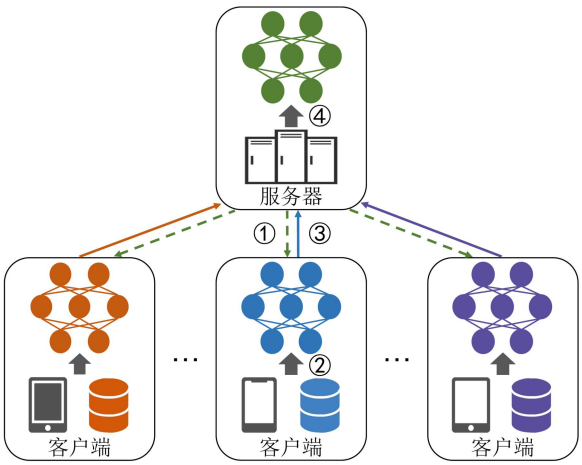

图 1 联邦学习的训练过程

Figure 1 The Training Process of Federated Learning

从训练过程来看，相比于传统的机器学习，联邦学习具有显著优势。第一是不需要将每个客户端的本地数据直接上传到中心服务器，联邦学习通过只共享模型参数来避免数据隐私泄露；第二是联合多个客户端来协同训练一个全局模型，一般情况下全局模型的效果比本地模型的效果更好。

然而，这种训练过程使得联邦学习面临了一些新的问题[19-20]，例如数据异构性、设备异构性、模型收敛慢、通信效率低下和算法健壮性差等问题，这些问题极大地限制了联邦学习的应用和发展。

2.2 资源受限场景

当前，联邦学习在许多领域都开始了广泛地应用探索。智慧医疗[7]领域中，联邦学习利用智能可穿戴设备采集用户的健康数据，协同学习一个预测模型，从而实现用户健康监测。城市安防[21]领域中，通过构建一个联邦分布的训练网络，联合分散的摄像头数据进行全天候监控，从而实现社会治安预警。

由此可知，联邦学习的应用领域中大都是智能终端设备，通过将联邦学习部署在大量智能终端设备上，利用设备采集到的海量数据进行联合建模。然而这些设备往往处于资源受限场景中，部署联邦学习的核心问题是如何实现资源管理[22]，达到降低通信开销、减少计算成本和优化资源分配等目的。目前资源受限场景中面临着以下问题与挑战：

1）设备资源受限。a）设备的计算资源有限，如何保证联邦学习在有限的计算能力下高效训练；b）设备的能量资源有限，如何保证联邦学习训练性能的同时实现低能耗；c）设备的不可靠性，如何保证在部分设备发生故障时联邦学习仍能正常运作；d）设备的异构性，如何保证不同性能的设备在训练过程中有效地协调工作。

2）网络资源受限。a）网络的通信带宽有限，如何保证联邦学习在有限的通信带宽下高效训练；b）网络的不稳定性，如何保证联邦学习在网络延迟或变化时仍能正常通信。

3）数据资源受限。a）数据的异构性，如何保证联邦学习利用非独立同分布数据实现联合训练；b）数据的个性化，如何保证不同场景下利用个性化数据训练出个性化模型。

综上所述，设备、网络和数据等资源受限问题严重限制了联邦学习的应用部署。为了解决资源受限场景中联邦学习的问题与挑战，如表 1 所示，国内外学者围绕架构高效、通信高效、计算高效和异构融合等四个技术思路展开了大量的研究工作。

表 1 资源受限场景中的联邦学习

Table 1 Federated Learning in Resource Constrained Scenarios

技术思路	研究方法	解决方案
架构高效	去中心化架构	Gossip 协议、Ping 请求、IPFS 协议、BlockChain 网络
	分层体系架构	中心服务器-边缘服务器-边缘设备
通信高效	减少参数传输	Dropout 策略、局部表征学习、模型更新反馈、最优客户端采样
	压缩模型梯度	蒸馏、量化、稀疏化、二值化等压缩策略
	减少通信轮数	最大平均偏差、周期性平均、损失差异阈值、局部更新模型
计算高效	模型网络拆分	Split Learning 及其衍生方案
	异步模型更新	陈旧性函数、随机分布式更新、深度强化学习、异步在线学习
	计算资源分配	客户端动态选择、深度 Q 学习、无线电资源管理、低复杂度迭代
异构融合	联邦元学习	基于度量、基于模型、基于优化
	联邦迁移学习	基于样本、基于特征、基于模型、基于关系
	联邦知识蒸馏	离线蒸馏、在线蒸馏、自蒸馏
	联邦多任务学习	硬共享、软共享、层次共享

3 面向架构高效的联邦学习

资源受限场景通常是一个复杂多样的环境，一方面体现在参与设备的异构性，参与设备可以是移动设备或物联网设备等不同类型设备，另一方面体现在网络环境的复杂性，网络环境可以是 4G、5G 或 WiFi 等不同速率网络。资源受限场景中构建一个通用的联邦学习算法主要考虑以下两个方面：

1）健壮性。资源受限场景中的参与设备或网络环境存在异构性，不同参与设备的性能差异巨大，不同网络环境的传输速率不同。因此要保证联邦学习能够实现不同性能设备在不同网络速率下仍能够有效地协调工作。

2）可靠性。资源受限场景中的参与设备或网络环境存在不可靠性，在训练过程中，部分设备可能会因为设备故障或网络中断等因素而退出训练。因此要保证联邦学习在部分设备退出训练后仍能够维持正常工作。

如何实现联邦学习的架构高效是解决资源受限场景中上述问题的首选方案[23]，如表 2 所示，目前面向架构高效的联邦学习方案主要包括：去中心化架构和分层体系架构等方案。

3.1 去中心化架构

传统的联邦学习是在一个中心服务器的协调下利用多个客户端联合建模，然而中心化的联邦学习很难有效部署在资源受限场景中。因为客户端所处的网络环境不稳定，随时都可能与中心服务器断开连接，此外一旦中心服务器出现单点故障，联邦学习的训练过程就会终止。如图 2 所示，去中心化架构可以解决资源受限场景中联邦学习的上述问题，实现联邦学习的高可靠性。

基于 Gossip 协议，Hu 等人[24]提出了一种去中心化联邦学习算法。Gossip 协议是一个基于流行病传播方式的节点或者进程之间交换信息的协议，目前广泛应用于分布式系统中[25]。利用 Gossip 协议的传播特性实现了联邦学习中模型的分发和上传，以此达到了取代中心服务器的效果。然而由于两个客户端之间将模型作为消息传播会导致严重的通信瓶颈，为了解决这个问题，将客户端的模型分段，使得通信开销分散到多个链路上。虽然总的通信开销并没有减少，但是提升了训练速度。

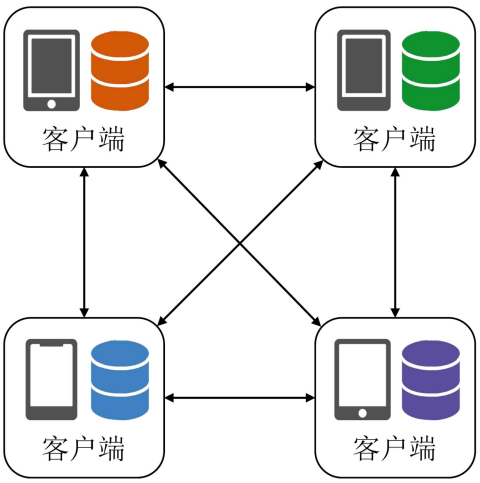

图 2 联邦学习的去中心化架构

Figure 2 Decentralized Architecture of Federated Learning

基于 Ping 请求，Roy 等人[26]提出了一种去中心化联邦学习算法 BrainTorrent。每个客户端上维护一个版本向量，这个向量包含了本地模型版本和最近一次聚合的其他客户端模型版本。每轮训练中，随机选择一个客户端向其他所有客户端发送 Ping 请求，通过判断模型版本的大小来决定是否执行模型更新，BrainTorrent 的这种策略可以代替中心服务器实现模型聚合。此外利用高度动态更新，可以使得客户端的训练过程更加健壮。

表 2 面向架构高效的联邦学习

Table 2 Federated Learning for Architecture Efficient

研究方法	文献	解决方案	贡献
去中心化架构	[24]	利用 Gossip 协议的传播特性实现模型参数的分发和上传	减少训练时间
	[26]	随机选择一个客户端向其他所有客户端发送 Ping 请求	提升训练的健壮性
	[27]	将模型分片后利用 IPFS 协议交换分片梯度来完成模型更新	提升算法的鲁棒性
	[29]	利用 BlockChain 网络代替中心服务器，并加入了验证和激励机制	增强算法的实用性
分层体系架构	[30]	分层联邦算法中允许多个边缘服务器执行部分模型聚合	通信和计算高效
	[31]	提出了一种高效的资源调度算法，以解决计算和通信资源分配问题	训练成本最小化
	[32]	本地更新阶段采用模型网络拆分，全局聚合阶段降低模型更新频率	降低全局通信成本
	[33]	在小区用户和移动基站之间引入小区基站，以解决通信距离长问题	减少端到端延迟

基于 IPFS 协议，Pappas 等人[27]提出了一种去中心化联邦学习算法 IPLS。IPFS（InterPlanetary File System）是一个旨在创建持久且分布式存储和共享文件的网络传输协议[28]，将 IPFS 协议融入联邦学习中实现了去中心化。首先将模型进行分片，每个客户端存储模型的一些分片，然后客户端之间通过交换分片梯度来完成模型更新。其中模型分片旨在保证系统的鲁棒性，为了防止某些客户端在训练的过程中突然断开连接所采取的措施。

基于 BlockChain 网络，Kim 等人[29]提出了一种去中心化联邦学习算法 BlockFL。用 BlockChain 网络来代替中心服务器，并且加入了验证和激励机制来鼓励用户提供训练数据和上传本地模型。BlockFL 克服了单点故障问题，并对本地训练结果进行了验证，将联邦学习训练范围扩展到公共网络中不可信客户端。此外由于 BlockChain 网络会引起延迟问题，因此还研究了端到端学习延迟，通过调整块的生成速率来使延迟最小化，从而增强了算法的实用性。

综上所述，基于 Gossip 协议、Ping 请求、IPFS 协议和 BlockChain 网络等方案可以实现联邦学习的去中心化架构，但是这些方案却增加了通信开销和计算开销。例如各个客户端需要额外的通信过程以保障去中心化训练，这会导致通信开销大幅度提升。此外将中心服务器的聚合工作交付给每个客户端，这会额外增加客户端的计算开销。

3.2 分层体系架构

传统的联邦学习中各种训练操作过度依赖于中心服务器，一旦资源受限场景中海量边缘设备同时跟中心服务器通信，中心服务器容易出现性能瓶颈问题，此外边缘设备和中心服务器之间的通信距离过于遥远。如图 3 所示，联邦学习的“中心服务器-边缘服务器-边缘设备”分层体系架构可以减轻中心服务器的负担。首先利用边缘服务器聚合来自边缘设备的更新，然后利用中心服务器聚合来自边缘服务器的更新。这种算法降低了中心服务器的计算和通信开销，解决了过度依赖中心服务器的瓶颈问题。

为了实现计算和通信高效，Liu 等人[30]提出了一种分层联邦学习算法 HierFAVG。基于云服务器的联邦学习，优点是客户端总数达到百万，可以提供海量数据集，缺点是客户端与云服务器的通信速度慢和不可预测。基于边缘服务器的联邦学习，优点是服务器放置在距离设备最近的边缘，所以通信延迟低，缺点是服务器可以访问的客户端数量有限，这将导致不可避免的训练性能损失。HierFAVG 基于以上两种架构提出了“云-边缘-客户端”联邦学习，不但结合了两种架构的优点，还弥补了上述两种架构的缺点，实现了计算和通信高效。

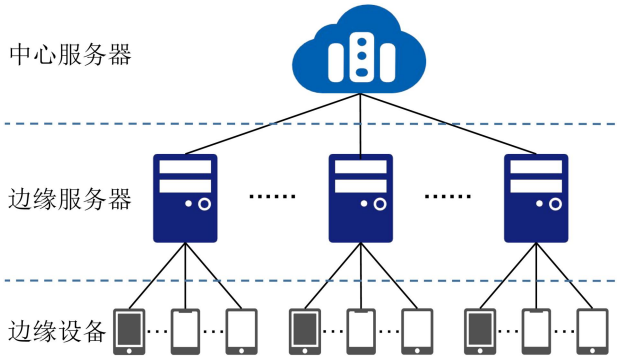

图 3 联邦学习的分层体系架构

Figure 3 Hierarchical Architecture of Federated Learning

基于资源调度算法，Luo 等人[31]提出了一种分层联邦学习算法 HFEL，旨在解决资源分配和边缘关联问题。HFEL 中实现了一种高效的资源调度算法，它可以分解为两个子问题：为每个边缘服务器安排一组预定的设备进行资源分配，以及对跨边缘服务器的设备用户进行边缘关联。实验结果表明，HFEL 资源调度算法相比于基线算法，不但节省了全局训练成本，还取得了更优越的模型性能。

在边缘计算场景中，Ye 等人[32]提出了一种分层联邦学习算法 EdgeFed。在本地更新阶段，采用模型网络拆分的策略，资源受限的边缘设备专注于低层训练，把更多的计算任务分配给资源丰富的边缘服务器，从而使得边缘设备训练更轻、更快。在全局聚合阶段，由于边缘设备和边缘服务器之间的带宽一般高于边缘服务器和中心服务器之间的带宽，采用降低更新频率的策略，目的是降低全局通信成本和减少网络带宽影响。

在跨异构蜂窝网络中，Abad 等人[33]提出了一种分层联邦学习算法 HFL。在小区用户和移动基站之间引入小区基站，避免了小区用户和移动基站因通信距离长而无法实现联邦训练。HFL 中具有本地数据集的小区用户聚集在小区基站周围，以分散的数据集进行联邦随机梯度下降。与此同时，这些小区基站会定期与移动基站通信，以寻求在共享模型上取得共识。为了进一步减少这种分层架构的通信延迟，HFL 使用了梯度稀疏化和周期性平均，并设计了资源分配方案以最小化端到端延迟。

综上所述，采用“中心服务器-边缘服务器-边缘设备”分层体系架构虽然减轻了对中心服务器的依赖程度，但是却带来了一些其他问题，例如计算负担、带宽压力和资源调度等问题。因此还需要借助模型网络拆分、降低更新频率和资源调度算法等方法来实现分层联邦学习的高效训练。

4 面向通信高效的联邦学习

联邦学习中，客户端与中心服务器需要不断地通信来交换大量模型参数，这会产生巨大的通信开销[34]，动辄万计的客户端很容易对网络环境造成严重的通信负担，联邦学习的通信效率变成了限制整体训练速度的主要因素。

资源受限场景中的网络环境非常复杂，一方面是网络带宽的局限性，另一方面是网络环境的不稳定性，因此通信效率成为联邦学习在资源受限场景中应用部署的一个重要瓶颈。如表 3 所示，目前面向通信高效的联邦学习方案主要包括：减少参数传输、压缩模型梯度和减少通信轮数等方案。

4.1 减少参数传输

联邦学习中不共享客户端的本地数据，实际共享的是本地模型参数，由于不同客户端上的本地模型不同，模型参数对服务器上的全局模型的整体贡献也不相同。因此每次可以只传输贡献度高的参数，即减少不必要的参数传输，以此降低联邦学习上行或下行传输的数据量，从而实现通信高效。

为了降低通信成本，Caldas 等人[35]提出了一种联邦学习算法 Federated Dropout。首先在全局模型的子集合中选择出一个简化的子模型，然后将子模型发送给客户端，最后客户端利用子模型训练出本地模型。利用 Dropout 策略有效减少了服务器和客户端之间传输的模型参数，不但可以减少联邦学习的通信成本，还可以降低本地训练的计算成本。

基于局部表征学习，Liang 等人[36]提出了一种联邦学习算法 LG-FedAvg。该算法将局部表征学习和全局模型训练相结合，局部表征学习[37]旨在提取对训练重要的高维、紧凑的特征，每个客户端只需要学习局部表征，从而减少了模型的参数数量。实验结果表明，LG-FedAvg 降低通信成本的同时，不但取得了优越的性能，还实现了个性化模型学习。

Wang 等人[38]实现了一种客户端选择性聚合的联邦学习算法 CMFL。该算法的核心思想是减少与全局收敛不相关的本地更新，即通过为客户端提供模型更新反馈信息，每个客户端检查本地模型的更新是否与全局模型的趋势一致。通过避免将趋势不一致的模型更新上传到服务器，CMFL 可以在保证训练收敛的同时，大幅度减少联邦学习的通信开销。

Chen 等人[39]则利用最优客户端采样方案来减少联邦学习的通信开销。在每一轮通信中，所有参与的客户端都会执行本地更新，但是只有那些具有重要更新的客户端才能将模型更新发送给服务器。其中重要性用模型更新的范数来衡量，即该算法利用最优客户端采样来限制将更新发送给服务器的客户端数量，从而降低联邦学习的通信开销。

综上所述，减少参数传输方案是一种最直接和最简单的通信高效方案。然而如何衡量客户端模型参数的重要性和贡献度却成为了一个亟需解决的难题，即在保证模型公平性前提下，判断哪些模型参数是必要或不必要传输。

4.2 压缩模型梯度

联邦学习中一般利用复杂的深度神经网络模型训练，复杂模型虽然具有优越性能，但是面临高额的内存空间占用和通信资源消耗，这是联邦学习难以有效地部署在资源受限场景的一个重要原因。由于模型或梯度存在冗余，可以利用蒸馏、量化、稀疏化和二值化等策略来压缩模型或梯度，减少客户端和服务器间传输的比特数，从而实现通信高效。

表 3 面向通信高效的联邦学习

Table 3 Federated Learning for Communication Effcient

研究方法	文献	解决方案	贡献
减少参数传输	[35]	基于 Dropout 策略从全局模型的子集合中选出较小子模型	上下行通信量分别减少 28 倍和 14 倍
	[36]	利用局部表征学习提取对训练重要的高维、紧凑的特征	通信参数量减少 50%
	[38]	客户端检查本地模型的更新是否与全局模型的趋势一致	通信效率提升 13.97 倍
	[39]	利用最优客户端采样限制将本地更新发送给服务器的客户端数量	客户端参与数量减少 90%
压缩模型梯度	[40]	利用知识蒸馏压缩模型从而大幅度减少模型参数	通信开销减少 26 倍
	[41]	利用 Top-k 梯度稀疏化压缩技术实现下游压缩	通信开销减少 8.73 倍
	[43]	利用本地数据学习到一个稀疏二值掩码	通信开销减少 34.48 倍
	[44]	利用 Count Sketch 算法实现模型压缩	模型参数压缩 3.90 倍
减少通信轮数	[47]	利用最大平均偏差约束保留全局模型从而实现双流联邦学习	通信轮数减少 23.4%
	[49]	服务器周期性对从客户端接收到的模型参数进行全局聚合	上行通信量减少
	[50]	当客户端中模型两次更新的损失差异大于阈值时才更新模型	通信开销减少 11.30 倍
	[51]	雾节点实现本地聚合，服务器选择最优雾节点实现全局聚合	通信延迟减少 85%

Jeong 等人[40]实现了一种联邦蒸馏算法 FD，其中每个客户端将自己视为学生，并将其他客户端的平均模型输出视为其教师的输出。教师模型与学生模型的差异为学生提供了学习的方向，利用联邦蒸馏不但可以减少通信开销，还能降低本地训练模型的计算开销。此外为了解决数据不满足独立同分布（Independent and Identically Distributed，IID），还提出了联邦增强，其中客户端共同训练一个对抗生成模型，通过增强其本地数据以产生 IID 数据集。

在资源受限环境中，Sattler 等人[41]特意为联邦学习设计了一种稀疏三元压缩算法（Sparse Ternary Compression，STC）。STC 利用一种新的机制扩展了现有的 Top-k 梯度稀疏化压缩技术[42]，以实现下游压缩以及权重更新的三元化和最优 Golomb 编码。实验结果显示，STC 支持联邦优化向高频低比特带宽通信的范式转变，特别是在带宽受限环境中。

为了降低通信成本，提高模型的训练和推断效率，Li 等人[43]提出了一种联邦学习算法 FedMask。相比于联邦学习现有的压缩技术，FedMask 的最大区别在于利用移动设备的本地数据学习到一个稀疏的二值掩码，并且在服务器和移动设备之间仅传递二值掩码，而非本地模型的大量参数，从而降低了通信成本。此外通过将本地数据的个性化信息嵌入二值掩码来解决数据一致性问题。

为了解决通信效率和收敛速度这两个难题，Rothchild 等人[44]提出了一种新颖的联邦训练方法 FetchSGD。首先利用 Count Sketch[45]来对模型参数进行压缩，然后根据 Count Sketch 的可合并性在服务器上对模型进行聚合。由于在模型压缩和解压缩的过程中会造成一些信息的损失，导致模型在训练过程中难以收敛，因此使用了误差累积的方式来使模型能够重新快速收敛。

综上所述，压缩模型梯度方案利用模型或梯度存在冗余这一特点来减少传输的比特数，从而实现了通信高效。然而对模型或梯度进行压缩的同时，会使得模型的精确度在一定程度上受到影响。因此需要探索压缩度和精确度之间的关系，权衡两者利弊以取得联邦学习的性能最优化。

4.3 减少通信轮数

联邦学习的训练过程中，客户端从服务器下载全局模型进行本地计算，然后将本地模型参数上传给服务器。由于达到模型收敛需要进行较多轮数的联合训练，使得客户端和服务器之间的通信轮数增多，导致通信成本急剧增大。通过增加客户端的本地训练次数，即减少服务器的全局训练次数，使得客户端和服务端之间的通信轮数减少[46]，这种方案可以直接降低通信开销，从而实现通信高效。

Yao 等人[47]将最大平均偏差（Maximum Mean Discrepancy）约束[48]引入联邦学习的训练迭代中，提出了一种双流联邦学习算法。该算法使得客户端在每一轮的本地训练中保留全局模型作为参考，通过最小化全局模型与本地模型之间的损失来保证学习到其他客户端数据之外的特征，从而加快模型收敛和减少通信轮数。实验结果表明，该算法性能优于基线算法，并且通信轮数减少了 23.4%。

基于周期平均和模型量化，Reisizadeh 等人[49]提出了一种联邦训练方案 FedPAQ。FedPAQ 允许参与训练的客户端跟服务器同步之前执行本地训练，仅将活跃客户端的更新上传到服务器，然后在服务器上周期性地平均所有接收到的模型，周期性平均导致大幅减少通信轮数，从而降低训练过程的总通信开销。此外在每轮通信中只向服务器上传本地模型的量化版本，模型量化虽然降低了通信开销，但是带来了精度降低的问题。

与上述方案类似，Chen 等人[50]基于损失差异和模型量化提出了一种异步训练方案。只有当客户端模型两次更新的损失差异大于阈值时，才会更新客户端模型，使得更新频率降低和通信轮数减少，最终减少了全局通信开销。为了进一步降低通信量，在参数传输前采用基于搜索的量化方法，来寻找比特位和偏差的最优组合，实现模型参数压缩，大幅度减少了每轮通信开销。

在雾计算领域，Saha 等人[51]实现了一种联邦学习算法 FogFL。FogFL 中首先引入了雾节点作为本地聚合器，然后部分聚合来自雾节点附近边缘节点的局部更新模型，最后在边缘节点和雾节点之间经过一定数量的通信轮数和局部聚合后，中心服务器选择一个最优的雾节点进行全局聚合。该算法通过减少全局聚合轮数，不但降低了边缘节点的通信延迟和功能消耗，还提高了系统的可靠性。

综上所述，减少通信轮数方案可以大幅度的降低通信开销。与此同时，这种方案增加了客户端的本地训练，即将更多的计算任务放在客户端上进行，而这会使得客户端的计算负担急剧增加。因此需要在通信高效和计算高效之间找到一个平衡点。

5 面向计算高效的联邦学习

联邦学习中，多个客户端并行地训练本地模型，然后将本地模型整合起来得到全局模型。训练过程中存在两个关键难点，一方面是异构性，不同客户

端的计算能力不一致，落后的客户端会拖慢整个训练过程，另一方面是大模型，深度神经网络拥有庞大的模型参数，需要大量算力才能学习到一个性能强大的模型。

资源受限场景中的客户端具有异构性、轻量化和低算力等特点，这些特点使得联邦学习在计算资源受限的客户端上部署变得更加艰难，因此如何实现计算高效成为资源受限场景中联邦学习的重大挑战之一。如表 4 所示，目前面向计算高效的联邦学习方案主要包括：模型网络拆分、异步模型更新和计算资源分配等方案。

5.1 模型网络拆分

模型网络拆分的核心思想是将深度神经网络模型拆分为两部分。如图 4 所示，模型网络被拆分到客户端和服务器上分别训练，并且将大部分训练任务部署到服务器上，解决了客户端的计算资源受限导致模型难训练的问题，从而实现计算高效。

Split Learning（SL）是一种实用性的模型网络拆分算法[52]。首先 SL 将模型网络进行拆分，客户端和服务器各自保留一部分子网络，然后只需要对本地子网络进行前向或反向计算，最后联合中间结果完成整体网络的训练。利用模型网络拆分可以大幅度减少客户端的计算量，随后提出的模型网络拆分方案大都是 SL 衍生而来的变体。

SL 中，模型网络在客户端和服务器之间实现拆分，因此 SL 间接提供了比联邦学习更好的模型隐私，此外将模型网络拆分节省了客户端的计算量，使得 SL 成为资源受限场景的优化方案。然而由于跨设备训练，SL 的执行速度比联邦学习慢。Thapa 等人[53]结合 SL 和 FL 两种算法提出了一种联邦拆分算法 SFL（SplitFed Learning）。它融合了两种算法的优点，并消除了它们的缺点，此外还通过加入差分隐私来增强数据隐私性和模型鲁棒性。

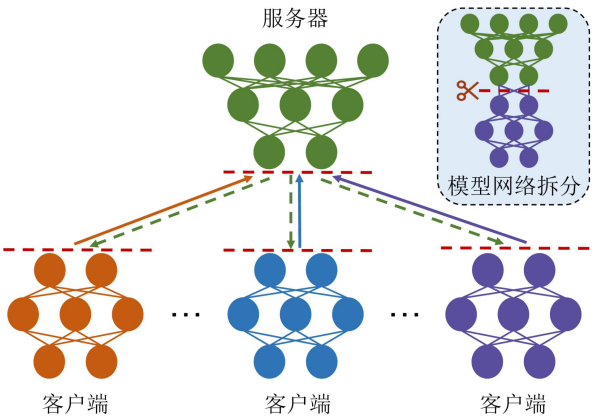

图 4 联邦学习的模型网络拆分

Figure 4 Model Split of Federated Learning

在 SL 的基础上，Park 等人[54]实现了一种面向多任务学习的联邦拆分算法 FESTA（Federated Split Task-Agnostic）。该算法将 SL、FL 和 ViT（Vision Transformer）[55]三种算法融合起来，最大限度地发挥它们各自的算法优势，同时实现了图像分类、图像分割和目标检测等多个医疗诊断任务。实验结果表明，FESTA 不但节省了客户端的大量计算资源和网络带宽，还提升了包括 COVID-19 诊断在内的单个医疗任务的性能。

表 4 面向计算高效的联邦学习

Table 4 Federated Learning for Computation Effcient

研究方法	文献	解决方案	贡献
模型网络拆分	[52]	提出了实用性的模型网络拆分算法 SL	客户端计算量降低 196.33 倍
	[53]	结合 SL 和 FL 提出了联邦拆分算法 SFL	降低客户端计算量，加快模型收敛速度
	[54]	结合 SL、FL 和 ViT 提出了多任务联邦拆分算法 FESTA	降低客户端计算量，提升多个模型性能
	[56]	结合 SL、FL 和 RL 提出了自适应联邦拆分算法 FedAdapt	降低客户端计算量，训练时间降低 57%
异步模型更新	[58]	利用陈旧性函数实现了异步联邦学习技术	加快模型收敛速度
	[59]	基于随机分布式更新提出了车联网环境下的异步联邦学习方案	加快模型收敛速度，提升数据隐私保护
	[60]	基于深度强化学习提出了车联网环境中的异步联邦学习方案	加快模型收敛速度，提升模型精度
	[61]	利用在线学习设计了异步联邦学习算法	提升计算效率，提升模型精度
计算资源分配	[62]	根据客户端资源动态地选择客户端参与并有效地执行联邦学习	减少模型训练时间
	[63]	利用深度双 Q 网络使得服务器从任何状态学习并找到最优策略	能量消耗减少 31%，训练延迟减少 55%
	[65]	将带宽资源分配给信道状态较弱或计算能力较差的设备	减少设备能量消耗
	[66]	提出一种低复杂度迭代算法来实现节能传输和资源分配	能量消耗减少 59.5%

针对资源受限的物联网设备，Wu 等人[56]提出了一种自适应联邦拆分算法 FedAdapt。FedAdapt 通过将深度神经网络从计算资源受限的物联网设备拆分到服务器上，来加速物联网设备中的本地训练速度。此外 FedAdapt 还采用了强化学习（Reinforcement Learning，RL）和聚类优化等策略，来自适应地识别

每台物联网设备应该将哪些层拆分到服务器上，以应对计算异构性和带宽时变性的挑战。

综上所述，模型网络拆分方案虽然可以降低客户端的计算开销，但是由于客户端和服务器之间不断地交换梯度和激活值，反而会大幅度增加通信开销[57]。因此在设计模型网络拆分方案时，还要额外考虑如何实现联邦学习的通信高效。

5.2 异步模型更新

异步模型更新是指客户端完成本地模型训练后，引入一些异步更新策略，无需等待其他客户端训练结束，就可以向服务器上传模型更新，并请求当前全局模型，以便快速进入下一轮训练。异步模型更新解决了客户端的差异性导致同步模型更新效率低的问题，从而实现计算高效。

为了缓解联邦学习算法在同步问题上的计算开销，Xie 等人[58]实现了一种异步联邦学习算法 FedAsync。该算法利用陈旧性函数（Staleness）自适应设定混合权重值来完成异步联邦优化过程，解决了传统联邦学习因为同步训练导致的效率低下、不可扩展和不灵活等问题，从而可以并行处理更多的客户端，大幅度提升了模型收敛速度。

在车联网环境下，Lu 等人[59]提出了一种异步联邦学习方案 DP-AFL，采用随机分布式更新机制以完成全局模型的更新。在每次迭代中，使用一个基于通信和计算资源的随机选择算法去选择一个车辆子集，模型更新仅限于该车辆子集，并且这个子集会随着车辆当前位置改变而重新选择。该方案不但解决了车辆的移动问题，还提升了模型的收敛速度。

基于区块链和强化学习，Lu 等人[60]提出了另外一种车联网环境中的异步联邦学习方案。该方案中，首先设计了一种新的混合区块链架构 PermiDAG，它由许可区块链和本地有向无环图（Directed Acyclic Graph，DAG）组成，以实现车联网中高效的数据共享，然后提出了一种异步学习方法，利用深度强化学习来选择参与设备，以最小化总代价来提升联邦学习的效率。实验结果表明，该方案提供了更高的模型精度和更快的收敛速度。

Chen 等人[61]设计了一种异步在线联邦学习算法 ASO-Fed，以解决异构客户端和落后客户端带来的不同计算负载问题。在 ASO-Fed 算法中，客户端通过执行在线学习来处理本地流数据。为了更好地捕捉客户端间的相关性，还使用了衰减系数，并通过迭代局部计算过程来平衡先前和当前的局部梯度，从而提升了联邦学习的预测性能和计算效率。

综上所述，异步模型更新方案虽然可以解决同步模型更新效率低的问题，但是由于每个客户端可以独立地向服务器申请全局模型，这会导致不同客户端获取的全局模型可能不一致，因而使得不同客户端的本地模型之间存在延迟现象。

5.3 计算资源分配

一般情况下资源受限场景是动态的和不确定的，不同客户端所拥有的计算资源存在不一致性，这会严重影响联邦学习的训练效率和收敛速度，因此计算资源分配成为了联邦学习实现高效训练的一个关键问题。如图 5 所示，实线代表计算过程，虚线代表通信过程，训练过程中需要对不同客户端进行计算资源分配，通过调度计算资源实现联合建模最优化，从而实现计算高效。

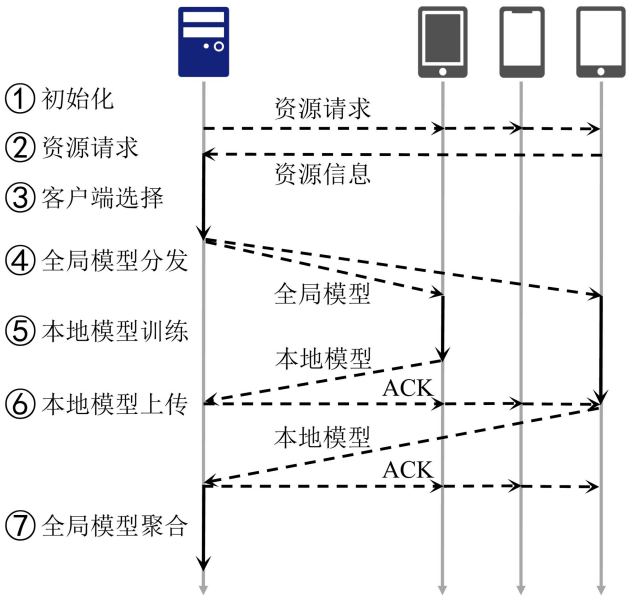

图 5 联邦学习的计算资源分配

Figure 5 Computational Resource Allocation of Federated Learning

基于客户端动态选择机制，Nishio 等人[62]提出了一种联邦学习算法 FedCS。联邦学习最初设计时利用所有客户端联合计算，没有考虑每个客户端的计算资源和信道条件。FedCS 解决了具有资源约束的客户端选择问题，它根据客户端的资源状况动态地选择客户端参与并有效地执行联邦学习，使得整个训练过程变得十分高效，并且减少了模型训练时间。

利用深度 Q 学习，Anh 等人[63]设计了一种移动群体机器学习（Mobile Crowd Machine Learning，MCML）。首先建立了 MCML 的随机化问题，然后利用深度双 Q 网络[64]构建了深度 Q 学习来实现服务器最优策略，其中深度双 Q 网络使得服务器在不具备任何网络动态先验知识的情况下学习并找到最优策略。实验结果表明，该算法在能量消耗和训练延迟方面均优于传统的联邦学习。

Zeng 等人[65]探索了联邦边缘学习的节能无线电资源管理（Radio Resource Management，RRM）。为了减少设备的能耗，设计并提出了带宽分配和设备调度的节能策略。该策略将更多的带宽资源分配给那些信道状态较弱或计算能力较差的设备，从而适应设备的信道状态和计算能力。实验结果表明，该方案在保证训练性能的同时降低了总能耗。

Yang 等人[66]研究了无线通信网络中联邦学习的节能传输和资源分配问题，并提出了一种低复杂度的迭代算法。该算法中每一步都输出时间分配、带宽分配、功率控制、计算频率和学习精度的封闭解。由于迭代算法需要初始可行解，算法中还构造了完成时间最小化问题，并基于二等分的算法来获得最优解。实验结果表明，与传统的联邦学习相比，所提出的算法减少了 59.5%的总能耗。

综上所述，计算资源分配方案不但提高了计算资源使用效率，还实现了联邦任务调度优化。但是这个方案主要面向小规模的客户端训练，一旦面对数量庞大的客户端，计算负载和计算时延会大幅度提升，因此在设计时需要提升该方案的可扩展性。

6 面向异构融合的联邦学习

由于终端设备、采集数据和应用场景的多样化问题，资源受限场景中部署联邦学习面临极其严峻的异构性挑战。首先参与训练的客户端为各种各样的智能终端设备，即设备异构性；其次不同客户端采集的本地数据为非独立同分布（Non-IID），即数据异构性；最后各个客户端的应用场景不一致导致模型参数或模型结构不同，即模型异构性。

个性化联邦学习（Personalized Federated Learning）利用个性化技术可以有效解决这些异构性挑战[67]。常见的个性化技术包括元学习、迁移学习、知识蒸馏和多任务学习等技术。如图 6 所示，在设备、数据或模型等各个方面进行个性化处理，提取个性化知识，以减轻异构性的影响，获得高质量的个性化模型。如表 5 所示，目前面向异构融合的联邦学习方案主要包括：联邦元学习、联邦迁移学习、联邦知识蒸馏和联邦多任务学习等方案。

6.1 联邦元学习

元学习（Meta Learning），又称学会学习的学习（Learning to Learn），它的目的是从已有任务中学习一种学习方法或元知识，从而实现新任务的快速学习，新任务模型是一个具有高度适应性和泛化性的新模型。元学习可以分为三种类型：基于度量的元学习、基于模型的元学习和基于优化的元学习。

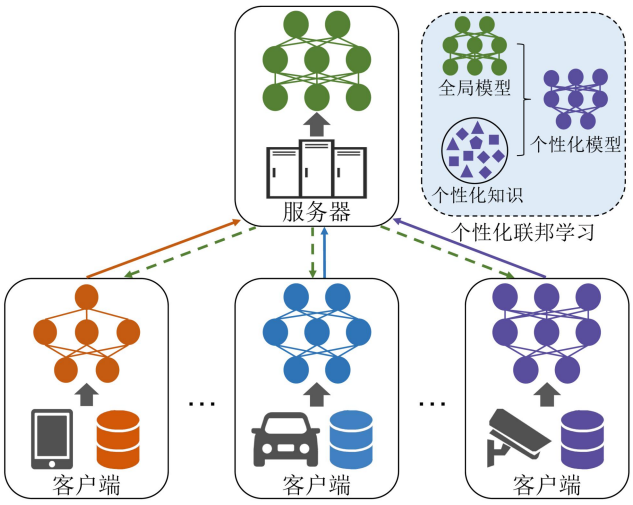

图 6 个性化联邦学习

Figure 6 Personalized Federated Learning

目前基于模型无关元学习算法（Model-Agnostic Meta-Learning，MAML）的改进算法是元学习领域的主要研究方向[68]。MAML 的目标在于学习到一个适配的参数，使得对于新任务进行微调时，能够在尽可能少的迭代更新后得到较好的性能，而对于联邦学习来说，各个客户端希望得到的全局模型在本地数据上进行微调后，能够得到一个优秀的个性化模型。由此可知，联邦学习和元学习这两种算法在本质上是相通的。

Jiang 等人[69]探讨了联邦学习和元学习的联系。一般来说 MAML 的执行过程分为元训练和元测试两个阶段，其中元训练阶段建立了多任务的全局模型，元测试阶段则将全局模型用于单独的任务。如果将联邦学习的训练过程视为元训练，而将个性化过程视为元测试，FedAvg 则可以被解释成一种元学习算法 MAML。在此基础上作者提出了一种 FedAvg 改进方案，该方案中将联邦学习分为训练和微调两个阶段。在微调阶段中，经过仔细地微调可以生成一个精度更高的全局模型，同时更易于实现更好、更稳定的个性化功能。

Chen 等人[70]设计了一种联邦元学习算法 FedMeta，可以解决移动设备分布式协作训练中存在的数据异构性和模型异构性等问题。该算法中将 MAML 算法应用于联邦学习中，以一个更灵活的方式共享元学习器，而不是像传统的联邦学习中共享全局模型，从而实现了隐私保护。数据集 LEAF 上的实验结果表明，FedMeta 在准确性、收敛速度和通信成本方面都取得了显著的改善。此外将 FedMeta 应用到工业推荐任务中的实验结果表明，每个客户端都有高度个性化的记录。

Lin 等人[71]提出了一种云平台辅助的协作学习

算法，可以实现物联网应用的实时智能决策。首先通过联邦元学习在一组源边缘节点上训练模型，然后利用云平台将模型转移到目标边缘节点，从而以模型更新实现快速适应新任务。此外研究了联邦元学习的收敛性，并检验了微调模型在目标边缘节点的自适应性。为了避免联邦元学习可能存在的漏洞以及面临对抗性攻击的脆弱性，基于分布式鲁棒优化算法，还提出了一种联邦元学习的健壮版本。

Zheng 等人[72]首次实现了一种基于联邦元学习的信用卡欺诈检测方案。由于信用卡交易数据集是非常倾斜的，欺诈交易的样本比合法交易的样本少得多，因此解决信用卡欺诈检测问题变得极具挑战性。该方案的创新点在于将客户端的模型分割为特征提取模型和关系模型两部分，首先输入数据通过特征提取模型将数据映射至低维特征空间，并通过参数学习将不同标签的数据的映射差异加大，然后利用关系模型学习输入数据之间的相似程度。由此可知，该方案中联邦元学习仅表现在客户端，服务器上仍是传统的联邦平均算法。

综上所述，联邦元学习通过优化全局模型，可以实现快速个性化，此外它只需少量的数据样本就可以快速学习和个性化适应。但是由于联邦元学习通常使用复杂的训练过程，会额外耗费一定的计算开销，在资源受限场景下会面临算力欠缺的困境。

6.2 联邦迁移学习

迁移学习（Transfer Learning）是指两个不同领域知识的一种迁移过程[73]，利用源领域中学习到的知识来提升目标领域上的学习任务。根据迁移知识形式的不同，迁移学习可以分为四种类型：基于样本的迁移学习、基于特征的迁移学习、基于模型的迁移学习和基于关系的迁移学习。

表 5 面向异构融合的联邦学习

Table 5 Federated Learning for Heterogeneous Fusion

研究方法	文献	解决方案	贡献
联邦元学习	[69]	探讨联邦学习和元学习的联系，提出了 FedAvg 改进算法	模型异构性
	[70]	将 MAML 算法应用于联邦学习中，提出了联邦元学习算法 FedMeta	模型异构性，数据异构性
	[71]	基于联邦元学习提出了协作学习算法，实现了物联网应用的实时智能决策	模型异构性
	[72]	将模型分割为特征提取模型和关系模型两部分，实现了信用卡欺诈检测方案	模型异构性
联邦迁移学习	[74]	利用源域大量标注数据完成目标域模型的训练，实现了联邦迁移学习	数据异构性
	[75]	利用迁移学习构建个性化模型，实现了用于可穿戴医疗健康的联邦迁移学习算法 FedHealth	模型异构性
	[76]	利用域自适应提取鉴别信息，实现了用于脑电图分类的联邦迁移学习算法	数据异构性
	[77]	利用迁移学习构建自定义检测模型，实现了用于物联网入侵检测的联邦迁移学习算法 IoTDefender	数据异构性
联邦知识蒸馏	[81]	利用知识蒸馏使不同客户端能够独立地训练出本地模型，从而实现了联邦蒸馏算法 FedMD	模型异构性
	[82]	将集成蒸馏应用在联邦学习的模型融合中，提出了更强大和更灵活的联邦蒸馏算法 FedDF	模型异构性，数据异构性
	[83]	对高质量的全局模型进行采样，并从贝叶斯推理的角度实现了联邦蒸馏算法 FedBE	数据异构性
	[84]	服务器学习轻量级生成器并以无数据方式集成客户端信息，实现了联邦蒸馏算法 FedGen	数据异构性
联邦多任务学习	[86]	提出了分布式优化方法 MOCHA，解决了多任务学习的高通信成本、落后设备和容错等问题	模型异构性，设备异构性
	[87]	基于星型贝叶斯网络提出了联邦多任务学习算法，解决了多任务学习仅适用于非凸模型问题	数据异构性
	[88]	提出了检测网络异常和分析网络流量的联邦多任务学习算法	模型异构性
	[89]	设计了基于混合数据分布的联邦多任务学习算法，并分别实现了中心化和去中心化算法	数据异构性

迁移学习允许在机器学习的训练和测试中域和任务是不同的，联邦学习应用部署的一大难点是资源受限场景中面临着不同数据分布或任务模型等问题，这些问题也是迁移学习要解决的，因此可以将迁移学习和联邦学习结合起来。

为了解决联邦学习的异构局限性，Liu 等人[74]提出了联邦迁移学习（Federated Transfer Learning，FTL）。FTL 特别适合处理异构数据这种联邦问题，它允许在隐私保护的情况下进行知识共享，并且允许在网络中传送互补知识，还可以利用源域大量有标注的数据实现目标域模型的训练。此外 FTL 提供了一个端到端的解决方案，并证明了该解决方案在收敛性和准确性两方面的性能与无隐私保护训练方式的性能一样。

在可穿戴医疗健康领域中，Chen 等人[75]实现了第一个联邦迁移学习算法 FedHealth。该算法利用联邦学习实现医疗数据聚合，利用迁移学习构建个性化模型。具体来说，首先利用联邦学习训练一个全局模型，然后将全局模型分发至客户端，其中每个客户端都能够通过使用本地数据来改进和细化全局模型，从而构建个性化的本地模型。同时为了减少训练开销，只对指定层的模型参数进行微调，而不是对整个模型进行再训练。实验结果表明，FedHealth 对于可穿戴设备行为识别的准确率取得 5.3%提升。

针对脑电图分类，Ju 等人[76]提出了一种联邦迁移学习算法。该算法采用单试验协方差矩阵，基于域自适应技术，从多学科脑电数据中提取共同的鉴别信息。在 PhysioNet 数据集上评估了用于二级运动图像分类的性能，在避免实际数据共享的同时，联邦迁移学习方法在一个主题自适应分析中提升了 2.0%的分类准确率。此外在缺乏多学科数据的情况下，该算法比其他最先进的深度学习算法的分类准确率提升了 6.0%。

基于联邦迁移学习，Fan 等人[77]设计了一种 5G 物联网入侵检测算法 IoTDefender，解决了物联网网络由于异构性和多样化而不能使用单一入侵检测模型的问题。IoTDefender 利用联邦学习进行网络数据聚合，利用迁移学习构建自定义检测模型，它使所有物联网网络能够在不泄露隐私的情况下共享信息。此外 IoTDefender 还具有出色的泛化能力，极大地提高对未知攻击的检测能力，与传统的机器学习方法相比入侵检测的准确率提升了 3.1%。

综上所述，联邦迁移学习适用于联邦学习中各个客户端的数据分布或任务模型不同的情况，可以有效解决数据异构性或模型异构性等难题。然而联邦迁移学习是一种复杂度更高的联邦学习，综合的性能分析后，发现大量的计算和通信开销会导致端到端性能恶化[78]，所以需要设计一些优化方案来提升联邦迁移学习的整体性能。

6.3 联邦知识蒸馏

知识蒸馏（Knowledge Distillation）旨在通过大的教师模型来提高小的学生模型的性能[79]，即将训练好的教师模型包含的知识通过蒸馏提取到学生模型里面，从而达到知识迁移或模型压缩的目的。根据教师模型是否与学生模型同步更新，知识蒸馏可以分为三种类型：离线蒸馏、在线蒸馏和自蒸馏[80]。

知识蒸馏中教师模型和学生模型的结构往往不同，可以在不同的客户端上通过蒸馏得到不同的任务模型，有效缓解了联邦学习中模型异构性问题。知识蒸馏还可以从模型中提取知识，有利于提升模型对数据的表征能力，有效缓解了联邦学习中数据异构性问题。因此将知识蒸馏应用于联邦学习中可以有效地解决异构融合这一难题。此外知识蒸馏将模型压缩成可以适配不同设备的轻量级模型，大幅度降低了模型参数量。

为了使不同客户端能够利用知识蒸馏独立地训练出自己的本地模型，Li 等人[81]实现了一种联邦蒸馏算法 FedMD。在传统的联邦学习中，客户端发送本地模型参数给服务端，然而在 FedMD 中，首先发送本地模型在公共数据集上的预测分数，然后在服务端上集成这些分数得到一个全局共识，最后每个客户端的本地模型利用知识蒸馏去学习这个全局共识。这种方法有效的解决了数据异构性问题，每个客户端可以根据自身数据训练出适合自己的模型，此外只发送预测分数给服务器，避免了发送模型参数所产生的隐私泄露风险。

通过将集成蒸馏应用在联邦学习的模型融合阶段，Lin 等人[82]提出了一种更强大和更灵活的联邦蒸馏算法 FedDF，该算法允许不同客户端上的数据和模型不一致。具体来说，针对数据异构性问题，通过利用其他领域的未标记数据或者预先训练生成器的合成数据进行集成蒸馏，针对模型异构性问题，通过对每个客户端的异构模型进行集成蒸馏来实现模型融合。这种集成蒸馏方法不但降低了隐私泄露的风险，还允许在大小、精度或结构等方面可能存在不同的异构模型上进行灵活聚合。

从贝叶斯推理的角度出发，Chen 等人[83]实现了一种新的联邦蒸馏算法 FedBE，它通过对高质量的全局模型进行采样，并通过贝叶斯模型集成将它们组合起来。首先是模型分布，客户端将本地模型上

传给服务器，服务器利用 Gaussian 或 Dirichlet 分布来构建有效的模型分布。然后是模型采样，服务器对所构建好的模型分布进行采样，得到多个集成模型。最后是模型聚合，利用知识蒸馏将多个集成模型训练为单个全局模型，并将全局模型发送给客户端。实验结果验证了在数据非独立同分布下 FedBE 算法的卓越性能。

为了解决数据异构性的难题，Zhu 等人[84]通过生成学习实现了无数据联邦蒸馏算法 FedGen。其中服务器首先学习到一个轻量级生成器，并以无数据方式集成客户端信息，然后将其广播给客户端，客户端用学习到的知识作为归纳偏差来调节本地训练。与之前研究工作中仅完善全局模型不同，FedGen 使用提取的知识来调节本地模型的更新，实验结果表明，这种学习到的知识对本地模型增加了归纳偏差，从而在客户端数据非独立同分布下，可以利用较少的通信轮数使联邦学习具有更好的泛化性能。

综上所述，联邦知识蒸馏提供了最大程度的灵活性，以适应客户端个性化模型的构建，此外该方案在通信和计算方面也具有一定的优势。然而由于知识蒸馏中教师模型和学生模型之间的模型差距，有时候学生模型难以很好地学习教师模型。

6.4 联邦多任务学习

多任务学习（Multi-Task Learning）是指同时学习多个相关任务[85]，使这些相关任务在学习过程中共享知识，利用多个任务之间的相关性来改进模型在每个任务上的性能和泛化能力。多任务学习的主要挑战在于如何设计多任务之间的共享机制，目前常见的共享模式包括三种类型：硬共享模型、软共享模式和层次共享模式。

多任务学习中一直很难避免隐私泄露的问题，引入联邦学习机制可以有效地避免隐私泄露，并且能够提升模型性能。此外在联邦学习中，中央服务器在大规模客户端上协调单个模型的训练，此设置可以自然地扩展到多任务学习，因此可以将两者结合起来。联邦多任务学习的目标是在不同客户端上同时学习不同任务，在训练过程中，首先中心服务器利用客户端上传的模型参数，学习多个学习任务之间的模型关系，然后每个客户端可以用其本地数据和当前模型更新自己的模型参数，最终每个客户端可以获得高质量的个性化模型。

Smith 等人[86]展示了联邦多任务学习是解决联邦学习中统计问题的自然选择，并提出了一种分布式优化方法 MOCHA。该方法解决了分布式多任务学习的高通信成本、落后设备和容错等问题。为了应对高通信成本的问题，MOCHA 通过执行额外的本地计算的方式造成在联邦环境下的通信轮数更少。此外通过允许参与设备周期性地退出，MOCHA 初步实现了设备的容错性。由于物联网环境中设备异构性对联邦学习的性能有巨大影响，因此联邦多任务学习对于物联网应用具有重要研究意义。

尽管联邦多任务学习被证明可以有效的处理现实世界中统计异构性的数据集，但它仅适用于凸模型情况。因此 Corinzia 等人[87]引入了 VIRTUAL，这是一种用于通用非凸模型的联邦多任务学习算法。在 VIRTUAL 中，把服务器和客户端共同构建的联邦网络视为一种星形贝叶斯网络，并使用近似变分推理进行训练学习。实验结果证明了在许多 IID 和 Non-IID 的真实数据集上，该算法的性能优于目前现有算法，即针对现实世界中的异构性数据集，VIRTUAL 是非常高效的。

一般情况下网络流量的检测和分析工作大多集中在网络异常检测这一特定任务上，无法为网络管理员提供更有价值的信息。因此 Zhao 等人[88]提出了一种联邦多任务学习算法 MT-DNN-FL，用来检测网络异常和分析网络流量。MT-DNN-FL 将模型分为输入层、共享层和任务层等三个部分，利用上传共享层参数来实现模型聚合，利用不同结构的任务层来实现多任务学习，包括网络异常检测任务、VPN 流量识别任务和流量分类任务。与使用多种单任务学习方法相比，MT-DNN-FL 可以大幅度节省训练时间，此外在 ISCXVPN2016、ISCXTor2016 和 CICIDS2017 等三个代表性数据集上的实验结果表明，该算法的检测和分类性能优于基线算法。

联邦多任务学习可以通过构造适当的惩罚项来学习个性化模型，其中惩罚项可以捕捉到个性化模型之间的复杂关系，但避免了对本地数据分布的统计假设。Marfoq 等人[89]设计了基于混合数据分布的联邦多任务学习算法，包括中心化算法 FedEM 和去中心化算法 D-FedEM，并提出了一些统计假设。假设每个客户端上的数据集是 M 个概率分布的混合分布，这样每个客户端都可以从其他客户端共享知识。此外假设每个客户端的权重是 M 个权重参数的混合分布，因此可以拓展到簇状多任务学习。实验表明该算法比现有算法具有更高的准确率和公平性。

综上所述，联邦多任务学习通过捕捉不同客户端之间的模型关系来实现个性化模型。然而由于它为每个任务都生成一个模型，所有客户端都必须参与每一轮训练，一旦客户端出现异常就会影响训练过程，所以需要解决客户端容错相关的挑战。

7 总结与展望

7.1 总结

联邦学习虽然能够有效解决数据孤岛现象和实现数据安全融合，但是面临着重大的挑战和威胁，尤其是在资源受限场景中如何高效地部署。为了突破资源受限场景中的联邦学习技术，国内外学者从架构高效、通信高效、计算高效和异构融合等四个方面提出了大量的解决方案，总结如下：

1）面向架构高效的联邦学习，包括去中心化架构和分层体系架构等方案。去中心化架构方案中，越来越多学者将区块链融入到联邦学习中，不但实现了去中心化，还弥补了各自的不足，因此这种方案是联邦学习的研究热点之一。分层体系架构方案中，借鉴了边缘计算“中心服务器-边缘服务器-边缘设备”三层架构，不但解决了过度依赖中心服务器的问题，还满足了设备大规模增长的需求，因此这种方案在资源受限场景中具有广泛应用前景。

2）面向通信高效的联邦学习，包括减少参数传输、压缩模型梯度和减少通信轮数等方案。减少参数传输方案中，不同学者提出了多个衡量指标来判断模型参数的贡献度，因此这种方案缺乏普适性。压缩模型梯度方案中，主要利用模型或梯度存在大量冗余，因此这种方案是实现通信高效的典型解决方案。减少通信轮数方案中，增加客户端的本地训练可以降低通信开销，但是计算负担急剧增加，因此这种方案在计算资源受限设备上部署存在难度。

3）面向计算高效的联邦学习，包括模型网络拆分、异步模型更新和计算资源分配等方案。模型网络拆分方案中，将大部分训练任务部署在服务器上，成功解决了客户端的计算资源受限问题，因此越来越多学者围绕这种方案展开了深入研究。异步模型更新方案中，异步更新策略解决了同步模型更新效率低的问题，但是引入了模型延迟现象，因此这种方案需要进一步改进。计算资源分配方案中，主要通过调度客户端的计算资源来实现联合训练最优化，但是海量客户端部署时会增加资源调度的难度，因此这种方案的研究重点要转向如何提升扩展性。

4）面向异构融合的联邦学习，包括联邦元学习、联邦迁移学习、联邦知识蒸馏和联邦多任务学习等方案。联邦元学习方案中，MAML 算法是元学习的主要研究方向，因此将该算法应用于联邦学习是最常用的方案。联邦迁移学习方案中，当资源受限场景中各个客户端的数据样本和数据特征很少重叠时，联邦迁移学习特别适合处理异构数据，因此这种方案是异构融合的重要解决方案。联邦知识蒸馏方案中，不但可以解决异构性问题，还可以降低通信成本，因此这种方案在资源受限场景中具有实用性。联邦多任务学习方案中，不同客户端上学习到不同任务，每个客户端都可以获得个性化模型，因此这种方案在多任务场景中具有重要研究意义。

7.2 展望

尽管联邦学习已经取得了一系列重要的研究成果，但是落地应用目前主要面向资源丰富场景，例如金融风控领域。由于存在资源匮乏和环境复杂等局限性，资源受限场景中部署应用联邦学习尚处于探索阶段。资源受限场景中联邦学习仍有许多挑战性问题亟待解决，主要包括以下几个方面：

1）联邦学习的安全性。安全问题一直是联邦学习的核心问题，安全威胁来自外部攻击和内部攻击。针对外部攻击，联邦学习已有比较成熟的防御方案，例如差分隐私[90]、同态加密[91]和安全多方计算[92]等隐私保护技术。针对内部攻击，这是一项更具挑战性的工作，目前却没有一套成熟的防御方案。此外引入隐私保护技术会增加资源开销和训练时间，这是联邦学习部署在资源受限场景中的一大难题。

2）联邦学习的高效性。联邦学习通常需要在模型精度、隐私保护、通信高效和计算高效等多个维度进行权衡和取舍。然而目前的研究工作只能在部分维度实现性能提升，例如加入隐私保护机制会额外增加计算开销，利用模型压缩来实现通信高效反而会导致精度下降。因此如何在多个维度取得综合性能提升，整体实现联邦学习的高效性，是资源受限场景中联邦学习的未来重点研究方向。

3）联邦学习的可解释性。与传统机器学习的可解释性问题类似，如何对联邦学习进行解释是一个更大的挑战[93]。例如隐私损失的可解释性，不同的安全机制会带来不同程度的隐私保护，但是随着模型隐私的提升，模型性能会降低，因此需要找到一个可解释性的机制来解释联邦学习会在多大程度上泄露数据隐私。此外由于联邦学习的复杂性，缺乏可解释性可能会导致潜在的安全威胁。

4）联邦学习的激励机制。联邦学习的目的是协同训练模型和释放数据价值，如果想训练出一个强大的联邦模型，就需要大量的智能终端设备持续提供海量的训练数据。激励机制是目前的研究热点之一，指的是如何推动更多设备参与到联邦学习的训练中，并根据贡献度分配来激励参与设备。因此激励机制对联邦学习的发展至关重要，但是在落地应用中激励机制仍不够成熟完善。

参考文献

[1] Labrinidis A, Jagadish H V. Challenges and Opportunities with Big Data[J]. *Proceedings of the VLDB Endowment,* 2012, 5(12): 2032-2033.

[2] Fan J, Han F, Liu H. Challenges of Big Data Analysis[J]. *National Science Review*, 2014, 1(2): 293-314.

[3] Bukht R, Heeks R. Defining, Conceptualising and Measuring the Digital Economy[J]. *Development Informatics Working Paper*, 2017(68).

[4] Konečnỳ J, McMahan H B, Yu F X, et al. Federated Learning Strategies for Improving Communication Efficiency[EB/OL]. 2016: ArXiv Preprint ArXiv:1610.05492.

[5] McMahan B, Moore E, Ramage D, et al. Communication-Efficient Learning of Deep Networks from Decentralized Data[C]. *Artificial Intelligence and Statistics*. PMLR, 2017: 1273-1282.

[6] Yang W, Zhang Y, Ye K, et al. FFD: A Federated Learning Based Method for Credit Card Fraud Detection[C]. *International Conference on Big Data*. Springer, 2019: 18-32.

[7] Xu J, Glicksberg B S, Su C, et al. Federated Learning for Healthcare Informatics[J]. *Journal of Healthcare Informatics Research*, 2021, 5(1): 1-19.

[8] Liu Y, Huang A, Luo Y, et al. FedVision: An Online Visual Object Detection Platform Powered by Federated Learning[C]. *Proceedings of the AAAI Conference on Artificial Intelligence*. 2020, 34(8): 13172-13179.

[9] Pokhrel S R, Choi J. Federated Learning with Blockchain for Autonomous Vehicles: Analysis and Design Challenges[J]. *IEEE Transactions on Communications*, 2020, 68(8): 4734-4746.

[10] Khan L U, Saad W, Han Z, et al. Federated Learning for Internet of Things: Recent Advances, Taxonomy, and Open Challenges[J]. *IEEE Communications Surveys & Tutorials*, 2021, 23(3): 1759-1799.

[11] Yang Q, Liu Y, Chen T, et al. Federated Machine Learning: Concept and Applications[J]. *ACM Transactions on Intelligent Systems and Technology*, 2019, 10(2): 1-19.

[12] Li T, Sahu A K, Talwalkar A, et al. Federated Learning: Challenges, Methods, and Future Directions[J]. *IEEE Signal Processing Magazine*, 2020, 37(3): 50-60.

[13] Chen B, Cheng X, Zhang J, et al. Survey of Security and Privacy in Federated Learning[J]. *Journal of Nanjing University of Aeronautics Astronautics*, 2020, 52(5): 675-684.

（陈兵，成翔，张佳乐等. 联邦学习安全与隐私保护综述[J]. *南京航空航天大学学报*, 2020, 52(5): 675-684.）

[14] Li L, Yuan S, Jin Y. Review of Blockchain Based Federated Learning[J]. *Application Research of Computers*, 2021, 38(11): 3222-3230.

（李凌霄，袁莎，金银玉. 基于区块链的联邦学习技术综述[J]. *计算机应用研究*, 2021, 38(11): 3222-3230.）

[15] Lim W Y B, Luong N C, Hoang D T, et al. Federated Learning in Mobile Edge Networks: A Comprehensive Survey[J]. *IEEE Communications Surveys & Tutorials*, 2020, 22(3): 2031-2063.

[16] Niknam S, Dhillon H S, Reed J H. Federated Learning for Wireless Communications: Motivation, Opportunities, and Challenges[J]. *IEEE Communications Magazine*, 2020, 58(6): 46-51.

[17] Nguyen D C, Ding M, Pathirana P N, et al. Federated Learning for Internet of Things: A Comprehensive Survey[J]. *IEEE Communications Surveys & Tutorials*, 2021, 23(3): 1622-1658.

[18] Imteaj A, Thakker U, Wang S, et al. A Survey on Federated Learning for Resource-Constrained IoT Devices[J]. *IEEE Internet of Things Journal*, 2021, 9(1): 1-24.

[19] 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.

[20] Wang J, Charles Z, Xu Z, et al. A Field Guide to Federated Optimization[EB/OL]. 2021: ArXiv Preprint ArXiv:2107.06917.

[21] Jiang J C, Kantarci B, Oktug S, et al. Federated Learning in Smart City Sensing: Challenges and Opportunities[J]. *Sensors*, 2020, 20(21): 6230.

[22] Khan L U, Pandey S R, Tran N H. Federated Learning for Edge Networks: Resource Optimization and Incentive Mechanism[J]. *IEEE Communications Magazine*, 2020, 58(10): 88-93.

[23] Lo S K, Lu Q, Zhu L, et al. Architectural Patterns for the Design of Federated Learning Systems[J]. *Journal of Systems and Software*, 2022: 111357.

[24] Hu C, Jiang J, Wang Z. Decentralized Federated Learning: A Segmented Gossip Approach[EB/OL]. 2019: ArXiv Preprint ArXiv:1908.07782.

[25] Baraglia R, Dazzi P, Mordacchini M, et al. A Peer-to-Peer Recommender System for Self-Emerging User Communities Based on Gossip Overlays[J]. *Journal of Computer and System*

Sciences, 2013, 79(2): 291-308.

[26] Roy A G, Siddiqui S, Pölsterl S, et al. Braintorrent: A Peer-to-Peer Environment for Decentralized Federated Learning[EB/OL]. 2019: ArXiv Preprint ArXiv:1905.06731.

[27] Pappas C, Chatzopoulos D, Lalis S, et al. IPLS: A Framework for Decentralized Federated Learning[C]. *2021 IFIP Networking Conference*. IEEE, 2021: 1-6.

[28] Benet J. IPFS: Content Addressed, Versioned, P2P File System[EB/OL]. 2014: ArXiv Preprint ArXiv:1407.3561.

[29] Kim H, Park J, Bennis M, et al. Blockchained On-Device Federated Learning[J]. *IEEE Communications Letters*, 2019, 24(6): 1279-1283.

[30] Liu L, Zhang J, Song S, et al. Client-Edge-Cloud Hierarchical Federated Learning[C]. *2020 IEEE International Conference on Communications*. IEEE, 2020: 1-6.

[31] Luo S, Chen X, Wu Q, et al. HFEL: Joint Edge Association and Resource Allocation for Cost-Efficient Hierarchical Federated Edge Learning[J]. *IEEE Transactions on Wireless Communications*, 2020, 19(10): 6535-6548.

[32] Ye Y, Li S, Liu F. EdgeFed: Optimized Federated Learning Based on Edge Computing[J]. *IEEE Access*, 2020, 8: 209191-209198.

[33] Abad M S H, Ozfatura E, Gunduz D, et al. Hierarchical Federated Learning across Heterogeneous Cellular Networks[C]. *2020 IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, 2020: 8866-8870.

[34] Bonawitz K, Eichner H, Grieskamp W, et al. Towards Federated Learning at Scale: System Design[J]. *Proceedings of Machine Learning and Systems*, 2019, 1: 374-388.

[35] Caldas S, Konečny J, McMahan H B, et al. Expanding the Reach of Federated Learning by Reducing Client Resource Requirements[EB/OL]. 2018: ArXiv Preprint ArXiv:1812.07210.

[36] Liang P P, Liu T, Liu Z, et al. Think Locally, Act Globally: Federated Learning with Local and Global Representations[EB/OL]. 2020: ArXiv Preprint ArXiv:2001.01523.

[37] Zemel R, Wu Y, Swersky K, et al. Learning Fair Representations[C]. *International Conference on Machine Learning*. PMLR, 2013: 325-333.

[38] Wang L, Wang W, Li B. CMFL: Mitigating Communication Overhead for Federated Learning[C]. *2019 IEEE 39th International Conference on Distributed Computing Systems*. IEEE, 2019: 954-964.

[39] Chen W, Horvath S, Richtarik P. Optimal Client Sampling for Federated Learning[EB/OL]. 2020: ArXiv Preprint ArXiv:2010.13723.

[40] Jeong E, Oh S, Kim H, et al. Communication-Efficient On-Device Machine Learning: Federated Distillation and Augmentation under Non-IID Private Data[EB/OL]. 2018: ArXiv Preprint ArXiv:1811.11479.

[41] Sattler F, Wiedemann S, Müller K R, et al. Robust and Communication Efficient Federated Learning From Non-IID Data[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2019, 31(9): 3400-3413.

[42] Stich S U, Cordonnier J B, Jaggi M. Sparsified SGD with Memory[J]. *Advances in Neural Information Processing Systems*, 2018, 31.

[43] Li A, Sun J, Zeng X, et al. FedMask: Joint Computation and Communication-Efficient Personalized Federated Learning via Heterogeneous Masking[C]. *Proceedings of the 19th ACM Conference on Embedded Networked Sensor Systems*. 2021: 42-55.

[44] Rothchild D, Panda A, Ullah E, et al. FetchSGD: Communication Efficient Federated Learning with Sketching[C]. *International Conference on Machine Learning*. PMLR, 2020: 8253-8265.

[45] Jiang J, Fu F, Yang T, et al. SketchML: Accelerating Distributed Machine Learning with Data Sketches[C]. *Proceedings of the 2018 International Conference on Management of Data*. 2018: 1269-1284.

[46] Konečnỳ J, McMahan H B, Ramage D, et al. Federated Optimization: Distributed Machine Learning for On-Device Intelligence[EB/OL]. 2016: ArXiv Preprint ArXiv:1610.02527.

[47] Yao X, Huang C, Sun L. Two-Stream Federated Learning: Reduce the Communication Costs[C]. *2018 IEEE Visual Communications and Image Processing*. IEEE, 2018: 1-4.

[48] Dziugaite G K, Roy D M, Ghahramani Z. Training Generative Neural Networks via Maximum Mean Discrepancy Optimization[EB/OL]. 2015: ArXiv Preprint ArXiv:1505.03906.

[49] Reisizadeh A, Mokhtari A, Hassani H, et al. FedPAQ: A Communication-Efficient Federated Learning Method with Periodic Averaging and Quantization[C]. *International Conference on Artificial Intelligence and Statistics*. PMLR, 2020: 2021-2031.

[50] Chen X, Li J, Chakrabarti C. Communication and Computation Reduction for Split Learning using Asynchronous Training[C].

2021 IEEE Workshop on Signal Processing Systems. IEEE, 2021: 76-81.

[51] Saha R, Misra S, Deb P K. FogFL: Fog-Assisted Federated Learning for Resource-Constrained IoT Devices[J]. *IEEE Internet of Things Journal*, 2021, 8(10): 8456-8463.

[52] Vepakomma P, Gupta O, Swedish T, et al. Split Learning for Health: Distributed Deep Learning without Sharing Raw Patient data[EB/OL]. 2018: ArXiv Preprint ArXiv:1812.00564.

[53] Thapa C, Chamikara M A P, Camtepe S, et al. SplitFed: When Federated Learning Meets Split Learning[EB/OL]. 2020: ArXiv Preprint ArXiv:2004.12088.

[54] Park S, Kim G, Kim J, et al. Federated Split Vision Transformer for COVID-19 CXR Diagnosis using Task-Agnostic Training[J]. *Advances in Neural Information Processing Systems*, 2021, 34: 24617-24630.

[55] Dosovitskiy A, Beyer L, Kolesnikov A, et al. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale[J]. *International Conference on Learning Representations*, 2021.

[56] Wu D, Ullah R, Harvey P, et al. FedAdapt: Adaptive Offloading for IoT Devices in Federated Learning[J]. *IEEE Internet of Things Journal*, 2022.

[57] Wang J, Qi H, Rawat A S, et al. FedLite: A Scalable Approach for Federated Learning on Resource-Constrained Clients[EB/OL]. 2022: ArXiv Preprint ArXiv:2201.11865.

[58] Xie C, Koyejo S, Gupta I. Asynchronous Federated Optimization[EB/OL]. 2019: ArXiv Preprint ArXiv:1903.03934.

[59] Lu Y, Huang X, Dai Y, et al. Differentially Private Asynchronous Federated Learning for Mobile Edge Computing in Urban Informatics[J]. *IEEE Transactions on Industrial Informatics*, 2019, 16(3): 2134-2143.

[60] Lu Y, Huang X, Zhang K, et al. Blockchain Empowered Asynchronous Federated Learning for Secure Data Sharing in Internet of Vehicles[J]. *IEEE Transactions on Vehicular Technology*, 2020, 69(4): 4298-4311.

[61] Chen Y, Ning Y, Slawski M, et al. Asynchronous Online Federated Learning for Edge Devices with Non-IID Data[C]. *2020 IEEE International Conference on Big Data*. IEEE, 2020: 15-24.

[62] Nishio T, Yonetani R. Client Selection for Federated Learning with Heterogeneous Resources in Mobile Edge[C]. *2019 IEEE International Conference on Communications*. IEEE, 2019: 1-7.

[63] Anh T T, Luong N C, Niyato D, et al. Efficient Training Management for Mobile Crowd-Machine Learning: A Deep Reinforcement Learning Approach[J]. *IEEE Wireless Communications Letters*, 2019, 8(5): 1345-1348.

[64] Van Hasselt H, Guez A, Silver D. Deep Reinforcement Learning with Double Q-Learning[C]. *Proceedings of the AAAI Conference on Artificial Intelligence*. 2016, 30(1): 2094–2100.

[65] Zeng Q, Du Y, Huang K, et al. Energy-Efficient Radio Resource Allocation for Federated Edge Learning[C]. *2020 IEEE International Conference on Communications Workshops*. IEEE, 2020: 1-6.

[66] Yang Z, Chen M, Saad W. Energy Efficient Federated Learning Over Wireless Communication Networks[J]. *IEEE Transactions on Wireless Communications*, 2020, 20(3): 1935-1949.

[67] Kulkarni V, Kulkarni M, Pant A. Survey of Personalization Techniques for Federated Learning[C]. *2020 4th World Conference on Smart Trends in Systems, Security and Sustainability*. IEEE, 2020: 794-797.

[68] Finn C, Abbeel P, Levine S. Model-Agnostic Meta-Learning for Fast Adaptation of Deep Networks[C]. *International Conference on Machine Learning*. PMLR, 2017: 1126-1135.

[69] Jiang Y, Konečnỳ J, Rush K, et al. Improving Federated Learning Personalization via Model Agnostic Meta Learning[EB/OL]. 2019: ArXiv Preprint ArXiv:1909.12488.

[70] Chen F, Luo M, Dong Z, et al. Federated Meta-Learning with Fast Convergence and Efficient Communication[EB/OL]. 2018: ArXiv Preprint ArXiv:1802.07876.

[71] Lin S, Yang G, Zhang J. A Collaborative Learning Framework via Federated Meta-Learning[C]. *2020 IEEE 40th International Conference on Distributed Computing Systems*. IEEE, 2020: 289-299.

[72] Zheng W, Yan L, Gou C, et al. Federated Meta-Learning for Fraudulent Credit Card Detection[C]. *Proceedings of the 29th International Conference on International Joint Conferences on Artificial Intelligence*. 2021: 4654-4660.

[73] Pan S J, Yang Q. A Survey on Transfer Learning[J]. *IEEE Transactions on Knowledge and Data Engineering*, 2009, 22(10): 1345-1359.

[74] Liu Y, Kang Y, Xing C, et al. A Secure Federated Transfer Learning Framework[J]. *IEEE Intelligent Systems*, 2020, 35(4): 70-82.

[75] Chen Y, Qin X, Wang J, et al. FedHealth: A Federated Transfer

Learning Framework for Wearable Healthcare[J]. *IEEE Intelligent Systems*, 2020, 35(4): 83-93.

[76] Ju C, Gao D, Mane R, et al. Federated Transfer Learning for EEG Signal Classification[C]. *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society*. IEEE, 2020: 3040-3045.

[77] Fan Y, Li Y, Zhan M, et al. IoTDefender: A Federated Transfer Learning Intrusion Detection Framework for 5G IoT[C]. *2020 IEEE 14th International Conference on Big Data Science and Engineering*. IEEE, 2020: 88-95.

[78] Jing Q, Wang W, Zhang J, et al. Quantifying the Performance of Federated Transfer Learning[EB/OL]. 2019: ArXiv Preprint ArXiv:1912.12795.

[79] Hinton G, Vinyals O, Dean J. Distilling the Knowledge in a Neural Network[EB/OL]. 2015: ArXiv Preprint ArXiv:1503.02531.

[80] Gou J, Yu B, Maybank S J, et al. Knowledge Distillation: A Survey[J]. *International Journal of Computer Vision*, 2021, 129(6): 1789-1819.

[81] Li D, Wang J. FedMD: Heterogenous Federated Learning via Model Distillation[EB/OL]. 2019: ArXiv Preprint ArXiv:1910.03581.

[82] Lin T, Kong L, Stich S U, et al. Ensemble Distillation for Robust Model Fusion in Federated Learning[J]. *Advances in Neural Information Processing Systems*, 2020, 33: 2351-2363.

[83] Chen H, Chao W. FedBE: Making Bayesian Model Ensemble Applicable to Federated Learning[J]. *International Conference on Learning Representations*, 2021.

[84] Zhu Z, Hong J, Zhou J. Data-Free Knowledge Distillation for Heterogeneous Federated Learning[C]. *International Conference on Machine Learning*. PMLR, 2021: 12878-12889.

[85] Caruana R. Multitask Learning[J]. *Machine Learning*, 1997, 28(1): 41-75.

[86] Smith V, Chiang C K, Sanjabi M, et al. Federated Multi-Task Learning[J]. *Advances in Neural Information Processing Systems*, 2017, 30: 4424-4434.

[87] Corinzia L, Beuret A, Buhmann J M. Variational Federated Multi-Task Learning[EB/OL]. 2019: ArXiv Preprint ArXiv:1906.06268.

[88] Zhao Y, Chen J, Wu D, et al. Multi-Task Network Anomaly Detection using Federated Learning[C]. *Proceedings of the 10th International Symposium on Information and Communication Technology*. 2019: 273-279.

[89] Marfoq O, Neglia G, Bellet A, et al. Federated Multi-Task Learning under a Mixture of Distributions[J]. *Advances in Neural Information Processing Systems*, 2021, 34.

[90] Geyer R C, Klein T, Nabi M. Differentially Private Federated Learning: A Client Level Perspective[EB/OL]. 2017: ArXiv Preprint ArXiv:1712.07557.

[91] Zhang C, Li S, Xia J, et al. BatchCrypt: Efficient Homomorphic Encryption for Cross-Silo Federated Learning[C]. *Proceedings of the 2020 USENIX Conference on Usenix Annual Technical Conference*. 2020: 493-506.

[92] Bonawitz K, Ivanov V, Kreuter B, et al. Practical Secure Aggregation for Privacy-Preserving Machine Learning[C]. *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*. 2017: 1175-1191.

[93] Wang G. Interpret Federated Learning with Shapley Values[EB/OL]. 2019: ArXiv Preprint ArXiv:1905.04519.

王鹏举 于 2015 年在北京邮电大学电子科学与技术专业获得硕士学位。现在中国科学院大学网络空间安全专业攻读博士学位。研究领域为计算机视觉、人工智能安全。研究兴趣包括：联邦学习、隐私计算。

Email：wangpengju@iie.ac.cn

卢江虎 于 2020 年在南昌大学材料物理专业获得学士学位。现在中国科学院信息工程研究所电子信息专业攻读硕士学位。研究领域为计算机视觉、人工智能安全。研究兴趣包括：联邦学习、模型表征与压缩。

Email：lujianghu@iie.ac.cn

葛仕明 于 2008 年在中国科学技术大学电路与系统专业获得博士学位。现任中国科学院信息工程研究所副研究员。研究领域为计算机视觉、人工智能安全。研究兴趣包括：联邦学习、低质量视觉理解。

Email：geshiming@iie.ac.cn

刘博超 于 2020 年在山东大学电子信息工程专业获得学士学位。现在中国科学院大学网络空间安全专业攻读博士学位。研究领域为计算机视觉、隐私学习。研究兴趣包括：差分隐私、隐私模型发布。

Email：liubochao@iie.ac.cn