



专题：通信与 AI 融合

人工智能在 5G 系统中的应用综述

章坚武¹, 王路鑫¹, 孙玲芬², 章谦骅^{3,4}, 单杭冠⁴

(1. 杭州电子科技大学, 浙江 杭州 310018; 2. 英国普利茅斯大学, 英国 普利茅斯 PL48AA;
3. 之江实验室智能网络研究中心, 浙江 杭州 311121; 4. 浙江大学, 浙江 杭州 310007)

摘要: 随着 5G 的不断发展, 万物互联时代即将来临。海量设备连接、海量业务请求、超高网络负载、复杂动态的网络环境等对 5G 系统优化提出了巨大的挑战。面对这些技术难点, 人工智能 (AI) 算法表现其独特的优势。首先对 5G 系统中基于深度学习的 AI 算法相比于传统算法的优势进行介绍; 随后, 针对多接入边缘计算和毫米波大规模多输入多输出 (mmWave massive MIMO) 系统中的 AI 算法应用进行详细的阐述, 并对比分析了各种方法的优劣; 最后, 根据已有研究, 总结了 AI 算法在 5G 实际场景中存在的不足, 并对未来研究方向提出了展望。

关键词: 5G; 人工智能; 强化学习; 边缘计算; 毫米波大规模 MIMO

中图分类号: TN929

文献标识码: A

doi: 10.11959/j.issn.1000-0801.2021109

An survey on application of artificial intelligence in 5G system

ZHANG Jianwu¹, WANG Luxin¹, SUN Lingfen², ZHANG Qianye^{3,4}, SHAN Hangguan⁴

1. Hangzhou Dianzi University, Hangzhou 310018, China

2. University of Plymouth, Plymouth PL48AA, UK

3. Research Center for Intelligent Network, Zhejiang Lab, Hangzhou 311121, China

4. Zhejiang University, Hangzhou 310007, China

Abstract: With the continuous development of 5G, the era of the internet of everything is coming. Problems such as massive device connections, massive application requests, ultra-high network load and complex dynamic network environment pose great challenges to the optimization of 5G systems in the context of the internet of everything. Facing these challenges, artificial intelligence (AI) shows its unique advantages. Firstly, the advantages of deep learning driven AI algorithms in 5G system compared with conventional algorithms were briefly introduced. Then, the application of AI algorithms in multi-access edge computing (MEC) and mmWave massive multiple-input multiple-output (MIMO) system were described in detail, with advantages and disadvantages of each method being compared and analyzed. Finally, according to the existing research, the shortcomings of AI algorithms in 5G application scenarios were summarized and the future research directions were forecasted.

Key words: 5G, artificial intelligence, reinforcement learning, edge computing, mmWave massive MIMO

专题组稿专家: 赵慧玲, 章坚武, 张同须

收稿日期: 2021-04-02; 修回日期: 2020-05-15

通信作者: 章坚武, jwzhang@hdu.edu.cn

基金项目: 国家自然科学基金资助项目 (No.U1866209, No.61772162); 英国皇家学会国际合作交流项目 (No.IEC\NSFC\181300)

Foundation Items: The National Natural Science Foundation of China (No.U1866209, No.61772162), The Royal Society, International Exchanges 2018 Cost Share (China) Grant (No.IEC\NSFC\181300)

1 引言

随着移动业务量的不断增加,互联网时代对移动通信技术的需求越来越紧迫。5G 提供了一种高度灵活和可扩展的网络技术,使得人与人、人与物、物与物在任何时间、任何地点都能实现相互连接。它提供了一个弹性的云原生核心网络,对网络切片提供端到端的需求。围绕 5G 的三大应用场景——增强型移动宽带(eMBB)、大连接物联网(mMTC)、超可靠低时延通信(uRLLC)将为未来的社会、经济发展提供巨大动力^[1]。在未来 5G 技术的支持下,越来越多的相关技术(例如物联网、车联网、远程医疗、虚拟现实与增强现实、无人驾驶飞行器、智慧交通、工业 4.0 等)将会从根本上改变人类生活。为应对这种发展趋势,无线通信系统必须能够具备支持超高网络负荷、海量设备接入与低时延大连接的能力^[2-3]。然而大量的数据处理和计算将极大地提高计算复杂度,同时灵活的空中接口和网络切片技术的引入也将对无线系统设计提出巨大挑战。

人工智能(artificial intelligence, AI)作为一种新型技术,已经在图像处理、自然语言识别、电子游戏等领域有着广泛的应用^[4-6]。为解决 5G 系统存在的问题与挑战,学术界开始把研究目光放在人工智能技术上。人工智能技术具备强大的学习和预测能力、良好的非线性拟合和自适应能力,善于挖掘高纬度和高冗余数据中复杂的特征并进行高效的海量数据处理^[7]。现有研究表明,针对 5G 系统中存在的难以建模、难以求解或具有统一模式的问题,现有的传统解决方案并不能起到良好的作用,然而人工智能技术可以通过其本身具备的学习预测和自适应能力以及挖掘数据特征的能力,帮助理解无线内容,识别未知模式,降低系统复杂度,并产生超越传统解决方案的实际效果^[8]。

2 人工智能在移动边缘计算领域的研究

随着移动终端数量的激增,现有中心化网络

逐渐无法支持各终端的任务需求。多接入边缘计算(multi-access edge computing, MEC)技术在网络边缘部署具有缓存、存储、计算能力的资源节点,通过将业务本地化、缓存加速、本地分流等技术,有效降低了网络回传的带宽需求,有效减轻了核心网在数据传输方面受到的压力,应对 5G 网络面临的高负荷、高带宽、低时延^[9]。关于 MEC 的研究主要集中在 3 个方面。

(1) 计算卸载与资源分配

在 MEC 技术中,计算卸载扮演了一个重要的角色。计算卸载策略决定了 MEC 系统在效率与可实现性中的表现。通过将计算任务完全或部分卸载至物理位置上接近且计算能力合适的 MEC 服务器,用户对任务计算的体验(例如设备能耗、计算时延等),将会有显著提升^[10-11]。目前,学术界在计算卸载量与资源分配问题上,使用的方法主要围绕启发式算法和博弈论。然而在热点区域,例如商场、学校等大型人员流动密集或存在大量终端设备的场所,若继续使用此类算法求解最佳卸载策略,将会产生较高的计算复杂度。并且对不同用户产生的不同类型任务,此类算法大多只能使用统一的卸载标准,难以给出个性化卸载策略。

(2) 移动性管理

MEC 移动性问题包括终端用户的位置移动导致的数据到应用的路径更改以及由于负载均衡或计算节点性能比不足产生的应用迁移^[12]。由于终端用户移动轨迹难以预测且 5G 基站覆盖范围较小,切换将频繁且无规律发生。这对 5G 系统的移动性管理提出了极高的要求。另外,由于网络边缘部署的小型计算节点计算能力有限,应用迁移也会时而发生。

(3) 安全性问题

传统的无线网络中,重要的网元设备均被部署在核心网机房内,并且处于一个相对封闭的环境中,安全性能得到良好的保障,主要安全问题仅存在于接入网中。MEC 系统中,业务具有本地化的特性,数据在进入核心网前就可能被终结,



系统极易受到攻击^[13]。因此，增加网络安全建设（例如入侵检测）变得尤为重要。

2.1 基于 AI 的计算卸载策略

计算卸载与资源分配策略决定了 MEC 系统的性能。针对工业物联网（industrial internet of thing, IIoT）中机器型代理（machine type agent, MTA）计算卸载与资源分配问题，参考文献[14]提出了一种基于多代理深层确定性策略梯度（MADDPG）的深度学习算法，对系统时延进行了优化。该算法只在执行阶段使用本地信息，保证了 IIoT 在大连接场景下的安全性。由于该算法执行时不需要代理间的相互通信，参考文献[15]提出了一种物联网环境中基于深度神经网络（deep neural network, DNN）的分布式深度学习驱动任务卸载算法。该研究对时延与能耗进行加权操作，求得的最佳卸载策略兼顾了移动业务的传输时延与设备能耗，在物联网场景下具有积极意义。此外，该算法利用均方误差函数最小化损失函数值，得到最佳卸载策略，实现系统整体开销最优化。参考文献[16-17]提出了一种基于分层式机器学习的 MEC 任务分配框架。与以往研究不同的是，该框架不仅在 MEC 服务器端使用 DNN，同时在传感器设备端部署浅层神经网络（shallow neural network, SNN），当数据传入时会先在本地根据 SNN 进行卸载策略处理，随后将需要卸载的数据上传至 MEC 服务器进行更精确的计算。这种架构的优势在于，当存在大量数据时，MEC 服务器的计算负荷将被本地分担，显著降低数据处理时间。劣势在于系统对数据处理的准确度将会有所降低。

参考文献[18]针对均衡时延与能耗问题，提出了一种基于深度强化学习（deep reinforcement learning, DRL）的算法架构。与其他现有研究相比，这种结合式的深度强化学习弥补了基于深度学习（deep learning, DL）模型需要大量已打标签的训练数据的缺点，同时也弥补了大规模 MEC 场景下基于强化学习（reinforcement learning, RL）

模型无法实现全覆盖且最终结果不稳定的劣势。

参考文献[19]针对多用户、多信道、多 MEC 场景提出了一种基于 DDPG（deep deterministic policy gradient）算法的计算卸载与资源分配策略。该策略不仅考虑用户时延与能耗的均衡，同时为每个用户设置了一个系统开销上限。这样操作相较于其他同类研究，系统整体的 QoE（quality of experience）将会提高，然而对于系统中的某个个体来说，所得到的卸载策略也许不是最佳卸载策略。另外，DDPG 算法能够处理优化变量的连续动作空间，以此解决兼顾系统时延、能耗以及各用户开销上限的问题。

参考文献[20]提出了一种基于深度强化学习的在线卸载算法，旨在最大化无线设备的计算速率。该算法由卸载动作生成和卸载决策更新两个交替过程组成。卸载动作的生成依赖于 DNN 的使用，被选择的卸载动作将会存放在重放内存中。在卸载决策更新过程中，重放内存中的一部分训练样本会被提取并重新用于 DNN 训练。这样操作的好处是：分批处理更新策略会大大降低系统复杂度；历史数据的重新使用减少了迭代更新过程中卸载策略参数化后的方差；随机抽样能够通过降低训练样本中的相关性加速收敛速度。

参考文献[21]针对“边-云”协同网络，设计了一个基于卷积神经网络（convolutional neural network, CNN）计算卸载方案，旨在降低系统时延。由于 CNN 是由多个层组成的，每个层都具有不同的数据处理能力，系统可以调度多个 CNN 层，在一个边缘服务器上处理一个任务的一部分，在其他服务器上执行剩余任务，实现部分卸载。但是当计算节点增加，CNN 网络层数会加深，在这种情况下采用 BP 传播修改参数会使靠近输入层的参数改动较慢，因此在大连接场景下性能会有所下降。

参考文献[22]针对车联网场景提出了一种基于 DQN（deep Q network）的多用户计算卸载与资源分配方案，考虑到任务执行时延与能耗的均衡，实现了任务总成本的下降。相较于其他方法

而言, DQN 能够在没有任何先验信息的情况下使车辆与环境进行交互, 并从环境中学习、调整卸载决策, 从而实现长期回报最佳化。这对类似车联网这种时变动态的场景来说非常重要。参考文献[23]在参考文献[22]的基础上, 对不同车辆进行了任务优先级的划分, 提出了一种基于 AHP-DQN 的计算卸载与资源分配算法。该算法通过引入层次分析法 (AHP) 对车辆终端的任务进行预处理, 使得 MEC 服务器能够轻松地合理分配计算资源。同时重新定义了奖励函数, 并且对任务处理速率进行加权作为新的优化目标, 降低了算法在迭代过程中的计算复杂度。

上述基于人工智能的计算卸载算法特点见表 1。

2.2 基于 AI 的边缘移动性管理

在 MEC 系统中, 移动性管理的方案主要有两种: 第一, 用户移动性较低的场景中, 可以通过调整基站功率, 保障用户处于移动状态时能够接受到连续不中断的服务, 该方案被称作功率控制方案; 第二, 当用户从一个基站覆盖区域切换到一个新的基站覆盖区域时, 计算节点间通过迁移虚拟机保障服务的连续性, 并且为用户选择一条新的合适路径与计算节点相连, 该方案被称作虚拟机迁移方案。在功率控制方案中, 传统的通过基站自适应式功率控制方案已经可以有效地保障

低移动性场景服务的连续性, 卸载应用成果传递有效率达 98% 以上^[24-25]。由于 5G 系统中基站覆盖范围较小, 用户的切换会频繁发生, 特别是车联网等终端处于高速移动的场景, MEC 系统需要及时作出正确的服务迁移。MEC 场景往往存在多个终端用户与多个 MEC 计算节点, 因此可选择的服务迁移路径较多, 例如当终端用户数为 5、边缘计算节点数为 3 时, 可选迁移路径数为 $5^3 = 125$ 。随着终端与计算阶节点的增加, 可选迁移路径将以指数形式增加, 这使得传统算法的计算复杂度急剧增加。在这种情况下, AI 技术成为一个研究重点。

参考文献[26]针对能耗敏感型的移动性用户提出了一种基于 DNN 的 EESA (energy efficient smart allocator) 算法。对每个用户所在的初始位置建立了距离矩阵, 并根据服务器的总 CPU 数、RAM 和磁盘空间及其各项的加权系数求得每一项对用户 QoS 的权重, 最终得出系统效用矩阵。最后利用三层循环神经网络 (recurrent neural network, RNN) 对效用函数值进行训练, 求得最佳卸载策略。现有对 MEC 系统能耗的研究主要集中在服务器 CPU 的分配, 而该文将 CPU、RAM 和磁盘空间共同纳入考量, 更符合 MEC 系统的实际场景。

参考文献[27]围绕深度强化学习提出了一种

表 1 各算法优化目标与卸载方式

人工智能算法	优化目标	卸载方式	参考文献
MADDPG	时延、信道利用率、信道成功接入率	全卸载	参考文献[14]
DNN	兼顾时延与能耗	全卸载	参考文献[15]
深度强化学习	兼顾时延与能耗	部分卸载	参考文献[18]
DNN、SNN	时延、任务错误率	部分卸载	参考文献[16-17]
improved DNN	计算速率	全卸载	参考文献[20]
DDPG	兼顾时延与能耗	全卸载	参考文献[19]
CNN	时延	部分卸载	参考文献[21]
DQN	兼顾时延与能耗	全卸载	参考文献[22]
AHP-DQN	兼顾时延与能耗	全卸载	参考文献[23]



新的动态移动性感知资源分配方案。由于在移动场景下，每个 MEC 计算节点中的任务量是动态的，会因为用户的切换而迁移到别的计算节点中，因此，通过传统方法找到动态网络的最优卸载策略是非常困难的。该文引入软件定义网络（software defined network, SDN）架构，通过 SDN 提供灵活、可靠的可用资源实时信息，集中控制器可以让系统中的每个单元做出最佳决策。该文在 SDN 控制器中部署了智能代理，代理通过 DQN 预测的 Q 值做出当前状态下用户的卸载策略。与现有研究不同的是，该文考虑了计算节点在计算期间的时变工作负载，当计算资源改变时，任务可以实时进行迁移，避免了不必要的排队时间，进一步最大化了系统增益。

参考文献[28]针对高速移动的车联网场景，提出了基于 DRL 的卸载规划方案。该方案能够最小化任务执行时延和系统能耗加权后的系统开销。在高速移动的车联网场景下，各终端任务特性不同且处于高速移动状态，此外，无线环境是动态变化的，因此需要考虑任务是否被卸载以及任务执行顺序。针对这种场景，该文通过马尔可夫决策过程（Markov decision process, MDP）和 DRL 进行建模，处理频繁切换带来的大量状态空间。并使用 PPO（proximal policy optimization）算法进行训练，该算法被证明具有良好的稳定性和可靠性^[29]。在提取任务队列特征方面，该文使用在 DNN 体系结构中嵌入 CNN 的方式获取近似的卸载策略和值函数。该文设计的方案能够在没有任何动态环境先验信息的情况下有效地学习最佳卸载调度策略。但是该文考虑的移动场景是一维移动场景，即车辆仅在直线上高速行驶，在实际场景中效果将有所降低。

参考文献[30]提出了移动感知深度强化学习（mobility-aware deep reinforcement learning, M-DRL）框架解决 MEC 系统中的移动服务提供问题。相较于其他研究，该文对终端用户的轨迹进行了预测。移动性预测是在移动服务提供过程

中实现高效服务迁移的关键点，如果能够预测未来用户在服务器间的一系列位置与时间，那么就可以攻击未来用户所经历的预期时延。为实现此目标，该文提出了一种称为 glimpse 的 seq2seq 模型进行轨迹预测，其中编码器与解码器使用长短期记忆（long short-term memory, LSTM）网络。由于多用户多 MEC 场景下状态空间维度较大，该文在 DRL 中使用神经网络来逼近 Q 值函数。最后结合 glimpse 行动预测模型与 DRL，形成 M-DRL 框架，从而降低系统的计算复杂度并以接近最优的性能处理 MEC 中的服务提供问题。

参考文献[31]针对车联网场景提出了一种基于 DRL 的卸载算法。与其他研究不同的是，该文不仅考虑了用户本地执行、任务卸载至边缘服务器执行，同时也考虑了用户与用户间相互卸载的情况（即 D2D）。这种做法的优势在于能够利用邻近空闲终端的计算能力，在提升系统效率的同时，提升频谱利用率。针对状态空间较大引起的高系统复杂度问题，该文提出了一种基于 continuous-control 的算法来对动作输出进行修正。同时，该文在 DRL 中使用了一种被称作 A3C 的改进 Actor critic（AC）算法来确定最优动作。AC 算法由两个结构不同的神经网络组成，可以进行单步更新，相较于传统的 Policy Gradient 回合更新要快，在学习阶段的学习效率相较于 DQN 算法更加高效。

2.3 基于 AI 的 MEC 安全性

由于 MEC 本身的特性，用户本地业务的数据将会在核心网之外被终结，攻击者将能够通过边缘平台或应用攻击核心网，造成敏感数据的泄露、DDoS 攻击等。边缘安全威胁主要存在于：（1）基础设施安全，例如，攻击者能够非法访问物理服务器的 I/O 接口，从而获取诸如用户位置等敏感信息；（2）MEC 平台与 MEC 应用安全，例如当 MEC 平台与 MEC App 通信时，传输数据可能会遭到篡改、拦截、重放；（3）MEC 编排和管理系统安全；（4）用户面功能（UPF）安全等^[32]。

由于无线信道具有开放广播的特性,边缘网络在对抗恶意终端设备的假冒攻击方面十分薄弱。攻击者可以通过合法的身份与 MEC 设备建立连接,这也将进一步导致中间人攻击和 DDoS 攻击的出现。对此,参考文献[33]设计了一种基于 DQN 的物理层假冒攻击检测方法。该文根据建立的假冒攻击模型,在接收侧建立基于信道状态信息(channel state information, CSI)的假设检验,随后把两次连续的 CSI 间的欧几里得距离作为检验统计量,并利用 DQN 比较不同阈值的预期效果,筛查具有不合法特征的发送端身份。由于边缘系统中用户并不总处于静止状态,MEC 与用户间的信道环境也是动态的,利用 DQN 算法能够在动态环境中找到最优卸载策略。同时利用物理层特性的检验方法几乎不会为终端带来计算负荷。参考文献[34]提出了一种孤立随机森林算法与 MEC 技术相结合的思路对大数据环境下的 DDoS 攻击进行检测。将算法的训练任务分散到各个边缘计算节点,模型的训练速度得到了大幅度提升。将中心云的训练任务分流到各个边缘节点,有效地减少了系统在检测阶段的耗时。通过该方法,MEC 系统能够在保证高检测性能的同时,有效应对网络数据增长带来的中心云检测压力。参考文献[35]提出了一种基于 DL 的流量预测与数据卸载机制和网络攻击检测(DLTPDO-CD)技术,用于在 MEC 系统中预测流量、卸载数据和检测网络攻击。该文首先使用 Bi-LSTM 进行流量负荷预测,这一过程有助于实现有效的数据卸载;其次将 ASCE (adaptive sampling cross entropy) 模型应用于最大化卸载用户的吞吐量;随后将 DBN (deep belief network) 模型作为网络攻击检测器,通过 BMO (barnacles mating optimizer) 算法优化 DBN 模型的学习率;最后,应用 softmax 分类层执行分类任务。实验结果证明,该模型对 MEC 系统中网络攻击的检测效率提升显著,有效地保障了 MEC 系统的安全性。参考文献[36]在 IoT 场景下提出了一种基

于生成对抗性网络(generative adversarial network, GAN)和深度 Q 网络(DQN)的代理检测安全机制,旨在从连接信息中检测未经请求的代理连接。这种 DQN-GAN 安全机制既有 DQN 对复杂随机环境适用性,也具有 GAN 的强鲁棒性,对于提升大数据量的 IoT 场景安全性有很好的帮助。

另外,在 MEC 卸载的传输过程中,用户的位置信息、移动状态、兴趣偏好和物理数据等敏感信息暴露在攻击者面前,这大大降低了 MEC 系统的隐私性。为解决 MEC 系统隐私性问题,参考文献[37]基于随机森林分类器和 k -匿名算法,提出了一种有效保证敏感数据隐私的分类-匿名模型。随机森林在许多现有算法中获得了较高的精度,使得它非常适合应用在 IoT 场景的海量复杂数据处理。在边缘环境下,根据用户特性可以将数据属性分为公共属性与个体属性。根据随机森林分类完毕之后, k -匿名算法会对敏感数据进行处理。最后通过构建 Third-Game 模型,分析博弈双方在交互阶段引起的用户敏感数据泄露问题。随机森林是在随机性的帮助下由多个决策树组成的,能够较好地对抗由大量不平衡的数据带来的平衡误差,收敛到一个较小的泛化误差,因此 CA 模型能够更好地进行样本预测,保护用户隐私数据。

参考文献[38]针对物联网场景提出了一种基于深度 PDS-learning 的隐私感知计算卸载策略。该文构建了一个与用户-节点间距离相关的隐私损失函数,用其表征系统安全性,并将隐私感知卸载问题描述为一个马尔可夫决策过程。PDS-learning 利用额外的先验统计信息加速 MDP 的学习过程,该文通过将 PDS-learning 与 DQN 相结合,结合两种技术的优点,有效解决了空间复杂性高和学习速率慢的问题。该算法能够使物联网设备以更快的速度学习一个良好的隐私感知卸载策略,相较于传统 DQN 算法有明显优势。

参考文献[39]针对边缘计算中的隐私泄露问题提出了一种新型的双时间尺度 RL 算法。在较



慢的时间尺度上, 该算法与传统 PDS-VE 算法相似; 在较快的时间尺度上, 该算法通过计算估计过渡概率学习未知的先验信息。当状态转换的先验信息可用时, 与其他传统的 RL 算法相比, 传统的 PDS-VE 算法可以实现更快的学习。然而在实践中, 移动设备可能无法获得先验信息。该文提出的双时间尺度 RL 算法可以使移动设备在缺失先验信息的情况下, 以与 PDS-VE 算法相似的速度学习最优的隐私感知卸载策略。

参考文献[40]构建了雾辅助三层安全计算体系结构, 其中开发了一种隐私感知的数据处理机制以对抗窃听攻击。该机制首先将公钥和私钥由受信任的认证方分别分配给云和终端设备, 并在传感层进行加密处理; 卸载部分传输到雾层, 并根据 Paillier 加密的加性同态性质分别对两部分聚合密文进行定义; 云层将两部分聚合结果进行二次聚合, 得到最终聚合结果; 最后在云层进行解密。这种方式使得窃听者即使成功窃听了通信链路中传输的数据, 也只能获得相应的密文, 从而在传输阶段有效地保护了用户的隐私信息。

参考文献[41]从安全服务开销的角度出发, 对 MEC 系统的安全性进行了研究。该文通过建立一个定量的模型来度量安全开销与安全数据大小、CPU 周期数、节点工作频率之间的关系。并通过遗传算法对系统安全性和能耗进行优化, 从而实现在确保系统安全性的前提下, 计算能耗最低。

3 AI 与毫米波大规模 MIMO 结合

与 4G 相比, 5G 网络的目标包括: 每个地理区域的移动数据流量增加 1 000 倍, 用户数据增加 100 倍, 连接设备增加 100 倍, 网络能耗降低到 1/10, 端到端时延减少到 1/5^[42], 这要求 5G 系统拥有更高的数据传输速率^[43]。针对此问题, 学界确定了 3 种方法: 部署更多的基站; 发掘更多带宽资源; 使用更多天线数量的 MIMO 系统。这 3 种方法对应产生了超密集组网技术、毫米波技术和

大规模 MIMO 技术。毫米波大规模 MIMO 技术一方面结合了毫米波技术可用带宽大的前景, 另一方面从大规模 MIMO 技术中获取大量收益, 能够支持 5G 系统中大量高速服务和带宽需求。

尽管毫米波大规模 MIMO 具有巨大的潜力, 但在物理层方面, 毫米波信道对抵抗阻塞和传播损耗非常敏感, 尤其是在室内或建筑物密集的城市环境中, 这种现象更为明显^[45]。

3.1 AI 结合混合预编码

为提升毫米波系统的传输容量和可靠性。现有的解决方案大致有全连接混合预编码^[46-52]、子连接混合预编码^[53-54]、自适应连接混合预编码^[55-58]。然而, 现有混合预编码方案的基本局限性是计算复杂度高, 不能充分利用空间信息。随着 AI 技术在 5G 系统中的应用越来越广泛, 研究人员开始关注结合 AI 的混合预编码方案, 从而在提升系统性能的同时, 降低系统复杂度, 提高鲁棒性。

3.1.1 全连接混合预编码方案

在全连接混合预编码方案中, 每个射频链固定地连接到所有天线上, 通过设计适当的预编码方案就可以实现较高的频谱效率。全连接混合预编码收/发两端结构如图 1 所示。

参考文献[59]在编码器中使用并行复卷积块, 在解码器中使用全连接层, 信道参数使用复杂且真实的值。将这种复杂的信道参数作为输入应用到卷积神经网络中, 卷积运算在本质上也将变得复杂, 从而帮助训练网络达到更高的精度。但是这种做法的缺陷在于训练时间比较长, 算法效率有所降低。参考文献[60]提出了一种基于 DNN 的混合预编码方案, 每一次获得更优解码器的预编码器的选择被视为 DNN 的映射关系。在训练过程中, DNN 能够获取预编码方案的结构信息, 有助于降低计算复杂度。但由于 DNN 本身存在的问题, 该方法的精确度相较于参考文献[59]的方案略显逊色。参考文献[61]提出了 HPNet 结构, 如图 2 所示, 该结构包括三层卷积层和四层全连接层,

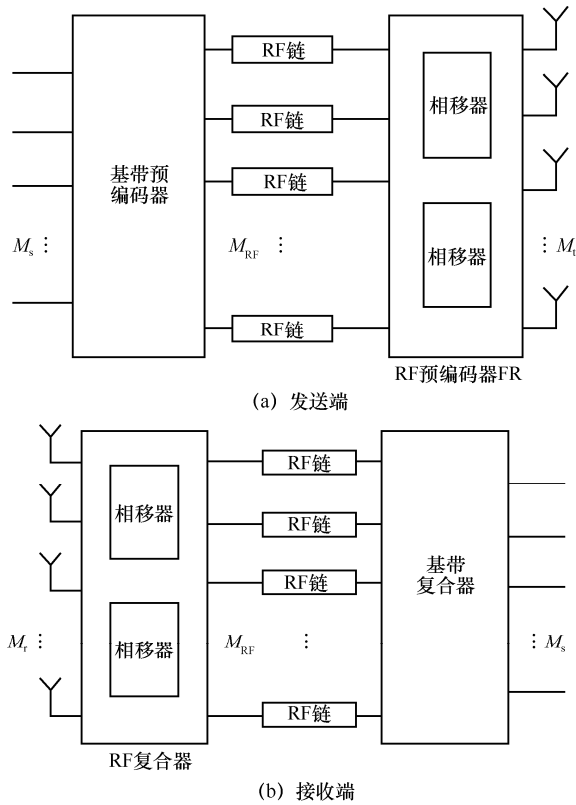


图1 全连接混合预编码收发端结构

其中，三层卷积层用来提取属于毫米波信道的通信特征，每个卷积层后面是批标准化(batch normalization, BN)层和一个整流线性单元，通

过这种结构避免训练过程中的过拟合和梯度消失现象；四层全连接层实现了非线性映射。实验结果表明，该方法的复杂度明显优于非AI混合预编码的方法，但是在频谱效率方面略逊于现有方法。

3.1.2 子连接混合预编码

子连接混合预编码中，每个射频链固定连接在天线的某一部分上，可以比全连接结构消耗更少的功率，但会损失一定的频谱效率。针对该种预编码存在的缺陷问题，参考文献[62]提出了一种基于深度学习量化相位(deep learning quantized phase, DLQP)的混合预编码方法，如图3所示。在预编码前使用了一种基于深度学习压缩感知的信道估计方案得到估计信道向量。在预编码阶段，首先使用估计信道向量训练混合预编码神经网络(training hybrid precoding neural network, THPNN)，其中考虑了近似相位量化；随后估计信道向量被输入部署混合预编码神经网络(deployment hybrid precoding neural network, DHPNN)中，从而得到模拟预编码器向量；最后将所有用户的模拟预编码向量叠加得到模拟预编码矩阵，通过迫零算法得到数字预编码矩阵。

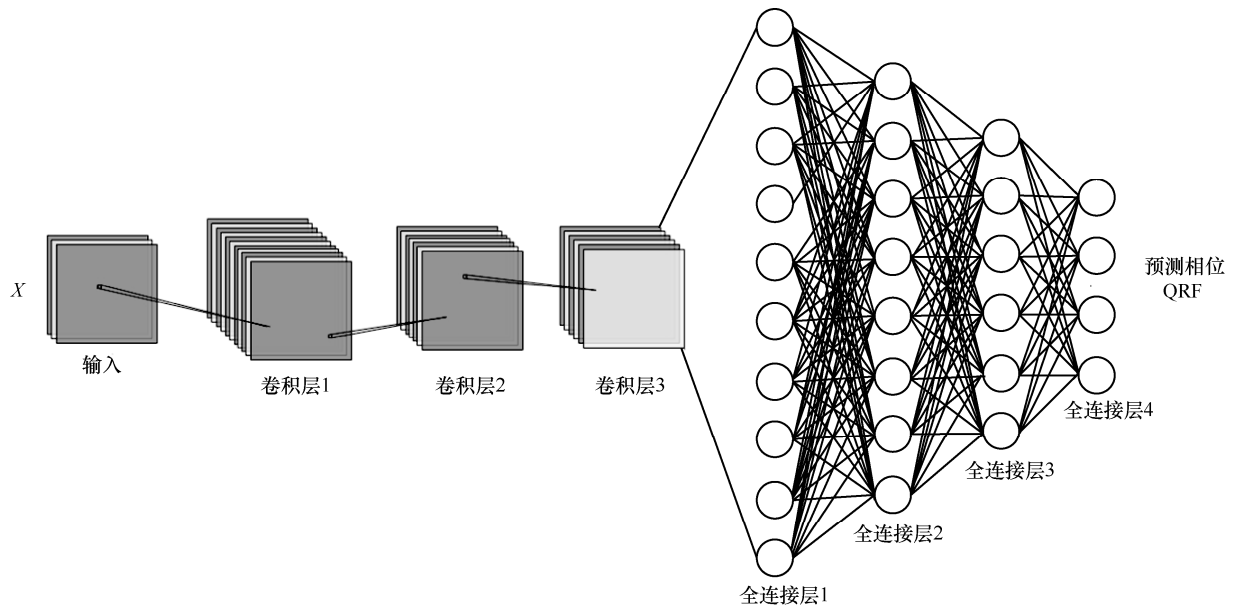


图2 HPNet 结构

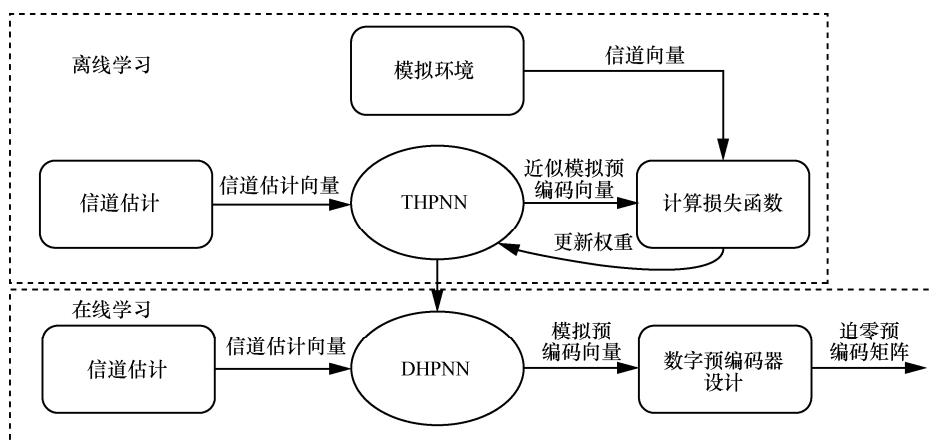


图3 基于 DLQP 的混合预编码方法

实验结果表明，该方法一定程度上减轻了由于子连接结构带来的频谱效率损失问题。

参考文献[63]针对实际场景中多路发射信号到达接收机存在相对时延，当多路信号叠加时会产生频率选择性衰落问题，该文运用交叉熵优化的思想智能搜索最优的模拟预编码器，同时以量化比特的相移器代替高精度相移器，从而实现能源效率方面的优化。实验结果表明，该方法可以接近最优的可达和速率，并具有较高的能量效率。

上述这些种方案考虑的子连接形式为固定子阵列，这种连接形式中，每个射频链连接到固定天线子集，虽然拥有低功耗和低复杂度的特性，但是在一定程度上牺牲了系统性能。参考文献[64]中设计了针对多用户的动态子阵列结构，如图4所示，将两个DNN集成到多用户混合预编码框架，并通过映射操作和训练解决了多用户选择和动态子阵列划分问题。DNN中隐含层的非线性变换具有拟合未知复杂函数的能力，因此即使在多用户动态子阵列结构中也能够保证较低的复杂度。实验结果表明，该方法具有较好的可达和速率。

3.1.3 自适应连接混合预编码

自适应连接结构利用射频链与天线之间的动态变化进行连接，在实现混合预编码方面与其他连接方式相比具有良好的灵活性，可以实现更好的波束增益。但是这种连接方式存在的问题是，

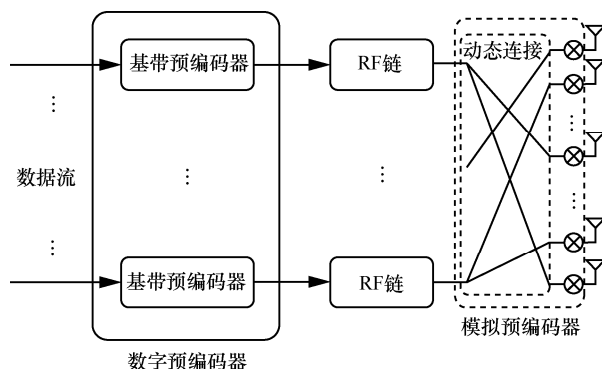


图4 针对多用户的动态子阵列混合预编码结构

需要高精度的相移器才能够实现与全数字预编码在可达和速率性能上的相似效果，对此参考文献[65]提出了一种新的卷积神经网络组合器结构，这种结构可以在硬件限制和瞬时信道状态信息(CSI)不完善的情况学习如何优化组合器以最大化频谱效率。但是在预编码阶段，为了降低复杂度，基带预编码器是直接由等效信道矩阵分解而来的，这在一定程度上降低了精确性。参考文献[66]提出了一种基于机器学习的自适应交叉熵优化算法。在模拟预编码器方面使用自适应连接，用自适应连接结构代替子连接结构中的开关和反向器，这种方式很大程度上降低了硬件复杂度和算法复杂度。此外，将机器学习应用到自适应结构中可以弥补在使用低精度相移器时不能很好地实现阵列增益的问题，提升结构的可达和速率。

3.2 AI 结合 CSI 反馈

大规模 MIMO 系统的潜在优势基于基站上获得的 CSI，因此 CSI 的准确性直接影响了大规模 MIMO 系统的性能。然而，在目前大多数蜂窝系统所采用的频分复用（FDD）系统中，由于上下行链路之间的互易性较弱，下行 CSI 难以被准确获取。因此，在 FDD 系统中，CSI 将首先在用户端通过导频信号进行估计，然后反馈给基站。但是大规模 MIMO 系统中天线数较多，导致大量反馈开销的产生，增加了 CSI 矩阵维数，因此在大规模 MIMO 系统中的 CSI 反馈实现难度非常大。为解决这类问题，应该对 CSI 矩阵进行有效的压缩，此时深度学习等人工智能算法能够发挥巨大作用。

2018 年，参考文献[67]最先提出将 DL 应用到 CSI 反馈中，设计了一种名叫 CsiNet 的 CSI 编解码网络。CsiNet 结构如图 5 所示。在该网络中，编码器不再使用随机投影，而是用训练数据学习从原始信道矩阵到压缩码字的转换。该算法不依赖于信道分布方面的知识，而是直接从训练数据中学习如何有效利用信道结构；而解码器中，CsiNet 学习从码字到原始信号的逆变换，该变换是非迭代的，且算法速度相比迭代算法有数量级上的提升。终端使用编码器将信道矩阵转换为码字，一旦码字反馈给基站，基站就能使用解码器恢复原来的信道矩阵，这大大降低了 CSI 反馈在

FDD MIMO 系统中实现的难度。在此基础上，参考文献[68]设计了一个扩展版本的 CsiNet+ 网络，网络结构如图 6 所示，该网络通过扩大卷积核大小进一步提升 CsiNet 的性能，并提出了串行多路压缩 CsiNet+（SM-CsiNet+）与并行多路压缩 CsiNet+（PM-CsiNet+）。实验结果表明，与原有 CsiNet 相比，CsiNet+ 网络使 CSI 反馈参数数量降低了 38.0% 和 46.7%。

但是上述两种方案侧重关注角延迟域稀疏性，而忽略了具有时变特性的大规模 MIMO 信道的时间相关性。这类方案在压缩率低时，分辨率会有所下降。针对此问题，参考文献[69]将 CsiNet 进行了改进，用 LSTM 网络代替 CNN，设计了一个叫作 CsiNet-LSTM 的 CSI 反馈方案。LSTM 网络具有提取时间特征的能力，通过使用 LSTM 对信道的时间相关性进行提取可以进一步提高分辨率。实验结果表明，CsiNet-LSTM 在 CSI 恢复质量上有较好的性能，且复杂度相较于现有基于压缩传感的 CSI 反馈方法更为简单。与 CsiNet 相比，CsiNet-LSTM 在降低压缩比方面表现出了优秀的鲁棒性，但是需要更多的参数，占用更多内存空间。随后，参考文献[70]在 LSTM 网络之后增加了一项注意力（Attention）机制^[71]，提出一种叫作 LSTM-Attention CsiNet 网络结构。注意力机制将 LSTM 网络的输出作为输入，通过自动地对特

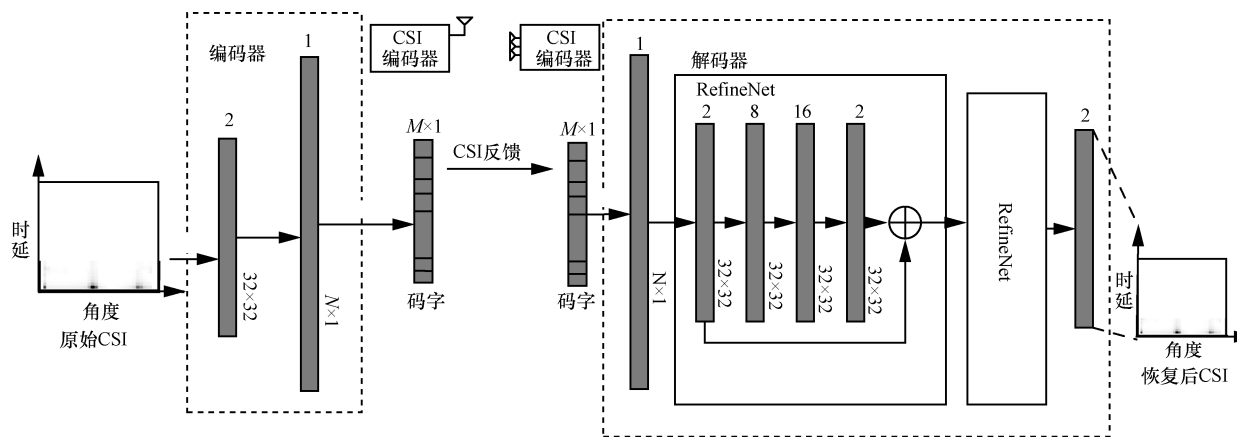


图 5 CsiNet 结构

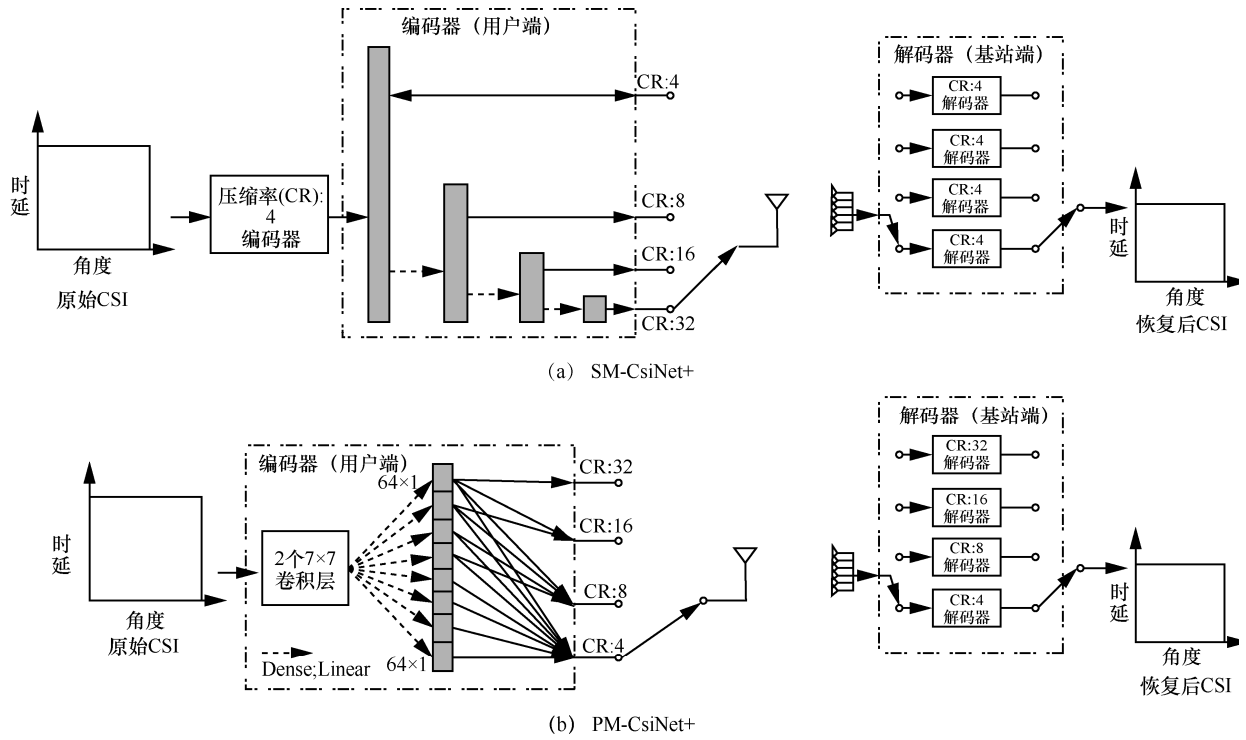


图6 扩展的 CsiNet+网络结构

征信息赋予优先级和权重，从而在参数数量较少的情况下也能够提取出重要的特征信息。通过这种方法，算法能在保证精确度的同时，兼顾内存空间的占用情况，并在提升系统性能方面表现出不俗的能力。

虽然 DL 方法在减少大规模 MIMO 的 CSI 矩阵维度方面实现了显著的性能提升，但目前使用单精度（32 bit）编码反馈系数的 CSI 反馈仍然消耗了太多的带宽。为了减少每个码字所需的比特数，参考文献[72]对维数降低后的编码进行了低比特量化，提出了一种叫作 CQNet 的通用 CSI 压缩框架，用于在有限带宽约束下联合优化 CSI 压缩、码字量化和 CSI 恢复。文章将 CQNet 分别与 CsiNet+[68]和 DualNet-MAG[73]对 CQNet 框架进行效果评估，结果表明，9 bit 均匀量化器在不对量化码字进行神经网络再训练的情况下，可以实现与 32 bit 高精度反馈相似的精度，且在压缩维度较低的情况下两种网络都能够表现出较强的鲁棒性。然而在 CSI 重建阶段，粗量化精度会导致 CSI

重建精度的降低。在实际场景中，相比于在高振幅时步长大、低振幅时步长小的非均匀量化器，均匀量化器的性能会相对较差。对此，参考文献[74]在量化阶段采用了一个 μ 律非均匀量化器来离散压缩 CSI，并引入基于神经网络的偏移网络消除量化失真。实验结果表明，与使用均匀量化器的 CSI 反馈方案相比，非均匀量化器方法无论是在室内还是在室外都能拥有更好的量化信噪比（signal-to-quantization noise ratio, SQNR）。同时，由于神经网络的参与，偏移网络中叠加的全连接层细化了去量化器的输出，从而减小了量化失真，使信道矩阵更容易恢复。

由于实际场景中存在大量噪声，仅仅对 CSI 矩阵进行降维仍然会使得码字中存在冗余。参考文献[75]提出了一种叫作 AnciNet 的网络结构应对带噪声的 CSI 压缩反馈，该网络由预降噪模块与 Anci 块增强反馈模块组成。其中，预降噪模块部署在 CNN 神经网络的前端，在编码前对噪声进行初步抑制；Anci 块增强反馈模块则对压缩与解

压缩过程中的残留噪声进行处理。通过这两个模块, CNN 能够使基站端从噪声中更好地恢复 CSI。与参考文献[67-68]相比, 该方法在归一化 MSE (NMSE) 指标方面有明显提升。但是在参数数量方面, AnciNet 相较于 CsiNet 参数数量更多, 这也意味着算法训练的时间更久, 回传 CSI 的时间越长。

上述各个网络的性能对比见表 2~表 4。NMSE 值越小, 代表基站对信道信息恢复得越准确; 参数数量越少, 代表算法所需训练时间越短。

4 其他与 AI 相关的研究

除前两节所提到的问题, AI 技术在 5G 系统的众多领域都有或多或少的应用。本节简单地对 5G 系统中基于 AI 的算法及方案进行介绍。

4.1 AI 结合信道均衡

信道均衡用来解决由信道和系统缺陷造成的失真问题, 这些问题包括由放大器和数据转换器引起的码间干扰 (ISI) 和非线性。参考文献[76]提出了一种新的方法来开发认知引擎 (CE), 该方法基于遗传算法和强化学习, 用于高频波段的 MIMO 系统中接收端信号均衡。参考文献[77]提出了三维射线发射 (3D-RL) 毫米波通道模型, 其中充分考虑了拓扑和材料特性来生成人工神经网络 (artificial neural network, ANN) 的数据集, 并且训练了一个多层感知 (multi-layer perceptron, MLP) 神经网络, 该神经网络将毫米波信道的最小二乘估计作为输入, 联合学习信道估计与信道均衡过程。参考文献[78]提出了一个基于 DL 的聚合物理层模块 (DeepWiPHY), 联合执

表 2 各网络 NMSE 性能对比 (单位: dB)

	压缩率 (CR)	CsiNet ^[67]	CsiNet+ ^[68]	SM-CsiNet+ ^[68]	PM-CsiNet+ ^[68]	LSTM-CsiNet ^[68]
室内	4	-17.36	-27.37	-27.90	-27.60	—
	8	-12.70	-18.29	-18.49	-17.70	—
	16	-8.64	-14.14	-13.45	-12.25	-23.06
	32	-6.24	-10.43	-9.89	-8.24	-22.33
	64	-5.84	—	—	—	-21.24
室外	4	-8.75	-12.40	-11.91	-12.02	—
	8	-7.61	-8.72	-8.25	-8.10	—
	16	-4.51	-5.73	-5.31	-5.07	-9.86
	32	-2.81	-3.4	-3.22	-3.00	-9.18
	64	-1.93	—	—	—	-8.83

表 3 各网络在不同压缩率下的参数数量对比

网络	4	8	16	32	64
CsiNet ^[67]	2 103 904	1 055 072	530 656	268 448	137 344
CsiNet+ ^[68]	2 122 340	1 073 508	549 092	286 884	—
LSTM-CsiNet ^[69]	10 247 148	—	7 484 688	7 024 272	6 794 064
LSTM-Attention ^[70]	2 924 912	—	582 704	281 936	141 152
AnciNet ^[75]	2 289 334	—	716 086	453 878	322 774



表 4 不同信道一噪声比(CNR)与压缩率(CR)下 CsiNet 与 AnciNet 的 NMSE 性能对比 (单位: dB)

压缩率	CNR (dB)	CsiNet ^[67]	AnciNet ^[75]
16	0	-2.76	-8.70
	5	-6.24	-11.64
	10	-8.71	-13.31
	15	-9.56	-14.42
	20	-10.03	-14.84
	25	-10.19	-14.99
32	0	-1.94	-7.47
	5	-4.94	-9.26
	10	-7.23	-9.88
	15	-7.83	-10.42
	20	-8.09	-10.55
	25	-8.17	-10.60
64	0	-0.75	-6.49
	5	-3.45	-7.65
	10	-4.80	-7.92
	15	-5.28	-8.30
	20	-5.45	-8.37
	25	-5.52	-8.40

行信道估计、共相位误差和采样率偏移(CpeSro)校正和信道均衡。与高度模块化的传统接收机相比, DeepWiPHY 可以同时执行多个补偿, 这能够使系统获得更好的性能, 也可以处理非线性射频损耗。但是, 由于神经网络的特性, 在经过基于神经网络的信道均衡后, 噪声通常不再是独立同分布的高斯白噪声, 这将导致信道解码器性能的下降^[79]。如何解决这一问题将是接下来研究的重点。

4.2 AI 结合同同时同频全双工自干扰消除

同时同频全双工(co-frequency co-time full duplex, CCFD)系统实现了无线电通信双方在同一频带内同时发送和接收信号, 这在理论上可以使现有的无线通信系统的容量与频谱效率翻倍^[80]。但

是, 在这种情况下接收天线将会捕获来自自身发射天线的自干扰信号(SI), 因此解决自干扰问题是实现 FD 技术的关键。参考文献[81]首次提出用机器学习的方法解决自干扰问题, 对传输信号的实部和虚部分别建立了两个独立的支持向量回归(SVR)模型进行学习预测, 实验结果表明这种方式能够很大程度上降低复杂度。参考文献[82]提出了一种具有级联结构的双层前馈神经网络解决 FD-MIMO 系统中的自干扰消除问题, 该算法同时实现了对自干扰消除效果的提升和系统复杂度的降低。由于硬件总是存在偏差的, 因此接收信号与发射信号的采样频率并不能严格一致, 这将导致训练过程中迅速出现过拟合现象, 严重影响自干扰消除效果。为解决这个问题, 参考文献[83]提出了一种采用机器学习的实时自干扰解决方案, 并使用了一种内插值算法来帮助发射和接收数据完成匹配。

5 未来发展方向

5.1 AI 与 5G 资源管理

在 5G 大连接场景下, 特别是物联网场景, 无线资源管理会由于用户量大、业务量大、系统不确定性高等问题变得异常困难。如何动态地、合理地分配信道资源, 实现功率控制与切换管理等将是保障 5G 系统平稳高效运行的重点。另外, 诸如边缘计算资源分配、网络切片资源管理等也要求系统具有一个良好且简单的资源管理算法。深度学习、强化学习等 AI 算法能够很好地对大量数据进行处理, 但是对于 5G 系统来说, 现有的 AI 算法依然存在较高的计算复杂度和计算时延。因此, 如何对大量数据进行预处理, 将训练数据集进行简化以提高算法效率成为未来一个重要的研究方向。

5.2 AI 与端到端通信

端到端通信被视为 5G 系统中的一种优化方案, 通过终端设备的短程直接连接通信, 实现了

系统频谱利用率和吞吐率的提升,降低了系统时延。为了充分发挥 D2D 的优势,需要解决包括设备发现、模式选择、干扰管理、功率控制、安全性保障、无线资源分配等一系列问题^[84]。针对以上问题,目前学术界提出了一系列基于 AI 算法的 D2D 通信优化方式^[84-86],且理论上能够实现系统性能的提升。然而,在实际应用中,由于端用户本身计算能力的限制,使用复杂的 AI 算法会导致较高的计算开销(例如时延、能耗等)。因此如何保证 AI 算法能够在端设备上以低成本运行是端到端通信中一个亟待解决的问题。

5.3 AI 与 5G 安全性

安全性问题一直是无线通信系统的一个研究重点。特别是,在 5G 中,连接从“人与人”上升到了“人与物”“物与物”,这在带来万物互联优势的同时,多样化的应用场景也使得安全性问题变得异常复杂。另外,由于边缘计算、网络切片等技术的提出,开放、多元的网络层结构也增加了 5G 网络安全信任体系的复杂度。对于这种高度复杂且多变的应用场景,AI 技术被认为是一种有效解决问题的途径,也将成为未来研究的一个重点。

5.4 AI 与 6G

随着无线通信技术的发展,学术界针对 6G 的研究正在慢慢展开。与前几代技术相比,6G 将彻底改变无线通信的发展,将“连接万物”发展到“连接智能”^[87]。6G 系统将采用先进的人工智能技术,及时高效、随时随地采集、传输和学习数据,生产大量创新性应用和智能服务。并且,无处不在的人工智能将赋予 6G 一种高度灵活的架构,将“以人为中心”的开发概念落实到网络的各个方面,取代“以数据为中心”“以机器为中心”和“以应用为中心”的服务模式^[88]。

6 结束语

本文总结了人工智能技术在 5G 中的应用。具

体总结了与 AI 相关算法在移动边缘计算技术与毫米波大规模 MIMO 技术中的相关研究;分析了各研究的创新点和优缺点;最后基于当前的发展趋势,提出了 AI 与 5G 结合的未来发展方向。

目前,AI 技术已经在各个领域有着广泛的应用,且被证明具有巨大优势。越来越多针对 5G 技术的研究也开始倾向于使用 AI 技术。然而在实际场景中,设备功耗、计算能力、应用场景等多种因素使得 AI 算法在 5G 系统中真正部署有着一定困难。但是由于传统算法在面对多变复杂且数据量大的问题上有着明显的缺陷,AI 技术仍然具有巨大的发展空间和潜力。

参考文献:

- [1] GHOSH A, MAEDER A, BAKER M, et al. 5G evolution: a view on 5G cellular technology beyond 3GPP release 15[J]. IEEE Access, 2019(7): 127639-127651.
- [2] ZHANG C, HUANG Y H, SHEIKH F, et al. Advanced base-band processing algorithms, circuits, and implementations for 5G communication[J]. IEEE Journal on Emerging and Selected Topics in Circuits and Systems, 2017, 7(4): 477-490.
- [3] MA Z, ZHANG Z Q, DING Z G, et al. Key techniques for 5G wireless communications: network architecture, physical layer, and MAC layer perspectives[J]. Science China Information Sciences, 2015, 58(4): 1-20.
- [4] LIN C Z, LU J W, WANG G, et al. Graininess-aware deep feature learning for robust pedestrian detection[J]. IEEE Transactions on Image Processing, 2020(29): 3820-3834.
- [5] ZHU S, CAO R S, YU K. Dual learning for semi-supervised natural language understanding[J]. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2020(28): 1936-1947.
- [6] SYNNAEVE G, BESSIÈRE P. Multiscale Bayesian modeling for RTS games: an application to StarCraft AI[J]. IEEE Transactions on Computational Intelligence and AI in Games, 2016, 8(4): 338-350.
- [7] 刘留, 张建华, 樊圆圆, 等. 机器学习在信道建模中的应用综述[J]. 通信学报, 2021, 42(2): 134-153.
LIU L, ZHANG J H, FAN Y Y, et al. Survey of application of machine learning in wireless channel modeling[J]. Journal on Communications, 2021, 42(2): 134-153.
- [8] 尤肖虎, 张川, 谈晓思, 等. 基于 AI 的 5G 技术: 研究方向与范例[J]. 中国科学: 信息科学, 2018, 48(12): 1589-1602.
YOU X H, ZHANG C, TAN X S, et al. AI for 5G: research di-



- rections and paradigms[J]. *Scientia Sinica (Informationis)*, 2018, 48(12): 1589-1602.
- [9] 张建敏, 杨峰义, 武洲云. 多接入边缘计算 (MEC) 及关键技术[M]. 北京: 人民邮电出版社, 2019.
- ZHANG J M, YANG F Y, WU Z Y. Multi-access edge computing (MEC) and key technologies[M]. Beijing: Posts & Telecom Press, 2019.
- [10] NING Z L, HU X P, CHEN Z K, et al. A cooperative quality-aware service access system for social Internet of vehicles[J]. *IEEE Internet of Things Journal*, 2020, 7(7): 6663.
- [11] 谢人超, 廉晓飞, 贾庆民, 等. 移动边缘计算卸载技术综述[J]. *通信学报*, 2018, 39(11): 138-155.
- XIE R C, LIAN X F, JIA Q M, et al. Survey on computation offloading in mobile edge computing[J]. *Journal on Communications*, 2018, 39(11): 138-155.
- [12] 陈山枝, 王胡成, 时岩. 5G 移动性管理技术[M]. 北京: 人民邮电出版社, 2019.
- CHEN S Z, WANG H C, SHI Y. Mobile management technology for 5G[M]. Beijing: Posts & Telecom Press, 2019.
- [13] 杨志强, 栗栗, 杨波, 等. 5G 安全技术与标准[M]. 北京: 人民邮电出版社, 2020.
- YANG Z Q, SU L, YANG B, et al. 5G security technologies and standards[M]. Beijing: Posts & Telecom Press, 2020.
- [14] CAO Z L, ZHOU P, LI R X, et al. Multiagent deep reinforcement learning for joint multichannel access and task offloading of mobile-edge computing in industry 4.0[J]. *IEEE Internet of Things Journal*, 2020, 7(7): 6201-6213.
- [15] WU H M, ZHANG Z R, GUAN C, et al. Collaborate edge and cloud computing with distributed deep learning for smart city Internet of Things[J]. *IEEE Internet of Things Journal*, 2020, 7(9): 8099-8110.
- [16] YANG B, CAO X L, LI X F, et al. Mobile-edge- computing-based hierarchical machine learning tasks distribution for IIoT[J]. *IEEE Internet of Things Journal*, 2020, 7(3): 2169-2180.
- [17] YANG B, CAO X L, LI X F, et al. Joint communication and computing optimization for hierarchical machine learning tasks distribution[C]//Proceedings of 2019 IEEE Symposium on Computers and Communications (ISCC). Piscataway: IEEE Press, 2019: 1-6.
- [18] JIANG F B, WANG K Z, DONG L, et al. Stacked autoencoder-based deep reinforcement learning for online resource scheduling in large-scale MEC networks[J]. *IEEE Internet of Things Journal*, 2020, 7(10): 9278-9290.
- [19] NATH S, LI Y Z, WU J X, et al. Multi-user multi-channel computation offloading and resource allocation for mobile edge computing[C]//Proceedings of ICC 2020 - 2020 IEEE International Conference on Communications (ICC). Piscataway: IEEE Press, 2020: 1-6.
- [20] HUANG L, BI S Z, ZHANG Y J A. Deep reinforcement learning for online computation offloading in wireless powered mobile-edge computing networks[J]. *IEEE Transactions on Mobile Computing*, 2020, 19(11): 2581-2593.
- [21] SONG X, CHAI R, CHEN Q B. Joint task offloading, CNN layer scheduling and resource allocation in cooperative computing system[C]//Proceedings of Communications and Networking. [S.l.:s.n.], 2020.
- [22] LI J, GAO H, LV T, et al. Deep reinforcement learning based computation offloading and resource allocation for MEC[C]//Proceedings of 2018 IEEE Wireless Communications and Networking Conference (WCNC). Piscataway: IEEE Press, 2018: 1-6.
- [23] 赵海涛, 张唐伟, 陈跃, 等. 基于 DQN 的车载边缘网络任务分发卸载算法[J]. *通信学报*, 2020, 41(10): 172-178.
- ZHAO H T, ZHANG T W, CHEN Y, et al. Task distribution offloading algorithm of vehicle edge network based on DQN[J]. *Journal on Communications*, 2020, 41(10): 172-178.
- [24] MACH P, BECVAR Z. Cloud-aware power control for cloud-enabled small cells[C]// Proceedings of 2014 IEEE Globecom Workshops (GC Wkshps). Piscataway: IEEE Press, 2014: 1038-1043.
- [25] WANG S Q, ZAFER M, LEUNG K K. Online placement of multi-component applications in edge computing environments[J]. *IEEE Access*, 2017(5): 2514-2533.
- [26] ALI Z, KHAF S, ABBAS Z H, et al. A deep learning approach for mobility-aware and energy-efficient resource allocation in MEC[J]. *IEEE Access*, 2020(8): 179530-179546.
- [27] DIN N, CHEN H P, KHAN D. Mobility-aware resource allocation in multi-access edge computing using deep reinforcement learning[C]// Proceedings of 2019 IEEE Intl Conf on Parallel & Distributed Processing with Applications, Big Data & Cloud Computing, Sustainable Computing & Communications, Social Computing & Networking. Piscataway: IEEE Press, 2019: 202-209.
- [28] ZHAN W H, LUO C B, WANG J, et al. Deep-reinforcement-learning-based offloading scheduling for vehicular edge computing[J]. *IEEE Internet of Things Journal*, 2020, 7(6): 5449-5465.
- [29] SCHULMAN J, WOLSKI F, DHARIWAL P, et al. Proximal policy optimization algorithms[EB]. 2017.
- [30] WU C L, CHIU T C, WANG C Y, et al. Mobility-aware deep reinforcement learning with glimpse mobility prediction in edge computing[C]// Proceedings of ICC 2020 - 2020 IEEE International Conference on Communications (ICC). Piscataway: IEEE Press, 2020: 1-7.
- [31] WANG J F, LV T, HUANG P M, et al. Mobility-aware partial computation offloading in vehicular networks: a deep reinforcement learning based scheme[J]. *China Communications*, 2020, 17(10): 31-49.
- [32] ABBAS N, ZHANG Y, TAHERKORDI A, et al. Mobile edge computing: a survey[J]. *IEEE Internet of Things Journal*, 2018, 5(1): 450-465.

- [33] 杨建喜, 张媛利, 蒋华, 等. 边缘计算中基于深度 Q 网络的物理层假冒攻击检测方法[J]. 计算机应用, 2020, 40(11): 3229-3235.
- YANG J X, ZHANG Y L, JIANG H, et al. Detection method of physical-layer impersonation attack based on deep Q-network in edge computing[J]. *Journal of Computer Applications*, 2020, 40(11): 3229-3235.
- [34] 陈佳, 欧阳金源, 冯安琪, 等. 边缘计算构架下基于孤立森林算法的 DoS 异常检测[J]. 计算机科学, 2020, 47(2): 287-293.
- CHEN J, OUYANG J Y, FENG A Q, et al. DoS anomaly detection based on isolation forest algorithm under edge computing framework[J]. *Computer Science*, 2020, 47(2): 287-293.
- [35] GOPALAKRISHNAN T, RUBY D, AL-TURJMAN F, et al. Deep learning enabled data offloading with cyber attack detection model in mobile edge computing systems[J]. *IEEE Access*, 2020(8): 185938-185949.
- [36] KAYODE O, TOSUN A S. Deep Q-network for enhanced data privacy and security of IoT traffic[C]//*Proceedings of 2020 IEEE 6th World Forum on Internet of Things (WF-IoT)*. Piscataway: IEEE Press, 2020: 1-6.
- [37] XIONG J B, ZHAO M F, BHUIYAN M Z A, et al. An AI-enabled three-party game framework for guaranteed data privacy in mobile edge crowdsensing of IoT[J]. *IEEE Transactions on Industrial Informatics*, 2021, 17(2): 922-933.
- [38] HE X F, JIN R C, DAI H Y. Deep PDS-learning for privacy-aware offloading in MEC-enabled IoT[J]. *IEEE Internet of Things Journal*, 2019, 6(3): 4547-4555.
- [39] DONG J Y, GENG D Q, HE X F. Privacy-aware task offloading via two-timescale reinforcement learning[C]//*Proceedings of 2020 IEEE/CIC International Conference on Communications in China (ICCC)*. Piscataway: IEEE Press, 2020: 220-225.
- [40] CHEN S G, YOU Z H, RUAN X K. Privacy and energy co-aware data aggregation computation offloading for fog-assisted IoT networks[J]. *IEEE Access*, 2020(8): 72424-72434.
- [41] HUANG B B, LI Z J, TANG P, et al. Security modeling and efficient computation offloading for service workflow in mobile edge computing[J]. *Future Generation Computer Systems*, 2019, 97: 755-774.
- [42] CHEN T, MATINMIKKO M, CHEN X F, et al. Software defined mobile networks: concept, survey, and research directions[J]. *IEEE Communications Magazine*, 2015, 53(11): 126-133.
- [43] BUSARI S A, HUQ K M S, MUMTAZ S, et al. Millimeter-wave massive MIMO communication for future wireless systems: a survey[J]. *IEEE Communications Surveys & Tutorials*, 2018, 20(2): 836-869.
- [44] LI Q C, NIU H N, PAPATHANASSIOU A T, et al. 5G network capacity: key elements and technologies[J]. *IEEE Vehicular Technology Magazine*, 2014, 9(1): 71-78.
- [45] UWAECHIA A N, MAHYUDDIN N M. A comprehensive survey on millimeter wave communications for fifth-generation wireless networks: feasibility and challenges[J]. *IEEE Access*, 2020(8): 62367-62414.
- [46] AYACH O E, RAJAGOPAL S, ABU-SURRA S, et al. Spatially sparse precoding in millimeter wave MIMO systems[J]. *IEEE Transactions on Wireless Communications*, 2014, 13(3): 1499-1513.
- [47] YU X H, SHEN J C, ZHANG J, et al. Alternating minimization algorithms for hybrid precoding in millimeter wave MIMO systems[J]. *IEEE Journal of Selected Topics in Signal Processing*, 2016, 10(3): 485-500.
- [48] CHEN C E. An iterative hybrid transceiver design algorithm for millimeter wave MIMO systems[J]. *IEEE Wireless Communications Letters*, 2015, 4(3): 285-288.
- [49] JIN J N, ZHENG Y R, CHEN W, et al. Hybrid precoding for millimeter wave MIMO systems: a matrix factorization approach[J]. *IEEE Transactions on Wireless Communications*, 2018, 17(5): 3327-3339.
- [50] ALKHATEEB A, LEUS G, HEATH R W. Limited feedback hybrid precoding for multi-user millimeter wave systems[J]. *IEEE Transactions on Wireless Communications*, 2015, 14(11): 6481-6494.
- [51] NI W H, DONG X D. Hybrid block diagonalization for massive multiuser MIMO systems[J]. *IEEE Transactions on Communications*, 2016, 64(1): 201-211.
- [52] LIANG L, XU W, DONG X D. Low-complexity hybrid precoding in massive multiuser MIMO systems[J]. *IEEE Wireless Communications Letters*, 2014, 3(6): 653-656.
- [53] DAI L L, GAO X Y, QUAN J G, et al. Near-optimal hybrid analog and digital precoding for downlink mmWave massive MIMO systems[C]//*Proceedings of 2015 IEEE International Conference on Communications (ICC)*. Piscataway: IEEE Press, 2015: 1334-1339.
- [54] GAO X Y, DAI L L, HAN S F, et al. Energy-efficient hybrid analog and digital precoding for mmWave MIMO systems with large antenna arrays[J]. *IEEE Journal on Selected Areas in Communications*, 2016, 34(4): 998-1009.
- [55] YU X H, ZHANG J, LETAIEF K B. A hardware-efficient analog network structure for hybrid precoding in millimeter wave systems[J]. *IEEE Journal of Selected Topics in Signal Processing*, 2018, 12(2): 282-297.
- [56] PARK S, ALKHATEEB A, HEATH R W. Dynamic subarrays for hybrid precoding in wideband mmWave MIMO systems[J]. *IEEE Transactions on Wireless Communications*, 2017, 16(5): 2907-2920.
- [57] JING X R, LI L H, LIU H Q, et al. Dynamically connected hybrid precoding scheme for millimeter-wave massive MIMO systems[J]. *IEEE Communications Letters*, 2018, 22(12): 2583-2586.



- [58] PAYAMI S, GHORAISHI M, DIANATI M. Hybrid beamforming for large antenna arrays with phase shifter selection[J]. *IEEE Transactions on Wireless Communications*, 2016, 15(11): 7258-7271.
- [59] SIDHARTH C, HIREMATH S M, PATRA S K. Deep Learning based Hybrid Precoding for mmWave Massive MIMO system using ComcepNet[C]//*Proceedings of 2020 International Conference on Communication and Signal Processing (ICCSPP)*. Piscataway: IEEE Press, 2020: 1317-1321.
- [60] HUANG H J, SONG Y W, YANG J, et al. Deep-learning-based millimeter-wave massive MIMO for hybrid precoding[J]. *IEEE Transactions on Vehicular Technology*, 2019, 68(3): 3027-3032.
- [61] CHAI M Y, TANG S H, ZHAO M, et al. HPNet: a compressed neural network for robust hybrid precoding in multi-user massive MIMO systems[C]//*Proceedings of GLOBECOM 2020 - 2020 IEEE Global Communications Conference*. Piscataway: IEEE Press, 2020: 1-7.
- [62] MA W Y, QI C H, ZHANG Z C, et al. Sparse channel estimation and hybrid precoding using deep learning for millimeter wave massive MIMO[J]. *IEEE Transactions on Communications*, 2020, 68(5): 2838-2849.
- [63] MIR T, ZAIN SIDDIQI M, MIR U, et al. Machine learning inspired hybrid precoding for wideband millimeter-wave massive MIMO systems[J]. *IEEE Access*, 2019(7): 62852-62864.
- [64] JIANG J, YANG Y. Deep learning assisted hybrid precoding with dynamic subarrays in mmWave MU-MIMO system[C]//*Proceedings of 2020 IEEE/CIC International Conference on Communications in China (ICCC)*. Piscataway: IEEE Press, 2020: 256-261.
- [65] BAO X L, FENG W J, ZHENG J L, et al. Deep CNN and equivalent channel based hybrid precoding for mmWave massive MIMO systems[J]. *IEEE Access*, 2020(8): 19327-19335.
- [66] 甘天江, 傅友华, 王海荣. 毫米波大规模 MIMO 系统中基于机器学习的自适应连接混合预编码[J]. *信号处理*, 2020, 36(5): 677-685.
- GAN T J, FU Y H, WANG H R. Machine learning-based adaptive connection hybrid precoding for mmWave massive MIMO systems[J]. *Journal of Signal Processing*, 2020, 36(5): 677-685.
- [67] LI J L, ZHANG Q, XIN X J, et al. Deep learning-based massive MIMO CSI feedback[C]//*Proceedings of 2019 18th International Conference on Optical Communications and Networks (ICOON)*. Piscataway: IEEE Press, 2019: 1-3.
- [68] GUO J J, WEN C K, JIN S, et al. Convolutional neural network-based multiple-rate compressive sensing for massive MIMO CSI feedback: design, simulation, and analysis[J]. *IEEE Transactions on Wireless Communications*, 2020, 19(4): 2827-2840.
- [69] WANG T Q, WEN C K, JIN S, et al. Deep learning-based CSI feedback approach for time-varying massive MIMO channels[J]. *IEEE Wireless Communications Letters*, 2019, 8(2): 416-419.
- [70] LI Q, ZHANG A H, LIU P C, et al. A novel CSI feedback approach for massive MIMO using LSTM-attention CNN[J]. *IEEE Access*, 2020(8): 7295-7302.
- [71] KARIM F, MAJUMDAR S, DARABI H, et al. LSTM fully convolutional networks for time series classification[J]. *IEEE Access*, 2018(6): 1662-1669.
- [72] LIU Z Y, ZHANG L, DING Z. An efficient deep learning framework for low rate massive MIMO CSI reporting[J]. *IEEE Transactions on Communications*, 2020, 68(8): 4761-4772.
- [73] LIU Z Y, ZHANG L, DING Z. Exploiting Bi-directional channel reciprocity in deep learning for low rate massive MIMO CSI feedback[J]. *IEEE Wireless Communications Letters*, 2019, 8(3): 889-892.
- [74] CHEN T, GUO J J, JIN S, et al. A novel quantization method for deep learning-based massive MIMO CSI feedback[C]//*Proceedings of 2019 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*. Piscataway: IEEE Press, 2019: 1-5.
- [75] SUN Y Y, XU W, FAN L S, et al. AnciNet: an efficient deep learning approach for feedback compression of estimated CSI in massive MIMO systems[J]. *IEEE Wireless Communications Letters*, 2020, 9(12): 2192-2196.
- [76] SAMUEL S, KRISTOPHER H. JUN G, et al. Marefat, Tamal Bose: machine learning based MIMO equalizer for high frequency (HF) communications[Z]. 2020.
- [77] CARRERA D F, VARGAS-ROSALES C, YUNGAICE-LA-NAULA N M, et al. Comparative study of artificial neural network based channel equalization methods for mmWave communications[J]. *IEEE Access*, 2021(9): 41678-41687.
- [78] ZHANG Y, DOSHI A, LISTON R, et al. DeepWiPHY: deep learning-based receiver design and dataset for IEEE 802.11ax systems[J]. *IEEE Transactions on Wireless Communications*, 2021, 20(3): 1596-1611.
- [79] YE H, LI G Y. Initial results on deep learning for joint channel equalization and decoding[C]//*Proceedings of 2017 IEEE 86th Vehicular Technology Conference (VTC-Fall)*. Piscataway: IEEE Press, 2017: 1-5.
- [80] NWANKWO C D, ZHANG L, QUDDUS A, et al. A survey of self-interference management techniques for single frequency full duplex systems[J]. *IEEE Access*, 2018(6): 30242-30268.
- [81] GUO H Q, WU S E, WANG H G, et al. DSIC: deep learning based self-interference cancellation for in-band full duplex wireless[C]//*Proceedings of 2019 IEEE Global Communications Conference (GLOBECOM)*. Piscataway: IEEE Press, 2019: 1-6.

- [82] CHEN Y T, MISHRA R K, SCHWARTZ D, et al. MIMO full duplex radios with deep learning[C]//Proceedings of 2020 IEEE International Conference on Communications Workshops (ICC Workshops). Piscataway: IEEE Press, 2020: 1-6.
- [83] GUO H Q, XU J H, ZHU S Y, et al. Realtime software defined self-interference cancellation based on machine learning for in-band full duplex wireless communications[C]//Proceedings of 2018 International Conference on Computing, Networking and Communications (ICNC). Piscataway: IEEE Press, 2018: 779-783.
- [84] IOANNOU I, VASSILIOU V, CHRISTOPHOROU C, et al. Distributed artificial intelligence solution for D2D communication in 5G networks[J]. IEEE Systems Journal, 2020, 14(3): 4232-4241.
- [85] XU J, GU X Y, FAN Z Q. D2D power control based on hierarchical extreme learning machine[C]//Proceedings of 2018 IEEE 29th Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC). Piscataway: IEEE Press, 2018: 1-7.
- [86] HAMDI M, YUAN D, ZAIED M. GA-based scheme for fair joint channel allocation and power control for underlaying D2D multicast communications[C]//Proceedings of 2017 13th International Wireless Communications and Mobile Computing Conference (IWCMC). Piscataway: IEEE Press, 2017: 446-451.
- [87] LIU Y, YUAN X L, XIONG Z H, et al. Federated learning for 6G communications: Challenges, methods, and future directions[J]. China Communications, 2020, 17(9): 105-118.
- [88] DANG S P, AMIN O, SHIHADA B, et al. What should 6G be?[J]. Nature Electronics, 2020, 3(1): 20-29.

[作者简介]



章坚武 (1961—)，男，博士，杭州电子科技大学通信工程学院教授、博士生导师，中国电子学会、中国通信学会高级会员，浙江省通信学会常务理事，主要研究方向为移动通信、多媒体信号处理与人工智能、通信网络与信息安全。



王路鑫 (1996—)，男，杭州电子科技大学通信工程学院硕士生，主要研究方向为 5G 接入边缘计算等。

孙玲芬 (1963—)，女，博士，英国普利茅斯大学副教授，英国高等教育学会会士，主要研究方向为多媒体和网络质量评估、QoS/QoE 管理与控制、VoIP 等。

章谦骅 (1990—)，男，浙江大学博士生，之江实验室智能网络研究中心工程师，主要研究方向为无线网络、物联网、智慧城市等。

单杭冠 (1982—)，男，浙江大学副教授、博士生导师，主要研究方向为与 5G 移动通信相关的网络设计和 QoS 保障等。