

智能无线通信技术研究概况

梁应敬¹, 谭俊杰¹, Dusit Niyato²

(1. 电子科技大学通信抗干扰技术国家级重点实验室, 四川 成都 611731;
2. 南洋理工大学计算机科学与工程学院, 新加坡 639798)

摘 要: 近年来, 人工智能技术已被应用于无线通信领域, 以解决传统无线通信技术面对信息爆炸和万物互联等新发展趋势所遇到的瓶颈问题。首先介绍深度学习、深度强化学习和联邦学习三类具有代表性的人工智能技术; 然后通过对这三类技术在无线通信中的无线传输、频谱管理、资源配置、网络接入、网络及系统优化 5 个方面的应用进行综述, 分析和总结它们在解决无线通信问题时的原理、适用性、设计方法和优缺点; 最后围绕存在的局限性指出智能无线通信技术的未来发展趋势和研究方向, 期望为无线通信领域的后续研究提供帮助和参考。

关键词: 人工智能; 无线通信; 深度学习; 深度强化学习; 联邦学习

中图分类号: TN929.5

文献标识码: A

doi: 10.11959/j.issn.1000-436x.2020145

Overview on intelligent wireless communication technology

LIANG Yingchang¹, TAN Junjie¹, Dusit Niyato²

1. National Key Laboratory of Science and Technology on Communications, University of Electronic Science and Technology of China, Chengdu 611731, China

2. School of Computer Science and Engineering, Nanyang Technological University, Singapore 639798, Singapore

Abstract: In recent years, artificial intelligence (AI) has been applied to wireless communications, in order to address the challenges introduced by data explosion and Internet of everything. Firstly, three core technologies of AI were introduced, including deep learning, deep reinforcement learning, and federated learning. Then, an overview of their applications on wireless communications was provided, from the aspects of wireless transmission, spectrum management, resource allocation, network access, network and system optimization. Based on the overview, the principle, applicability, design methodology, pros and cons on applying AI technologies to solve wireless communication problems were analyzed and summarized. Regarding the existed limitations, the future development trends and research directions on intelligent wireless communication technologies were pointed out, to hopefully provide useful help and reference for the future research in this field.

Key words: artificial intelligence, wireless communication, deep learning, deep reinforcement learning, federated learning

1 引言

人类社会步入信息爆炸时代, 各类智能通信终端和流量密集型应用的普及使人们对信息通信的

需求以前所未有的速度增长。思科公司的流量预测白皮书^[1]显示, 到 2022 年, 全球将会有超过 250 亿移动通信设备产生将近 390 EB 的月流量。此外, 虚拟现实 (VR, virtual reality) 等新型移动应用的出

收稿日期: 2020-03-19; 修回日期: 2020-06-22

基金项目: 国家自然科学基金资助项目 (No.61631005, No.U1801261); 国家重点研发计划基金资助项目 (No.2018YFB1801105); 高等学校学科创新引智计划基金资助项目 (No.B20064)

Foundation Items: The National Natural Science Foundation of China (No.61631005, No.U1801261), The National Key Research and Development Program of China (No.2018YFB1801105), The Programme of Introducing Talents of Discipline to Universities (No.B20064)

现更是对信息传输的时延和可靠性提出了更严苛的要求。

在无线通信需求剧增的同时,无线通信系统的发展却遭遇瓶颈问题。一直以来,无线通信系统的设计遵循模型驱动的理念,基于模块化和层次化的思想进行构建,其中使用的每一项技术都是在人工建立数学模型后通过各类优化方法推导得到的。也就是说,传统的无线通信技术通过充分发挥人类对无线通信的知识积累(即专家知识)来保证系统设计与运行的最优性。显然,这种最优性仅局限于被专家知识充分了解的环境中,而这一条件在未来的无线通信系统中将变得越来越难以成立。例如,无线通信系统正在向多频段、超宽带宽且超高频率的方向发展,其所处的无线环境变得极其复杂,难以再依赖专家知识建立广泛适用的信道模型,从而令基于传统信道模型的信道估计方法失效。再者,在未来万物互联时代网络规模剧增的同时,多维度的网络资源却需要通过细粒度的精确配置来满足各异的通信需求。这将导致网络中需要优化的参数数量呈指数级增长,使传统技术所依赖的各类优化算法因巨大的通信及计算开销而不再适用。

在这一背景下,近年来快速发展的人工智能(AI, artificial intelligence)技术为解决无线通信技术发展中遇到的困境带来了新的思路。机器学习在 AI 发展中发挥了重要作用,它能够直接从海量数据中学习所需的隐藏规律,并利用这些规律做出相应的预测或决策。AI 在计算机领域已经得到广泛的应用,主要包括深度学习(DL, deep learning)^[2]、深度强化学习(DRL, deep reinforcement learning)^[3]以及联邦学习(FL, federated learning)^[4]三类机器学习方法。其中,DL 利用深度神经网络(DNN, deep neural network)建立源数据与目标数据间的映射关系,从而能够根据未知源数据预测对应的目标数据,目前它主要被应用在计算机视觉^[5]和自然语言处理^[6]等领域。DRL 则是一类在动态环境中做出最优决策的机器学习方法,它首先通过与环境交互来记录经验,然后利用 DNN 分析环境存在的规律,最后据此做出最优决策,目前主要被应用在机器人控制^[7]等领域。无论是 DL 还是 DRL,它们都需要收集和海量数据,因而产生了数据隐私和安全性问题。为解决这一问题,FL 被提出用于保护数

据隐私的协同学习,目前主要被应用在数据敏感的学习任务中,如输入法联想词预测^[8]等。

AI 的数据驱动特性正好可以解决传统无线通信系统设计中因依赖专家知识和优化算法而产生的问题。此外,AI 的自动控制能力也契合了运营商一直以来致力于减少人工对网络管理和维护的干预,从而降低运作成本的目标。因此,AI 无疑会为无线通信的发展带来新的机遇。目前,AI 中的 DL、DRL 和 FL 等机器学习方法已经被成功用于解决无线通信中的某些问题,为无线通信技术的发展带来了深远影响。值得注意的是,这些机器学习方法最初都是针对计算机领域的特定任务而诞生的,其设计并没有考虑无线通信中的特点。为了最大化 AI 带来的性能增益,在利用 AI 发展无线通信技术时,需要根据无线通信的特点和所需完成的任务来选择合适的机器学习方法并进行针对性的设计。

本文通过介绍 AI 中的 DL、DRL 和 FL 三类主要的机器学习方法及其在无线通信领域应用的研究进展,分析 AI 在解决不同无线通信问题时的原理、适用性、设计方法和优缺点,并针对这些方法的局限性指出未来智能无线通信技术的发展趋势和研究方向。

2 背景知识

2.1 深度学习概述

DL 是近年来人工智能领域应用最广泛的一类机器学习方法,其基本思想是利用 DNN 来拟合源数据与目标数据的关系^[2,9]。

图 1 展示了一个简单的 DNN 结构。DNN 的基本单位是神经元,神经元间的连线表示信息的传递,而连线的箭头方向为信息传递的方向。此外,神经元的排列具有层次结构。根据信息传递的方向,该 DNN 中的神经元分别构成了一个输入层、多个隐含层(又称隐层)和一个输出层。每个神经元的输入信息由与之连接的上一层神经元的输出经缩放与求和得到,该缩放值称为权值。而后,信息再通过神经元的激活函数处理后传递至与之连接的下一层神经元。为了更好地理解,图 2 展示了属于图 1 所示 DNN 中隐层 1 某神经元的内部结构。其中,激活函数为 DNN 带来非线性表征能力,常用的激活函数有 Sigmoid^[10]、tanh、ReLU^[11]等。

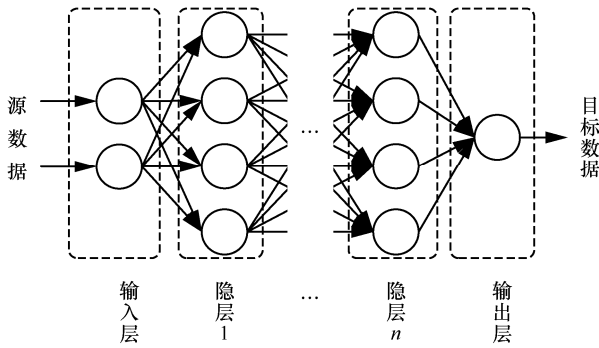


图 1 DNN 结构

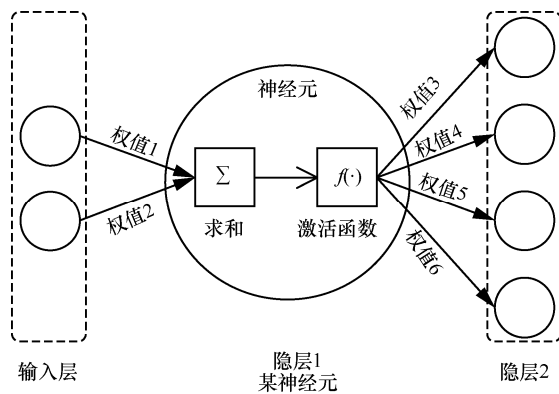


图 2 DNN 中隐层 1 某神经元的内部结构

若给定 DNN 中的神经元连接方式和权值，源数据通过输入层进入 DNN 后，经过神经元间的信息传递规则层层传递后在输出端可以得到相应的目标数据预测值。这一过程称为正向传播。若此时训练数据集中有此源数据对应的目标数据真实值，那么 DNN 可以比对真实值与预测值间的误差，并将误差值从输出层逐层向前传递，以此调整各个神经元的权值，这一过程称为误差反向传递。在 DL 的训练阶段，随机梯度下降算法利用训练数据集反复进行以上 2 个步骤，直至各神经元的权值收敛，则 DNN 的训练结束。完成训练后的 DNN 可以有效地表征源数据和目标数据间的关系，从而能根据未知的源数据预测其目标数据值。

值得注意的是，训练 DNN 仅能调整 DNN 中神经元的权值，而不能改变神经元的排列结构和连接方式，即 DNN 的结构。由图 1 可知，源数据和目标数据的维度决定了 DNN 的输入层和输出层，因此隐层的结构决定了 DNN 的结构。在图 1 DNN 的隐层中，所有的神经元都相互连接，这样的结构称为全连接层（fully-connected layer）。虽然全连接层是最简单的神经元排列结构，但是它理论上可以捕

捉数据间的所有特征。然而，采用全连接层会使权值的数量随着神经元个数的增加而呈指数上升，容易造成训练时间过长、泛化能力差等问题。因此，设计最优的 DNN 需要根据数据的特征或任务特点采用合适的隐层。根据所采用的不同隐层结构，常见的 DNN 有卷积神经网络（CNN, convolutional neural network）^[5]、循环神经网络（RNN, recurrent neural network）^[12]、长短记忆（LSTM, long short term memory）^[13]神经网络。其中，CNN 适用于处理具有局部相似特性的数据，而 RNN 和 LSTM 则擅长处理序列信息。此外，还有一类称为自动编码器（autoencoder）^[14]的特殊 DNN 结构，可以通过训练得到最优的编码器和解码器。

2.2 深度强化学习概述

作为机器学习的另一重要分支，DRL 的目标是在动态环境中做出最优决策^[15]。决策者被称为智能体。DRL 的技术内涵是在强化学习（RL, reinforcement learning）^[16]的基础上发展而来的。传统的 RL 技术可以分为基于值和基于策略这 2 种方法，其具体介绍请参见附录。如附录的分析可知，无论是哪种 RL 方法，都存在较大的局限性，因而难以被广泛应用。DRL 通过结合 DL 来克服传统 RL 技术中存在的局限性。

其中，基于值的 RL 方法需要建立表格来存储值函数，由此产生了维度爆炸问题。为此，DNN 被提出用于拟合值函数，也被称为深度 Q 网络（DQN, deep Q-network）。DQN 的输出记为 $Q(s, a; \theta)$ ，对应 RL 中的值函数 $Q(s, a)$ 。其中， θ 是 DQN 的权值。为了提高决策的精准度，智能体需要利用它与环境互动时记录下的经验对 θ 进行迭代训练和更新。

进一步地，文献[3]提出经验回放和拟静态目标网络（quasi-static target network）技术来提高 DQN 的训练效率和准确性。在经验回放中，智能体构建一个大小为 M 的先入先出（FIFO, first input first output）队列来存储经验，因此该队列被称为经验池 \mathcal{M} 。然后，每次更新权值 θ 时，智能体将从经验池中随机调用 B 个经验组成经验集合 \mathcal{B} 来进行批量梯度下降。在拟静态目标网络中，智能体将构建 2 个 DQN，一个用于实时训练（称为训练 DQN），另一个则被用于获得较稳定的目标值估计值（称为目标 DQN）。每隔 K 时刻，智能体将当前时刻的训练 DQN 同步为新的目标 DQN。

以上介绍的是 DeepMind 团队于 2015 年提出的深度 Q 学习 (DQL, deep Q-learning) 算法^[3]。此后, 在 DQL 算法基础上相继发展出了双深度 Q 学习 (double deep Q-learning) 算法^[17]和竞争深度 Q 学习 (dueling deep Q-learning) 算法^[18]。上述算法都需要借助值函数, 因而被称为基于值的 DRL 方法。

虽然基于值的 DRL 方法允许连续的状态空间, 但是它们在动作选取时依旧需要遍历所有动作的值函数。因此, 它们只能被用于解决具有离散动作空间的问题。为了既能实现连续动作的选取, 又能提高数据利用率和决策的稳定性, 人们结合基于策略和基于值这 2 种方法的特点提出了深度确定性策略梯度 (DDPG, deep deterministic policy gradient) 算法^[19]。DDPG 使用了 2 个分别称为动作家 (actor) 和评论家 (critic) 的 DNN。其中, 评论家用于评估值函数, 而动作家则用于建立值函数与动作的映射关系。相较于基于值的 DRL 方法, DDPG 直接从动作家获得动作, 而不是选择值函数最大的动作。具有动作家和评论家架构的方法称为基于动作评论家 (AC, actor-critic) 的 DRL 方法。除 DDPG 外, 近端策略优化 (PPO, proximal policy optimization)^[20]和异步优势动作评价 (A3C, asynchronous advantage actor-critic)^[21]也是基于 AC 的代表算法。

2.3 联邦学习概述

DL 和 DRL 都需要利用大量数据对 DNN 进行训练。传统的训练方法首先需要假设存在一个完整的训练数据集 \mathcal{D} , 然后通过优化 DNN 的权值使 DNN 可以对数据集 \mathcal{D} 进行最优拟合。其中数据集 \mathcal{D} 由源数据集 $\mathcal{X} = \{x_1, \dots, x_N\}$ 和相应的目标数据集 $\mathcal{Y} = \{y_1, \dots, y_N\}$ 构成。若 DNN 的权值记为 θ , 则训练 DNN 的目标是最小化损失函数

$$L(\theta) = \frac{1}{N} \sum_{i \in \mathcal{D}} L_i(\theta) \quad (1)$$

其中, $L_i(\theta)$ 是描述 DNN 在权值 θ 下对 x_i 的输出与其真实目标数据 y_i 间差异的损失函数。因此, θ 应该向 $L(\theta)$ 梯度下降的方向更新, 即

$$\theta' = \theta - \frac{1}{N} \sum_{i \in \mathcal{D}} \frac{\partial L_i(\theta)}{\partial \theta} \quad (2)$$

通过数据集 \mathcal{D} 和式(2)可以对 DNN 的权值进行迭代更新, 并最终获得最优的权值。然而, 在无线通信系统中, 建立这样一个完整的训练数据集往往

是难以实现的。一方面, 无线通信系统中的数据具有天然的分布式特性, 在数据收集的过程会产生巨大的通信开销。例如, 系统中各个用户的通信设备都会自主产生大量数据。另一方面, 这些数据有可能包含用户个人信息或归属不同的利益集团。出于对个人隐私和数据安全的考虑, 数据的拥有者不会轻易地将数据与不信任的第三方共享。基于这一背景, FL 被提出用于多个智能体在不需要共享训练数据的前提下对 DNN 进行协作训练^[4]。

通过观察式(2), 若将 \mathcal{D} 划分成多个子集 $\mathcal{D}_1, \dots, \mathcal{D}_j$, 则有

$$\theta' = \theta - \sum_{j=1}^J \frac{N_j}{N} \sum_{i \in \mathcal{D}_j} \frac{\partial L_i(\theta)}{\partial \theta} \quad (3)$$

其中, N_j 是数据集 \mathcal{D}_j 的大小。从式(3)可以看出, θ 基于数据集 \mathcal{D} 的更新可以分解为 θ 关于多个子数据集的更新。利用这一特性, FedAvg 算法^[8]被提出用于多个拥有本地数据集的智能体共同训练一个全局 DNN。其训练过程主要包含 2 个迭代的步骤, 即本地训练和全局更新。在本地训练中, 智能体首先从服务器下载最新的全局 DNN 权值, 然后利用本地数据集对 DNN 进行训练, 并将训练后的 DNN 权值与此前下载的全局 DNN 权值的差值 (又称权值更新值) 发送至服务器。在全局更新中, 服务器将从各智能体收集得到的权值更新值进行平均运算后再与全局 DNN 的权值进行合并。

2.4 小结

作为支撑 AI 的三类重要的机器学习方法, DL、DRL 和 FL 均有其特定的适用性。其中, DL 的本质是利用 DNN 对源数据和目标数据进行拟合, 建立源数据和目标数据之间的关系。因此, DL 适用于数据预测和判决问题。DRL 通过试错 (trial-and-error) 的过程在与环境不断交互中学习隐含的规律并据此做出最优决策, 因而适用于解决动态系统或动态环境中的决策问题。针对 DL 和 DRL 中存在的隐私问题, FL 通过仅传输 DNN 权值来避免数据泄露, 从而实现了原始数据的保护。因此, FL 适用于所有需要多智能体协作训练, 但对隐私保护有较高要求的问题。

下面, 本文将具体阐述这三类机器学习方法在无线通信系统中的应用, 包括无线传输、频谱管理、资源配置、网络接入、网络及系统优化 5 个方面,

以支持智能无线通信的实现。

3 无线传输

3.1 信道估计

在无线通信中，发送端向接收端发送的信号会受信道的影响发生畸变。因此，为了正确检测发送信号，接收端需要对信道状态信息（CSI, channel state information）进行估计并以此对信号进行均衡处理。信道估计通常包括2个步骤：首先发送端向接收端发送导频信号；然后接收端根据接收到的导频对CSI进行估计。未来无线通信系统的用户规模和天线阵列规模都将急剧增大，这会给传统的信道估计方法带来挑战。一方面，能用于精确估计信道的正交导频资源有限，而采用非正交导频会极大降低信道估计的准确性。另一方面，随着系统规模增大，经典信道估计算法的计算复杂度也在攀升。鉴于信道估计是一个从接收导频推测CSI的过程，DL可以有效地解决这一类问题。

受传统最小均方误差（MMSE, minimum mean square error）信道估计器的启发，文献[22]提出了一种基于DL的MMSE信道估计方案。作者首先根据MMSE估计器的结构设计DNN结构，在信道协方差矩阵为托普利兹矩阵（Toeplitz matrix）的假设下，利用CNN来降低DNN的计算复杂度。仿真结果表明，在协方差矩阵不是托普利兹矩阵的3GPP信道模型中，该方案依然能在较低的运算复杂度下获得比传统信道估计方法更低的信道估计误差。同样是从传统信道估计方法出发，文献[23]针对导频非正交分配下的信道估计问题，将DNN与最小二乘信道（LS, least square）估计器相结合。其中，DNN被用于对因采用非正交导频造成的导频污染进行降噪，而LS估计器被用于利用降噪后的导频进行信道估计。在处理高维信号时，作者证明了该方案能以更低的计算复杂度获得与MMSE相同的渐进性能。此外，文献[24]考虑了单频相移键控信号的信道估计问题。对于单频信号而言，其CSI由载频偏移值（CFO, carrier frequency offset）和定时偏移值（TO, timing offset）决定。因此，文献[24]提出使用DNN建立接收导频与CFO和TO间的映射关系，从而对CSI进行准确估计。

值得注意的是，文献[22-24]采用了传统导频设计方案。为了进一步提升性能，文献[25]利用DL对导频设计和信道估计进行联合设计。在该文献

中，作者考虑一个多输入多输出（MIMO, multi-input multi-output）系统中的多用户上行信道联合估计问题。为了得到符合信道特性的最优导频序列和信道估计器，作者基于自动编码器将DNN设计为导频设计模块和信道估计模块两部分。在利用真实CSI数据集对DNN进行离线训练后，得到的导频设计模块和信道估计模块可以分别在发送端和接收端进行在线部署。仿真结果显示，该方案产生的均方误差（MSE, mean square error）比经典线性最小均方误差（LMMSE, linear MMSE）算法产生的MSE低40%。

3.2 信道预测和反馈

除了在接收端需要进行信道估计外，为提升频谱效率，MIMO系统中的发送端也需要获取CSI对信号进行预编码，从而进行多信息流的并行传输。然而，这在采用频分复用（FDD, frequency division duplexing）的大规模MIMO系统难以进行实现。在大规模MIMO系统中，基站的天线数远大于用户数及用户终端的天线数。因此，从用户到基站的上行CSI在基站端容易估计得到。然而，基站到用户的下行CSI数据量庞大，若让用户终端进行信道估计并直接向基站进行反馈，回传链路会因此过载^[26]。与此同时，由于FDD系统中上下行信道频率不同，基站也无法从上行CSI直接得到下行CSI，即不存在信道互异性。为此，文献[27-28]尝试通过信道预测来解决这一问题。

虽然FDD系统中的上下行信道不存在严格的互异性，但是其信道存在共性。例如，因为信号经过的散射物体是一致的，所以其多径方向大致相同^[29]。基于这一原理，文献[27-28]提出使用DL来挖掘上行CSI与下行CSI之间的映射关系。其中，文献[27]设计了一个由卷积层和全连接层构成的DNN来提取上、下行CSI的映射关系。仿真结果显示，该方案在不同信道数据集上的性能差异较大，其稳定性欠佳。文献[28]进一步利用多径数量通常少于接收天线数的稀疏性设计了一个稀疏复值神经网络（SCNet, sparse complex-valued neural network）来对上、下行CSI的关联性进行分析。SCNet通过减少中间层的神经元个数来迫使DL提取上、下行CSI共同包含的关键共性信息，从而获得更高的预测准确率。

降低反馈开销是另一种解决思路。文献[30-31]利用自动编码器对CSI进行压缩，通过仅传输CSI压缩值来降低回传链路负载。其中，文献[30]考虑

了室内环境下的大规模 MIMO 系统。在该环境下,信道矩阵具有低秩特性。因此,作者首先对信道矩阵进行角度域的二维离散傅里叶变换后去除零元素,然后提出了一个名为 CsiNet 的自动编码器对非零元素进行压缩编码和重构。CsiNet 的编码器部分模仿压缩感知算法中的投影操作,由卷积层和全连接层构成,而 CsiNet 的解码器部分则由提炼网络 RefineNet 构成。仿真结果表明,CsiNet 能够获得比压缩感知算法更高的 CSI 重构质量。文献[31]进一步考虑了信道时变带来的影响。为了挖掘相邻时刻 CSI 的相关性,文献[31]在 CsiNet 基础上引入了 LSTM 层,提出了 CsiNet-LSTM。CsiNet-LSTM 对多个时刻的 CSI 进行同时处理,提高了对时变 CSI 的压缩和重构质量。

3.3 信号检测

信号检测是实现无线传输的重要组成部分,其目标是从接收信号恢复发送信号。传统的信号检测方法需要首先根据发送信号、信道和噪声的先验概率来构建接收信号的后验概率,然后依据后验概率进行判决。显然,若系统对信道和噪声的先验知识不完美,信号检测准确率将会降低。此外,当发送信号比较复杂时,例如在 MIMO、正交频分复用 (OFDM, orthogonal frequency-division multiplexing) 等技术的引入后,信号检测的复杂度会极大地提升。作为典型的判决问题,信号检测问题可以被 DL 高效求解。

在 MIMO 系统中,现有的文献提出了最大似然 (ML, maximum likelihood)、消息传递 (MP, message passing)、近似消息传递 (AMP, approximate MP) 等经典检测算法。这些算法基于迭代结构,其计算复杂度较高。文献[32-35]从经典的 MIMO 信号检测算法出发,将它们的迭代结构展开成 DNN 中的层结构,即 DNN 中的每一层对应原算法中的一次迭代。以文献[32]为例,作者利用 DL 将正交 AMP (OAMP, orthogonal AMP) 检测器展开。为了避免在直接计算后验概率时对高维矩阵进行积分,OAMP 算法需要迭代地近似接收信号的后验概率。其中,有部分迭代变量需要人工设置更新步长,而不合适的步长会导致过多的迭代步数且增加计算复杂度。为此,文献[32]设计了一种 OAMP-Net,它将每一次迭代运算展开为一层神经元的内部连接,而更新步长由神经元的权值决定。在给定 DNN 层数后,OAMP-Net

可以从大量数据中学习迭代变量每次更新的最优步长。仿真结果表明,OAMP-Net 可以在极低的运算复杂度下获得比 OAMP 和 LMMSE 更低的误比特率。基于类似的思想,文献[33-35]分别利用 DNN 对 MP、共轭梯度下降 (CGD, conjugate gradient descent) 和 ML 检测算法进行了展开,将其中需要人工调节的更新步长转变为可以从训练数据中学习得到的神经元权值。仿真结果显示,它们都能以更低的计算开销获得比原经典算法更优的检测准确率。

文献[32-35]提出的方案依赖于准确的信道估计。由于信道估计的误差会传播到信号检测中,信号检测的准确率将会降低,文献[36]提出将信道估计与信号检测进行联合处理,考虑的是 OFDM 系统,其中每个 OFDM 帧包含导频和发送数据。作者采用了一个由全连接层构成的 DNN 将接收到的 OFDM 帧直接映射为发送数据。仿真结果表明,该方案在导频数量较少或缺少循环前缀时可以获得比传统的 LS 和 MMSE 检测器更优的性能。更进一步,文献[37-39]利用自动编码器设计了端到端的信号编码与检测系统。虽然文献[37-39]分别考虑了单输入单输出 (SISO, single-input single-output)、单输入多输出 (SIMO, single-input multi-output) 和 1 bit OFDM 这 3 种不同的无线通信系统,但是它们所采用的方法是类似的。首先,一个 DNN 被设计用于模拟通信系统的编码和解码过程,其输入为发送端的原始信号,而输出则是接收端解码后的信号。然后,引起信号畸变的信道响应被抽象为 DNN 中的一层。在使用大量信道数据训练后,自动编码器可以根据发送信号和信道的内在特征对导频、信道估计和信号检测进行最优的联合设计。当训练完成后,自动编码器的编码模块和解码模块将分别部署于发送端和接收端来进行通信。仿真结果表明,由于基于 DL 的端到端的设计方案可以充分挖掘和利用原始信号及信道的先验知识,它们相较于传统方法可以取得明显的性能优势。然而,端到端的设计方案完全依赖于原始信号和信道的训练数据集,它们在面对未知环境下的泛化性值得探讨,而目前尚未有相关研究。

4 频谱管理

4.1 频谱感知

频谱是进行无线通信的基础,但频谱资源却是

有限和稀缺的。为了缓解激增流量和有限频谱之间的矛盾，认知无线电（CR, cognitive radio）被提出用于提高频谱利用率^[40]。CR 的基本思想是允许次用户（SU, secondary user）在拥有频谱的主用户（PU, primary user）空闲时使用频谱。显然，准确地判断 PU 是否在使用频谱是实现 CR 的关键，相关技术被称为频谱感知（spectrum sensing）技术。与传统的信号检测和调制识别方法类似，传统的频谱感知技术通过构建检验统计量的似然函数来对 PU 状态进行判断，例如根据接收信号功率进行判断的能量检测器（ED, energy detector）。由于依赖信道或噪声分布等先验知识，传统的频谱感知算法存在信噪比墙等问题，其准确率受限。

显然，频谱感知本质上是对主用户是否存在的判决问题，因而它可以被 DL 高效地求解，而 CNN 在其中得到广泛应用^[41-44]。例如，文献[41]考虑了一个单用户多天线的频谱感知系统。作者根据信号协方差矩阵与图片的相似性，采用 CNN 从接收信号的协方差矩阵提取检验统计量，然后基于该检验统计量进行判决。在独立同分布的信号模型下，该方案被证明与已知信号先验知识（如统计协方差矩阵）的 ED 等价。而在指数相关的信号模型下，仿真结果显示该方案性能远远优于 ED、最大特征值检测器等经典算法。文献[42]在文献[41]的基础上加入了对 PU 活动规律的识别，以进一步提高频谱感知的准确度。和基于检验统计量的方法不同，文献[43-44]利用 DL 直接对接收信号进行识别，避免了对检验统计量这一中间值的处理。其中，文献[43]考虑了对 3.5 GHz 雷达信号的识别问题。作者将 CNN 与 LSTM 结合，然后通过大量训练数据使 DNN 学习到接收信号频谱瀑布图序列和是否存在雷达信号之间的关系。文献[44]则考虑了多用户联合频谱感知的问题。为了使用 CNN 处理频谱信息，作者将多个 SU 在不同频点上的接收信号强度拼接成矩阵，然后再将该矩阵当成图像数据输入 CNN 中处理。

上述文献所提方案都需要利用当前时刻的接收信号进行判决。然而，频谱感知是一个对时效性要求严苛的任务。因此，利用频谱历史数据进行频谱态势的分析和预测是实现频谱感知的另外一种思路。文献[45]中提出用 LSTM 建立历史频谱占用信息与当前频谱占用情况的关系。仿真结果显示，该方案可以利用过去 60 min 的历史频谱数据以接

近 95% 的准确率预测当前时刻的频谱占用情况。文献[46]同样采用 LSTM 设计了基于 DL 的频谱占用情况预测方法。在该方法中，一种启发式算法被用于自动优化 DNN 超参数，包括 DNN 的神经元个数、层数、激活函数、学习速率以及权值初始值。仿真结果表明，优化后的 DNN 设计比没有优化或仅采用全连接层的 DNN 设计在地面数据集和卫星数据集上均能获得更高的准确性和稳定性。

4.2 信道接入

一般地，频谱会先被划分成多个信道，然后通信系统以信道为基本单位接入频谱。提高频谱利用效率有赖于精准和高效的信道接入方法。由于信道接入问题是对信道和接入时间的决策问题，DRL 提供了实现最优信道接入的决策工具。

文献[47-48]关注单信道的接入问题。其中，文献[47]考虑一个智能用户与其他多个采用不同接入策略的用户共享同一频段进行数据传输。当多于一个用户在同一时刻发送数据时，数据分组会因为碰撞而发送失败。因为各个用户采用的接入策略最终会反映在频谱占用情况中，所以作者将历史的频谱观察数据作为 DRL 的状态。通过对频谱数据的分析，DRL 最终能预测并充分利用频谱中的空闲时隙进行传输，以提高频谱的利用率。文献[48]考虑了长期演进（LTE, long term evolution）蜂窝移动通信系统与无线局域网（Wi-Fi）这 2 个异构系统在非授权频谱上的共享问题。传统的 LTE 与 Wi-Fi 共享技术需要基于 2 个异构系统的即时且完整的信息来进行频谱管理，从而使 LTE 系统在保护 Wi-Fi 系统（即满足 Wi-Fi 系统流量需求）的前提下充分占用空闲频谱。然而，这 2 个异构系统是相对独立的，难以进行信令交互。根据频谱信息与 Wi-Fi 流量等信息存在的关联性，作者提出利用 DRL 直接对频谱信息进行分析，通过预测 Wi-Fi 流量信息来智能地调整 LTE 和 Wi-Fi 的时隙分配。仿真结果表明，尽管缺少信令交互，该方案可以在充分保护 Wi-Fi 系统的前提下提高 LTE 系统的传输速率，且能够逼近完美已知信息下的理论最优性能。

在多信道的场景下，信道接入还需要对信道进行选择^[49-51]。文献[49]中考虑了一个单用户多信道的频谱接入问题。其中，各个信道的质量根据某种隐藏的规律动态变化，而用户在同一时刻仅能对某一信道的质量进行感知。因此，这是一个部分可观

察马尔可夫决策过程 (POMDP, partially observable Markov decision process) 问题。在使用 DQL 算法来解决这一问题时, 作者首先将状态设计为过去多个时刻选择的信道和相应的信道质量感知结果, 以弥补 POMDP 中的观察信息缺失问题, 然后将动作设置为当前时刻需要感知和发送数据的信道。当用户选择的信道质量比较好时, 数据分组能被成功发送, 奖赏值就设置为正值, 否则为负值。此外, 作者还针对环境规律可能发生变化的情况提出了自适应的重训练算法, 根据累积奖赏的下降是否超过某一阈值来感知环境规律的突变, 进而决定是否对 DQN 进行重新训练。仿真结果表明, 该方案可以逼近已知信道质量时得到的最优性能。文献[50]进一步研究了多用户情况下的信道动态接入问题。为了避免多用户同时使用同一信道而产生数据分组碰撞, 作者提出了一个基于 DRL 的分布式信道接入方案。其中, 每个用户都是独立的智能体。各个智能体将其发送记录 (即数据分组是否成功发送) 作为状态, 并且将发送当前数据分组选择的信道作为动作。若该数据分组被成功发送, 智能体将获得奖赏为 1, 否则为 0。为了获得更高的性能, 作者结合竞争深度 Q 学习和双深度 Q 学习提出了竞争-双深度 Q 学习算法。此外, LSTM 还被用于帮助分析状态中包含的多个时刻的数据。仿真结果显示, 所提方案可以在用户没有信令交互时得到避免发生碰撞的信道接入方式。

在有恶意干扰器的复杂电磁环境中, 例如军事通信中, 信道接入问题变得更复杂。文献[51]提出了基于 DQL 的抗干扰信道选择算法。在该文献中, 智能体将频谱瀑布图作为状态, 通过分析其中存在的干扰模式来主动预测并选择未受干扰的信道进行信息发送。为了在保证通信质量的同时减少信道切换, 智能体的奖赏由一个与信干噪比 (SINR, signal to interference plus noise ratio) 相关的回报值和信道切换开销值相减得到。此外, 作者在设计 DQN 时, 根据频谱瀑布图存在的递归特性将传统 CNN 改进为递归卷积神经网络 (RCNN, recursive convolutional neural network), 从而降低计算复杂度。

4.3 调制识别

监测频谱非法占用情况是频谱管理的另一项重要任务。为了检测非法信号并定位信号源, 往往需要对信号进行特征识别和解码。这就需要无线

信号的调制方式进行盲估计, 即调制识别。传统的调制识别算法基于似然函数进行求解, 需要根据发送信号的先验知识计算后验概率, 而后再进行判决。然而, 一般情况下发送信号的先验知识是难以获得的。此外, 当调制方式较多时, 后验概率的计算也会变得极其复杂。为解决这一问题, 人们通过使用 DL 来避免对似然函数的复杂求解过程, 转而利用 DNN 进行直接判决。

文献[52-53]提出了基于高阶累积量的信号特征提取方案, 它们利用 DL 建立已知信号特征与其调制方式之间的关系, 从而对调制方式进行识别。在文献[52]所提方案中, 接收信号首先被预处理为 4 阶和 6 阶累积量, 然后再经过一个全连接的 DNN 对调制方式进行识别。文献[53]则采用高阶累积量的比值作为信号的特征, 然后采用 CNN 对调制方式进行识别。文献[52-53]分别对 6 种和 11 种调制方式进行了测试。结果显示, 它们在多径信道和加性白噪声信道中均能实现较高的识别成功率, 且文献[52]所提方案甚至能够对存在频率偏移的非完美接收信号进行较高成功率的识别。

文献[52-53]提出的调制识别方案是基于人工提取的信号特征设计的。显然, 人工选择特征是一个烦琐的过程, 且选择的特征通常难以普遍适用。因此, 文献[54-61]提出利用 DL 对信号进行自动的特征提取和调制识别。其中, 文献[54-57]提出先将接收信号进行图像化预处理, 再采用 CNN 对图像进行分析。它们采用的图像化预处理方法略有不同: 文献[54-55]使用信号的星座图, 而文献[56-57]则采用信号的频谱瀑布图。值得一提的是, 文献[55]提出了一种双层 CNN 架构, 来同时使用信号的星座图和 IQ 原始数据: 原始信号经过第一层 CNN 处理后, 输出的数据连同信号的星座图再通过第二层 CNN 处理得到调制方式的预测值。另一方面, 采用同一调制方式的发送信号通常不是孤立出现的, 因而对接收信号的时间序列进行联合处理会带来更好的性能。基于这一思路, 文献[58-61]将 LSTM 用于调制识别中。其中, 文献[58-60]将 LSTM 与 CNN 相结合, 获得比仅基于 CNN 方案更好的识别成功率。文献[61]则在 LSTM 的基础上加入了注意力 (attention) 机制, 对接收信号的时间序列进行加权, 通过自适应地调整序列中各信号的重要性来解决噪声不确定下的调制识别问题。仿真结果表明, 该方案在 0~20 dB 的信噪比下获得

的识别准确率均接近已知噪声功率的传统 ML 算法的性能。

5 资源配置

5.1 功率分配

无线信道具有开放性，因此当多个用户共用同一时频资源传输信息时会产生复杂的干扰问题。功率分配是实现干扰管理的有效手段。

文献[62]研究了一个 CR 中的功率分配问题。其中，SU 和 PU 使用相同的信道同时进行传输，而 SU 的目标是通过调节其发射功率来使 PU 和 SU 的 SINR 都能大于某一阈值。为了实现这一目标，SU 首先在 PU 附近放置了多个接收功率传感器，然后利用 DRL 根据传感器的接收功率来推测 PU 传输模式和信道变化规律，并据此选择一个最优的发射功率。因此，DRL 的状态和动作分别为所有传感器的接收功率和 SU 的发射功率。当 SU 选择了一个功率进行发送后，若 PU 和 SU 的 SINR 都大于阈值，则奖赏为 1，否则为 0。最后，作者选择 DQL 算法用于实现该基于 DRL 的功率分配方案，其性能逼近需要 PU 和 SU 完全协作的经典优化算法。但是，该方案仅考虑了单用户的场景，对于涉及多用户的问题无法使用。

进而，文献[63]研究多用户蜂窝网络中的功率分配问题。在该问题中，各个用户通过调整发射功率来最大化整个系统的和速率。求解该问题存在 2 个难点：一是该优化问题是非凸的，难以获得最优解；二是用户之间的 CSI 数据量随着用户数的增加而呈指数增加，因而获取所有用户的 CSI 进行求解所要求的信令开销是难以承受的。为此，作者提出了一种分布式 DQL 算法，使各用户成为独立的智能体，根据自身可获取的局部信息选择发射功率。然而，多个智能体的存在会产生非平稳性的问题，即某个智能体的外部环境会因其他智能体的动作而产生变化，而并非完全由真正的环境变化规律决定。此时，DRL 就有可能无法收敛或者收敛性能较差。针对该问题，作者从 3 个方面进行解决。首先，作者在设计各用户的状态时，不仅包含自身的本地 CSI 等信息，也包含相邻用户的历史信息，从而使各用户能够推测未来的 CSI 以及其他用户的发射功率。其次，作者将系统和速率这一全局优化目标分解为各个用户的子目标，把用户自身速率与对其他用户产

生的干扰的差值作为奖赏值。这一设计让各用户的优化目标不会互相冲突，有利于算法收敛。最后，作者提出了一个集中式训练的框架，通过集中式训练让各用户具有相似的决策逻辑，从而保证算法的收敛性能。仿真结果显示，该分布式 DQL 算法得到的系统和速率可以超越经典的分式优化(FP, fractional programming)和加权 MMSE (WMMSE, weighted MMSE) 优化算法。

此外，文献[64-65]将功率分配问题和信道接入问题进行了结合，联合优化发射功率和信道。其中，文献[64]将文献[63]所提方案拓展到多信道的 D2D 通信系统中。和文献[63]相比，文献[64]除了将 DRL 的状态、动作和奖赏从单信道拓展为多信道外，还提出了完全分布式的 DQL 算法。在该算法中，各用户不再需要集中式训练，避免了用户向中央训练单元上报数据而产生的通信开销。此外，文献[64]还发现了通信开销与系统性能之间存在折中关系，当每个用户与更多的相邻用户共享历史信息时，整个系统获得的和速率会越大。文献[65]同样采用 DRL 来解决车联网中的功率分配和信道选择问题。在车联网中，通信时延关乎车辆的运行安全。为此，作者通过分别将系统容量和通信时延设计为构成 DRL 奖赏的回报值和惩罚值，使车联网可以在最大化系统容量的同时降低通信时延。仿真结果显示，与随机策略和启发式算法相比，基于 DRL 的方案在用户平均速率、系统容量和时延满足概率等方面均有明显的性能提升。

5.2 计算资源配置

在 VR 等计算密集型移动业务出现的同时，通信设备却在向小体积、低功耗的方向发展。移动边缘计算(MEC, mobile edge computing)因而被提出用于辅助计算能力和续航能力有限的终端以完成计算复杂度较高的业务。在 MEC 中，计算单元被配置在网络接入侧(一般为基站)，便于快速响应终端的计算请求。

然而实际上，MEC 的计算资源也是有限的，需要根据用户需求弹性配置。文献[66]研究了 MEC 中的服务配置问题，其将计算任务划分为多类服务，而基站在接受用户请求前需要先为各类服务进行资源划分和保留。为了实现最优的配置，基站首先需要对用户使用的各类服务偏好进行分析和预测。因为用户的服务使用记录是敏感数据，所以作者基于 FL 的思想提出了分布式 DL 算法，

让多个基站不需要共享用户信息就可对所有用户的服务偏好进行建模。

此外,向 MEC 卸载计算任务也是一个需要综合考虑费用、通信时延、计算时延、电能损耗等因素的复杂决策问题。文献[67]提出了一种分布式的计算任务卸载决策方案。该问题针对的是能量和计算任务随机到达的基于能量采集的物联网 (IoT, Internet of things) 设备。为了完成计算任务, IoT 设备可以选择将计算任务卸载到 MEC 中, 但会因此带来通信时延和 MEC 服务开销。若 IoT 设备选择本地执行计算任务, 则它需要为本地计算单元分配能量: 更多的能量产生计算时延更低, 反之亦然。此外, 计算任务需要在规定时间内完成。因此, 计算任务卸载和能量分配需要在兼顾时延要求的情况下降低 MEC 服务开销。为了学习和利用环境中的随机规律 (如计算任务达到规律) 做出最优决策, 作者提出了基于 DRL 的方案。特别地, 作者基于 FL 将 DQL 算法改进为联邦-深度 Q 学习算法, 让多个 IoT 设备不需要暴露隐私数据便可共同训练, 提高决策准确度。

6 网络接入

移动设备数量的激增也催生了超密集网络 (UDN, ultra dense network) 等技术的诞生, 通过增加基站的数量来提供高速和可靠的连接。这意味着同一用户终端在任何时候都能被多个基站的信号所覆盖, 而用户接入网络时需要选择一个基站进行连接。值得注意的是, 信道、用户需求及位置等因素是动态变化的。为持续地保证接入性能, 用户需要从全局和长期的角度对连接的基站进行动态选择和切换。

文献[68]考虑 UDN 中的基站接入问题。在传统的基站接入机制中, 若用户接收到其他基站的参考信号接收功率 (RSRP, reference signal received power) 比当前连接基站的 RSRP 高于某一阈值, 用户就切换到 RSRP 最大的基站。这一阈值的设置是为了解决切换中的乒乓效应, 但这在 UDN 中是不够的。在 UDN 中, 基站分布密集使每个基站只在其附近较小范围内的 RSRP 比较大。因此, 当用户在热点区域比较集中时, 该区域内的基站负载极高, 而其他区域的基站却非常空闲。此时, 与邻近的高负载基站相比, 用户连接远距离的空闲基站得到的通信质量可能会更

高。为了实现基站负载均衡, 作者提出在原有切换阈值上加入偏置值, 并利用 DRL 中的 A3C 算法对偏置值进行智能选择。其中, DRL 的状态设计为各基站的负载和边缘用户比例, 而奖赏被设计为所有基站最大负载的倒数, 即所有基站达到相同负载可以将奖赏最大化。仿真结果显示, 与基于 RL 或静态规则等传统负载均衡方案相比, DRL 可以有效地降低用户接入失败概率。

文献[69]结合信道接入对基站接入问题进行研究。其中, 作者考虑了一个由多个发射功率不同的基站所构成的多层蜂窝系统。特别地, 拥有更大发射功率的基站覆盖范围越大, 但同时也因为更可能被过多的用户连接而更容易出现拥塞和速率下降。此外, 当连接的基站发生切换时, 用户数据需要在不同基站间进行迁移, 会带来额外开销。为了在满足各用户的速率需求的同时降低切换次数, 作者提出了分布式的竞争-双深度 Q 学习算法, 让各用户成为独立的智能体, 在每个时刻自主选择接入的基站与信道。与文献[63]类似, 多个智能体的存在也会导致 DRL 的非平稳性问题。为解决该问题, 作者令所有用户共享同一状态, 其中包含表示各个用户的速率需求是否得到满足的指示值。除此之外, 各用户的奖赏值被设计为该用户的速率与发射功率、切换开销的差值, 使用户在尽量降低切换次数和对其他用户产生干扰的同时最大化自身速率。与 RL、贪婪算法等传统方法相比, 文献[69]提出的竞争-双深度 Q 学习算法在用户数较少时性能几乎一致, 但在用户数较多的复杂场景能够取得明显优势。

为了满足各种各样的通信需求, 当前的无线网络通常包含多种采用不同无线电接入技术 (RAT, radio access technology) 的无线通信系统, 形成异构网络。因为采用不同 RAT 所提供的服务特点及开销均不同, 所以用户接入网络时需要根据自身需求选择最优的 RAT。文献[70]考虑 LTE 和 Wi-Fi 这 2 种 RAT 的接入问题。显然, LTE 通过授权频谱提供服务, 因而服务费用较高。同时, 因为 LTE 能提供更高的速率, 所以通过 LTE 传输数据所需的单位能量开销较低。Wi-Fi 则与之相反。用户的目标是以最低的费用和能量开销在规定时间内完成文件传输。为求解该问题, 文献[70]使用 DQL 算法让用户学习最优的 RAT 切换控制策略。其中, 时间被分成了多个时隙。在每个时隙开始, DQL 算法将根据包

含了用户位置和文件剩余大小的状态，选择当前时隙的动作，即选择使用 LTE 还是 Wi-Fi 连接网络。在做出选择后，用户将会获得一个由费用、能量开销和文件传输失败惩罚值决定的奖赏值，用以引导 DQL 算法实现规定的目标。作者通过仿真证明了 DQL 算法相较于动态规划和启发式算法能够有效地降低费用及能量损耗。

7 网络及系统优化

在传统的无线通信系统中，网络和系统优化极大地依赖人工参与。例如，基站部署需要人工规划，而故障排除通常也需要工程师介入。因此，随着网络规模的扩大以及各类复杂新技术的加入，无线通信系统的管理和维护成本将会攀升。人工智能的引入有望实现无线通信系统的自优化，在提高系统性能的同时降低成本。

近年来，智能反射表面 (IRS, intelligent reflecting surface) 被提出用于优化无线网络的覆盖和速率。IRS 包含众多可调相位的被动反射单元。通过调整相位，无线信号的传播路径得以改变，从而增强特定位置处的信号接收强度。文献[71]提出使用 DRL 对 IRS 反射单元的相位进行优化。该文献考虑多输入单输出 (MISO, multi-input single-output) 下行蜂窝系统，而 IRS 被用于提高用户接收信号的信噪比 (SNR, signal-to-noise ratio)。由于 IRS 的反射单元数量众多，且其相位可以为连续值。因此，作者采用了能够输出连续动作空间的 DDPG。DRL 的状态包含了用户上一时刻的 SNR 及 IRS 各反射单元的相位，而用户接收信号的 SNR 则被用作 DRL 的奖赏。仿真结果显示，基于 DRL 的 IRS 优化方案所取得的性能逼近具有高计算复杂度且需获取完整 CSI 的传统半正定松弛 (SDR, semi-definite relaxation) 算法。

当蜂窝系统的基站数量越来越多时，其运行产生的能耗变得不可忽视。实际上，由于用户流量需求在时间、空间上是不均匀的，系统可以通过关闭或休眠基站来降低能耗。基于这一思考，文献[72-73]分别提出了基于 DRL 的基站管理技术。文献[72]中考虑了单个基站的场景，其目标是让基站根据网络流量来动态地选择休眠或唤醒，从而降低能量损耗。为了实现这一目的，作者提出了 2 种基于 DQL 算法的方案。在第一种方案中，过去多个时刻的网络流量元数据，包括数

据分组的种类、大小等信息以及相应的基站动作被用作 DRL 的状态，然后让 DQL 据此选择下一时刻的动作。第二种方案则是采用隐马尔可夫模型对历史流量进行建模并对下一个时刻的流量情况进行预测，然后让 DQL 根据预测值选择动作。为了使基站能够在服务质量和能量开销间取得折中，奖赏值包含了满足的请求数、等待请求数、失败请求数、基站运行能量开销和切换基站状态开销 5 个变量。仿真表明，通过隐马尔可夫模型的辅助，DQL 在大多数的流量数据集中可以更准确地预测流量并做出更好的决策。文献[73]进一步考虑了多基站的场景。其中，作者首先将 DL 用于分析和预测用户流量分布，然后再利用 DRL 根据 DL 的预测结果对基站进行开关控制。值得注意的是，所有基站的开关组合与基站数量呈指数关系。为了容纳庞大的动作空间，作者采用了 DDPG 算法，将每一种开关组合映射到连续动作空间中。在设计奖赏时，作者将用户体验、能耗和开关切换开销 3 个因素考虑其中，使 DRL 在优化能耗的同时最小化切换次数和保证用户体验。仿真结果显示，该方案的性能相较于 RL、随机策略等方法更接近穷搜算法得到的最优性能。

针对用户流量需求时空分布不均的问题，另外一种解决办法就是通过部署无人机 (UAV, unmanned aerial vehicle) 空中基站来为热点区域的用户提供流量卸载服务。UAV 的移动特性为基站部署带来了灵活性，但同时也使无线信道难以预测。文献[74]采用 DQL 算法对 UAV 基站进行智能部署。作者首先将 UAV 和用户的坐标设计为 DRL 的状态，然后将 UAV 在空间中各个移动方向作为动作。为了最大化系统容量，DRL 的奖赏被设计为所有用户从空中基站和地面基站得到的速率之和。仿真结果表明，该算法在 4 种用户分布模型下均能接近理论可达的最大速率，且其时间复杂度低于 RL 和爬山 (hill climbing) 算法。

此外，文献[75-76]分别研究了系统故障修复和异常检测问题。文献[75]假设无线通信系统在运行过程中会随机出现故障，而故障需要通过一系列排障操作来解决。作者提出使用 DQL 算法来学习如何通过排障操作来修复故障。其中，DRL 的状态是系统的故障数变化，而动作则是选择一个排障操作。为了使系统能在最短的时间内排除

故障, DRL 的奖赏被设计为一个与排除故障所耗时间负相关的函数。仿真结果显示, 所提方案有效地提高了系统的可用性。文献[76]则针对 IoT 网络提出了一种异常设备检测方法。作者首先对 IoT 设备进行了分类, 然后提出了一种 DL 算法, 让 IoT 网关对数据分组所属设备类型的时间序列进行建模, 即根据过去数据分组的设备类型序列预测下一数据分组可能为各个设备类型的概率。显然, 只用一个 IoT 网络的数据训练得到的模型泛化性较差, 但同时收集多个网络的数据分组既会产生巨大的开销也会产生隐私问题。因此, 作者利用 FL 让多个 IoT 网络的网关进行共同训练。当训练完成后, 各 IoT 网关通过比对所建立模型预测的数据分组序列和实际数据分组序列来判断网络中是否存在异常设备。

8 总结及未来研究展望

本文对 AI 中三类代表性的机器学习方法及其在无线通信领域的应用进行了介绍和分析。根据这三类方法的特点可知, 它们所适用的无线通信问题略有差别。其中, DL 主要适用于无线传输、频谱感知、调制识别等问题, 而 DRL 主要适用于信道接入、资源配置、网络接入等问题。与 DL 和 DRL 相比, FL 的应用场景则具有较大的弹性。FL 能够通过结合 DL 或 DRL 来解决各类问题, 同时为数据隐私提供保护。下面, 将分别对这三类方法在解决无线通信问题时的优缺点进行总结, 并围绕其局限性讨论可行解决方案, 供未来研究工作参考。

8.1 深度学习

在 DL 中, DNN 的训练和部署是分开的。训练阶段是离线进行的, 通过对训练数据的关联分析来建立最优的映射关系。训练完成后的 DNN 便可在部署。因此, DL 在 DNN 完成训练后能以极低的计算开销从源数据推断相应的目标数据, 适合在低功耗设备部署。

然而, 训练与部署的分离也使 DL 高度依赖训练数据集: 规模越大、越完整的数据集可以让 DL 训练出质量越高的 DNN 模型。但受限于链路速率、无线环境波动以及噪声的影响, 无线通信系统获得的数据集往往数据量比较少, 而且容易受噪声干扰甚至部分缺失。在计算机视觉等领域, 如 PCA Jittering、独立元分析等技术已经被成熟地应用于图像数据的增广和增强中, 将其推广到无线通信数据

的增广和增强应当是下一步需要关注的研究方向。

此外, 为了获得更好的性能, DNN 的结构以及数据预处理都需要根据所考虑具体问题的特点进行设计。与此同时, 合适的 DNN 结构往往需要根据数据特征进行人工调优^[77], 为 DL 的广泛应用增加了困难。虽然文献[46]尝试使用启发式算法对包括 DNN 结构在内的超参数进行优化, 但是该方法在其他问题以及场景下尚未得到验证。元学习 (meta learning) 是解决这一问题的可能途径, 它可以根据数据和任务特征自动选择 DNN 结构和其他 DL 超参数, 代表算法有 AutoML^[78]、Reptile^[79]等。目前, 元学习在无线通信领域仍有很大的研究空间。

8.2 深度强化学习

与 DL 相比, DRL 是一类在线训练和部署的机器学习方法, 不需要进行预先的离线训练。这使 DRL 不再需要像 DL 那样耗费大量资源来获得训练数据集。然而, 这也意味着 DRL 从开始运行到性能收敛需要经过大量的试错, 因而 DRL 在收敛过程中的性能无法得到保证。在 AlphaGo^[80]中, 围棋的棋谱被用于 DRL 的初期训练, 加快了训练的速度。为此, 应当研究如何结合无线通信中的专家知识 (如成熟的优化算法、模型等) 来加快 DRL 训练速度以及提升性能。

另一方面, 训练神经网络所需的计算开销大而移动设备通常是低功耗且电池寿命有限的, 难以承担 DRL 中对 DNN 的训练任务。利用 MEC 为移动设备进行低时延的训练任务卸载是可行的方案, 而其中存在的能量、时延、费用等因素的折中问题需要进一步研究。

最后, 随着具有 AI 算力的移动芯片的普及, 未来的无线通信系统必然将由众多智能设备构建而成, 因而分布式的多智能体决策方案将是大势所趋。然而, 尚未有能够解决 DRL 多智能体非平稳性问题的相关理论, 目前的研究仅停留在基于经验的启发式解决方案, 如文献[63,69]。实际上, 多智能体下的 DRL 收敛需要所有智能体能够学习到一种可以达到均衡点并不再改变的策略。均衡点是否存在以及均衡点的性能决定了 DRL 的收敛性能。作为描述和分析多个个体交互过程的理论, 博弈论以及随机博弈论^[81]可以为解决 DRL 的多智能体非平稳性问题提供理论框架, 是未来研究的重要课题。

8.3 联邦学习

FL 通过仅传输 DNN 权值来避免原始数据的泄露, 实现了对原始数据的保护。因此, FL 适用于所有需要多智能体协作训练, 但对隐私有较高要求的问题。作为新兴的 AI 技术, FL 在无线通信中的应用目前还处于比较早期的研究阶段, 也存在一些局限性。

1) 当 DNN 规模较大时, 需要传输的权值更新值数量庞大, 导致智能体与服务器之间的通信开销巨大, 对控制链路容量有限的无线通信系统造成巨大的负担。为了降低 FL 的通信开销, 智能体可以对权值更新值进行压缩^[82]或量化^[83]。目前, 尚未有采用 FL 的无线通信技术考虑交互 DNN 权值产生的通信开销。因此, 未来的研究工作应当量化并降低采用 FL 带来的通信开销。

2) FL 中各个智能体的数据来源可能是不一样的, 而具有偏见的数据源会使各智能体拥有的数据集有可能是非独立同分布的。例如, 室内基站的用户流量大部分来源于音/视频等娱乐服务, 而室外基站的用户流量则可能大部分来源于语音、导航等业务。文献[8]的仿真结果显示, 当各个智能体的本地数据并非服从相同分布(也称非独立同分布数据)时会对 DNN 的训练结果产生负面影响, 降低 DNN 的准确度。由于非独立同分布数据集在无线通信领域中是非常常见的, 未来的研究工作应当考虑这一问题并采取相应的手段来解决。

3) FL 将训练任务分配到各个智能体中, 而智能体和服务器是相对独立的, 使 FL 具有广阔的攻击面、抗攻击能力比较差。当某智能体或其与服务器之间的链路存在恶意攻击时, 智能体上传的权值更新值就有可能被截获或者被替换成恶意的更新值。对于前者, 攻击者可以基于截获的权值更新值来反推智能体的数据或数据的某些性质。对于后者, 攻击者可以利用恶意的更新值降低训练效果。目前, 存在同态加密^[84]、差分隐私^[85]、稳健合并算法^[86]等方法来加强 FL 的安全性。在无线通信中, 由于信息传输具有一定的开放性, 应用 FL 时应当采用强有力的安全保障手段, 而现有的方案在这一方面是欠缺的。

9 结束语

在当下信息爆炸和万物互联时代, 无线通信需求呈现指数增长的态势。为了满足用户激增的需

求, 无线通信系统在规模剧增的同时也加入了各类复杂技术, 增加了其设计、运行、管理和维护的难度, 使传统的无线通信技术捉襟见肘。在这一背景下, 以深度学习、深度强化学习和联邦学习为代表的人工智能技术为解决无线通信的瓶颈问题提供了可行的手段。本文对基于这三类方法的智能无线通信技术研究进行了综述。目前的研究表明人工智能在无线通信领域拥有巨大的发展潜力, 但也存在一些问题。针对存在的局限性, 本文为未来智能无线通信技术的研究指出了一些方向。

附录 强化学习概述

一个人在成长过程中会习得各种各样的技能, 且大多数是从生活中的积累经验中学习得到的。RL 正是一种模仿智慧体累积和学习经验的机器学习方法^[16]。

在 RL 中, 进行学习的智慧体被称为智能体或代理(agent)。一般而言, 智能体需要先观察其周遭的环境, 然后做出相应的动作来完成规定的任务。根据环境选择动作的过程称为决策。在智能体采取行动后, 其周围的环境可能因此发生变化, 并且智能体也能从环境中观察到是否完成、接近或者偏离其需要完成的目标。环境因智能体所采取的动作而发生变化的概率称为转移概率, 而衡量动作对完成任务作用的指标称为奖赏。智能体与环境之间进行的复杂交互过程在数学上通常可以被抽象为一个马尔可夫决策过程(MDP, Markov decision process)。MDP 由动作、状态、策略、转移概率和奖赏 5 个核心要素构成。

动作(action)。智能体决策的对象称为动作, 记为 a 。智能体被允许采取的所有动作构成了 MDP 的动作空间, 记为 A 。

状态(state)。智能体对其所处环境的观察和描述称为状态, 记为 s , 而所有可能出现的状态构成了状态空间, 记为 S 。从智能体的角度, 状态的改变反映了环境发生的变化。

策略。智能体根据状态选择动作的决策规则称为策略, 记为 π 。其中, $\pi(a|s)$ 表示智能体在状态 s 下选择动作 a 的概率。

转移概率。智能体采取动作 a 后使状态从 s 转移到 s' 的概率称为转移概率, 记为 $p_a(s, s')$ 。

奖赏(reward)。智能体在采取动作后会从环境中观察到该动作对完成规定任务的作用, 而衡量的指标被称为奖赏值。当智能体在状态 s 下采取动作 a 使状态变为 s' 时, 获得的奖赏值记为 $r_a(s, s')$ 。

基于以上的基本要素, MDP 问题被定义为如何找到一

个最优策略 π^* 来最大化时间尺度 T 内的期望累积奖赏

$$R = \mathbb{E} \left[\sum_{t=0}^T \gamma^t r_{a_t}(s_t, s_{t+1}) \right] \quad (4)$$

其中, $\gamma \in [0, 1]$ 被称为折扣因子。折扣因子决定了未来奖赏对智能体的重要程度。例如, 当 $\gamma=0$ 时, 智能体完全不考虑当前决策对未来奖赏的影响; 当 $\gamma=1$ 时, 智能体的目标则是最大化未来的所有奖赏值。一般来说, γ 越大, 则可以让智能体以更长远的眼光进行决策, 但是对未来奖赏的考虑同时也会增加决策难度。此外, 根据时间尺度 T 是有限值或无限值, MDP 又可以分为有限时间尺度的 MDP 和无限时间尺度的 MDP。

为了求解 MDP, RL 让智能体不断地与环境进行交互来获取经验, 并从经验中直接学习得到最优策略。其中, 借助值函数进行学习的方法称为基于值 RL 方法。以最广泛使用的 Q 学习 (Q-learning) 算法^[16]为例, 它为所有状态 $s \in S$ 和所有动作 $a \in A$ 都构建了值函数 $Q(s, a)$ 。 $Q(s, a)$ 的物理意义为智能体在状态 s 下采取动作 a 能获得的期望累积奖赏。由文献^[16]可知, 当智能体根据最优策略 π^* 进行决策时, 得到的值函数 $Q(s, a)$ 对于所有的 $s \in S$ 和 $a \in A$ 能达到最大值, 记为 $Q^*(s, a)$ 。相应地, 若值函数达到最大值时, 最优策略为

$$\pi^*(s) = \arg \max_a Q^*(s, a) \quad (5)$$

为了得到 $Q^*(s, a)$ 和相应的最优策略, Q 学习根据值函数的实际采样值和其预计值之间的时间差分 (temporal difference) 来迭代地更新值函数, 直至值函数收敛。具体的更新式为

$$Q(s, a) = Q(s, a) + \alpha \left[r_a(s, s') + \gamma \max_a Q(s', a) - Q(s, a) \right] \quad (6)$$

其中, α 是控制值函数更新速度的步长因子。在 Q 学习算法中, 动作的选择是根据 ϵ 贪婪规则进行的, 智能体以 $1-\epsilon$ 的概率选择当前状态下的值函数最大的动作, 即 $\arg \max_a Q(s, a)$, 否则在动作空间 A 中随机选择一个动作。前者的作用是让智能体充分利用已有的知识进行最佳的决策, 后者的作用则是让智能体探索未知但可能更好的动作。在实际中, ϵ 的选择需要兼顾两方面的影响并取得折中。

此外, 还有一类基于策略的 RL 方法。它们不需要借助值函数, 而是直接利用函数对策略进行拟合, 并通过调整函数的参数来优化策略。若用一个参数为 θ 的函数拟合策略 π , 则智能体在状态 s 下选择动作 a 的概率可以表示为 $\pi(a|s, \theta)$ 。此时, 对策略 π 的优化就转化为对参数 θ 的优化。为了优化 θ , 首先需要定义一个标量 $J(\theta)$ 来度量 θ 对应策略的性能, 然后 θ 应当以 $J(\theta)$ 梯度上升的方向更新, 即

$$\theta' = \theta + \alpha \widehat{\nabla J(\theta)} \quad (7)$$

其中, $\widehat{\nabla J(\theta)}$ 为 $J(\theta)$ 梯度的随机抽样值。因此, 这一类使用函数拟合策略并对函数的参数进行梯度更新的 RL 方法也被称为策略梯度法。

REINFORCE 算法^[16]是一种被广泛应用的策略梯度法。该算法使用 $\widehat{\nabla J(\theta)} = G_t \nabla \ln \pi(a_t | s_t, \theta)$ 作为 $J(\theta)$ 的随机抽样值。其中, G_t 为从 $t+1$ 时刻直至最终时刻 T 的累积奖赏记, 即

$$G_t = \sum_{k=t+1}^T \gamma^{k-t-1} r_{a_k}(s_k, s_{k+1}) \quad (8)$$

值得注意的是, 基于策略的 RL 方法, 包括 REINFORCE 算法, 其参数和策略的更新是以回合为单位进行的。也就是说, 智能体在同一个回合中采用相同的参数和策略来生成动作、状态和奖赏。

上述 2 种方法都有明显的缺点。对于基于值的方法, 值函数需要通过表格来存储。显然, 当状态空间或动作空间维度很大甚至为连续空间时, 将会产生维度爆炸问题而导致基于值的 RL 方法无法使用。对于基于策略的方法, 以回合为单位的更新机制导致数据利用率比较低, 也使智能体在不同回合下采用的策略差异大、产生的数据方差大。这两方面原因都导致其学习效率比较低。此外, 无线通信中的许多问题是无限时间尺度的, 这使基于策略的 RL 方法无法适用。

参考文献:

- [1] CISCO. Cisco visual networking index: global mobile data traffic forecast update, 2017–2022[R]. Cisco Public Information, 2017.
- [2] LECUN Y, BENGIO Y, HINTON G. Deep learning[J]. Nature, 2015, 521(7553): 436-444.
- [3] MNIH V, KAVUKCUOGLU K, SILVER D, et al. Human-level control through deep reinforcement learning[J]. Nature, 2015, 518(7540): 529-533.
- [4] YANG Q, LIU Y, CHEN T, et al. Federated machine learning: concept and applications[J]. ACM Transactions on Intelligent Systems and Technology (TIST), 2019, 10(2): 1-19.
- [5] LECUN Y, BOTTOU L, BENGIO Y, et al. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11): 2278-2324.
- [6] YOUNG T, HAZARIKA D, PORIA S, et al. Recent trends in deep learning based natural language processing[J]. IEEE Computational Intelligence Magazine, 2018, 13(3): 55-75.
- [7] AMARJYOTI S. Deep reinforcement learning for robotic manipulation-the state of the art[J]. arXiv Preprint, arXiv:1701.08878, 2017.
- [8] MCMAHAN H B, MOORE E, RAMAGE D, et al. Communication-efficient learning of deep networks from decentralized data[J]. arXiv Preprint, arXiv:1602.05629, 2016.
- [9] BISHOP C M. Pattern recognition and machine learning[M]. Berlin: Springer, 2006.

- [10] CYBENKO G. Approximation by superpositions of a sigmoidal function[J]. *Mathematics of Control, Signals and Systems*, 1989, 2(4): 303-314.
- [11] DAYAN P, ABBOTT L F. *Theoretical neuroscience: computational and mathematical modeling of neural systems*[M]. Cambridge: MIT Press, 2001.
- [12] MIKOLOV T, KARAFIÁT M, BURGET L, et al. Recurrent neural network based language model[C]//Eleventh Annual Conference of the International Speech Communication Association. Trier: DBLP, 2010: 1045-1048.
- [13] HOCHREITER S, SCHMIDHUBER J. Long short-term memory[J]. *Neural computation*, 1997, 9(8): 1735-1780.
- [14] LIU W, WANG Z, LIU X, et al. A survey of deep neural network architectures and their applications[J]. *Neurocomputing*, 2017, 234: 11-26.
- [15] LUONG N C, HOANG D T, GONG S, et al. Applications of deep reinforcement learning in communications and networking: a survey[J]. *IEEE Communications Surveys & Tutorials*, 2019, 21(4): 3133-3174.
- [16] SUTTON R S, BARTO A G. *Introduction to reinforcement learning*[M]. Cambridge: MIT press, 1998.
- [17] VAN H H, GUEZ A, SILVER D. Deep reinforcement learning with double q-learning[C]//Thirtieth AAAI Conference on Artificial Intelligence. Palo Alto: AAAI Press, 2016: 2094-2100.
- [18] WANG Z, SCHAUL T, HESSEL M, et al. Dueling network architectures for deep reinforcement learning[J]. *arXiv Preprint, arXiv:1511.06581*, 2015.
- [19] LILLICRAP T P, HUNT J J, PRITZEL A, et al. Continuous control with deep reinforcement learning[J]. *arXiv Preprint, arXiv:1509.02971*, 2015.
- [20] SCHULMAN J, WOLSKI F, DHARIWAL P, et al. Proximal policy optimization algorithms[J]. *arXiv Preprint, arXiv:1707.06347*, 2017.
- [21] MNH V, BADIA A P, MIRZA M, et al. Asynchronous methods for deep reinforcement learning[C]//International Conference on Machine Learning. New York: ACM Press, 2016: 1928-1937.
- [22] NEUMANN D, WIESE T, UTSCHICK W. Learning the MMSE channel estimator[J]. *IEEE Transactions on Signal Processing*, 2018, 66(11): 2905-2917.
- [23] BALEVI E, DOSHI A, ANDREWS J G. Massive MIMO channel estimation with an untrained deep neural network[J]. *IEEE Transactions on Wireless Communications*, 2020, 19(3): 2079-2090.
- [24] O'SHEA T, KARRA K, CLANCY T C. Learning approximate neural estimators for wireless channel state information[C]//2017 IEEE 27th International Workshop on Machine Learning for Signal Processing (MLSP). Piscataway: IEEE Press, 2017: 1-7.
- [25] CHUN C J, KANG J M, KIM I M. Deep learning-based joint pilot design and channel estimation for multiuser MIMO channels[J]. *IEEE Communications Letters*, 2019, 23(11): 1999-2003.
- [26] QING C, CAI B, YANG Q, et al. ELM-based superimposed CSI feedback for FDD Massive MIMO system[J]. *IEEE Access*, 2020, 8: 53408-53418.
- [27] ARNOLD M, DÖRNER S, CAMMERER S, et al. Enabling FDD massive MIMO through deep learning-based channel prediction[J]. *arXiv Preprint, arXiv:1901.03664*, 2019.
- [28] YANG Y, GAO F, LI G Y, et al. Deep learning-based downlink channel prediction for FDD massive MIMO system[J]. *IEEE Communications Letters*, 2019, 23(11): 1994-1998.
- [29] LIANG Y C, CHIN F P S. Downlink channel covariance matrix (DCCM) estimation and its applications in wireless DS-CDMA systems[J]. *IEEE Journal on Selected Areas in Communications*, 2001, 19(2): 222-232.
- [30] WEN C K, SHIH W T, JIN S. Deep learning for massive MIMO CSI feedback[J]. *IEEE Wireless Communications Letters*, 2018, 7(5): 748-751.
- [31] WANG T, WEN C K, JIN S, et al. Deep learning-based CSI feedback approach for time-varying massive MIMO channels[J]. *IEEE Wireless Communications Letters*, 2018, 8(2): 416-419.
- [32] HE H, WEN C K, JIN S, et al. A model-driven deep learning network for MIMO detection[C]//2018 IEEE Global Conference on Signal and Information Processing (GlobalSIP). Piscataway: IEEE Press, 2018: 584-588.
- [33] TAN X, ZHONG Z, ZHANG Z, et al. Low-Complexity message passing MIMO detection algorithm with deep neural network[C]//2018 IEEE Global Conference on Signal and Information Processing (GlobalSIP). Piscataway: IEEE Press, 2018: 559-563.
- [34] WEI Y, ZHAO M M, ZHAO M, et al. Learned conjugate gradient descent network for massive MIMO detection[J]. *arXiv Preprint, arXiv:1906.03814*, 2019.
- [35] SAMUEL N, DISKIN T, WIESEL A. Learning to detect[J]. *IEEE Transactions on Signal Processing*, 2019, 67(10): 2554-2564.
- [36] YE H, LI G Y, JUANG B H. Power of deep learning for channel estimation and signal detection in OFDM systems[J]. *IEEE Wireless Communications Letters*, 2017, 7(1): 114-117.
- [37] O'SHEA T, HOYDIS J. An introduction to deep learning for the physical layer[J]. *IEEE Transactions on Cognitive Communications and Networking*, 2017, 3(4): 563-575.
- [38] XUE S, MA Y, YI N, et al. Unsupervised deep learning for MU-SIMO joint transmitter and noncoherent receiver design[J]. *IEEE Wireless Communications Letters*, 2018, 8(1): 177-180.
- [39] BALEVI E, ANDREWS J G. One-bit OFDM receivers via deep learning[J]. *IEEE Transactions on Communications*, 2019, 67(6): 4326-4336.
- [40] LIANG Y C. *Dynamic spectrum management: from cognitive radio to blockchain and artificial intelligence*[M]. Berlin: Springer, 2020.
- [41] LIU C, WANG J, LIU X, et al. Deep CM-CNN for spectrum sensing in cognitive radio[J]. *IEEE Journal on Selected Areas in Communications*, 2019, 37(10): 2306-2321.
- [42] XIE J, LIU C, LIANG Y C, et al. Activity pattern aware spectrum sensing: a cnn-based deep learning approach[J]. *IEEE Communications Letters*, 2019, 23(6): 1025-1028.
- [43] LEES W M, WUNDERLICH A, JEAUVONS P J, et al. Deep learning classification of 3.5-GHz band spectrograms with applications to spectrum sensing[J]. *IEEE Transactions on Cognitive Communications and Networking*, 2019, 5(2): 224-236.
- [44] LEE W, KIM M, CHO D H. Deep cooperative sensing: cooperative spectrum sensing based on convolutional neural networks[J]. *IEEE Transactions on Vehicular Technology*, 2019, 68(3): 3005-3009.

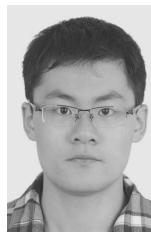
- [45] YU L, WANG Q, GUO Y, et al. Spectrum availability prediction in cognitive aerospace communications: a deep learning perspective[C]//2017 Cognitive Communications for Aerospace Applications Workshop (CCAA). Piscataway: IEEE Press, 2017: 1-4.
- [46] YU L, CHEN J, DING G, et al. Spectrum prediction based on Taguchi method in deep learning with long short-term memory[J]. IEEE Access, 2018, 6: 45923-45933.
- [47] YU Y, WANG T, LIEW S C. Deep-reinforcement learning multiple access for heterogeneous wireless networks[J]. IEEE Journal on Selected Areas in Communications, 2019, 37(6): 1277-1290.
- [48] TAN J, ZHANG L, LIANG Y C, et al. Intelligent sharing for LTE and Wi-Fi systems in unlicensed bands: a deep reinforcement learning approach[J]. IEEE Transactions on Communications, 2020, 68(5): 2793-2808.
- [49] WANG S, LIU H, GOMES P H, et al. Deep reinforcement learning for dynamic multichannel access in wireless networks[J]. IEEE Transactions on Cognitive Communications and Networking, 2018, 4(2): 257-265.
- [50] NAPARSTEK O, COHEN K. Deep multi-user reinforcement learning for distributed dynamic spectrum access[J]. IEEE Transactions on Wireless Communications, 2018, 18(1): 310-323.
- [51] LIU X, XU Y, JIA L, et al. Anti-jamming communications using spectrum waterfall: a deep reinforcement learning approach[J]. IEEE Communications Letters, 2018, 22(5): 998-1001.
- [52] XIE W, HU S, YU C, et al. Deep learning in digital modulation recognition using high order cumulants[J]. IEEE Access, 2019, 7: 63760-63766.
- [53] LUO B, PENG Q, COSMAN P C, et al. Robustness of deep modulation recognition under AWGN and Rician fading[C]//2018 52nd Asilomar Conference on Signals, Systems, and Computers. Piscataway: IEEE Press, 2018: 447-450.
- [54] LIN Y, TU Y, DOU Z, et al. The application of deep learning in communication signal modulation recognition[C]//2017 IEEE/CIC International Conference on Communications in China (ICCC). Piscataway: IEEE Press, 2017: 1-5.
- [55] WANG Y, LIU M, YANG J, et al. Data-driven deep learning for automatic modulation recognition in cognitive radios[J]. IEEE Transactions on Vehicular Technology, 2019, 68(4): 4074-4077.
- [56] KARRA K, KUZDEBA S, PETERSEN J. Modulation recognition using hierarchical deep neural networks[C]//2017 IEEE International Symposium on Dynamic Spectrum Access Networks (DySPAN). Piscataway: IEEE Press, 2017: 1-3.
- [57] ZENG Y, ZHANG M, HAN F, et al. Spectrum analysis and convolutional neural network for automatic modulation recognition[J]. IEEE Wireless Communications Letters, 2019, 8(3): 929-932.
- [58] HONG D, ZHANG Z, XU X. Automatic modulation classification using recurrent neural networks[C]//2017 3rd IEEE International Conference on Computer and Communications (ICCC). Piscataway: IEEE Press, 2017: 695-700.
- [59] ZHANG D, DING W, ZHANG B, et al. Automatic modulation classification based on deep learning for unmanned aerial vehicles[J]. Sensors, 2018, 18(3): 924.
- [60] ZHANG M, ZENG Y, HAN Z, et al. Automatic modulation recognition using deep learning architectures[C]//2018 IEEE 19th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC). Piscataway: IEEE Press, 2018: 1-5.
- [61] HU S, PEI Y, LIANG P P, et al. Deep neural network for robust modulation classification under uncertain noise conditions[J]. IEEE Transactions on Vehicular Technology, 2020, 69(1): 564-577.
- [62] LI X, FANG J, CHENG W, et al. Intelligent power control for spectrum sharing in cognitive radios: a deep reinforcement learning approach[J]. IEEE Access, 2018, 6: 25463-25473.
- [63] NASIR Y S, GUO D. Multi-agent deep reinforcement learning for dynamic power allocation in wireless networks[J]. IEEE Journal on Selected Areas in Communications, 2019, 37(10): 2239-2250.
- [64] TAN J, ZHANG L, LIANG Y C. Deep reinforcement learning for channel selection and power control in D2D networks[C]//2019 IEEE Global Communications Conference (GLOBECOM). Piscataway: IEEE Press, 2019: 1-6.
- [65] YE H, LI G Y, JUANG B H F. Deep reinforcement learning based resource allocation for V2V communications[J]. IEEE Transactions on Vehicular Technology, 2019, 68(4): 3163-3173.
- [66] QIAN Y, HU L, CHEN J, et al. Privacy-aware service placement for mobile edge computing via federated learning[J]. Information Sciences, 2019, 505: 562-570.
- [67] REN J, WANG H, HOU T, et al. Federated learning-based computation offloading optimization in edge computing-supported Internet of things[J]. IEEE Access, 2019, 7: 69194-69201.
- [68] XU Y, XU W, WANG Z, et al. Load balancing for ultradense networks: a deep reinforcement learning-based approach[J]. IEEE Internet of Things Journal, 2019, 6(6): 9399-9412.
- [69] ZHAO N, LIANG Y C, NIYATO D, et al. Deep reinforcement learning for user association and resource allocation in heterogeneous cellular networks[J]. IEEE Transactions on Wireless Communications, 2019, 18(11): 5141-5152.
- [70] ZHANG C, LIU Z, GU B, et al. A deep reinforcement learning based approach for cost-and energy-aware multiframe mobile data offloading[J]. IEICE Transactions on Communications, 2018, 7: 1625-1634.
- [71] FENG K, WANG Q, LI X, et al. Deep reinforcement learning based intelligent reflecting surface optimization for MISO communication systems[J]. IEEE Wireless Communications Letters, 2020, 9(5): 745-749.
- [72] LIU J, KRISHNAMACHARI B, ZHOU S, et al. DeepNap: data-driven base station sleeping operations through deep reinforcement learning[J]. IEEE Internet of Things Journal, 2018, 5(6): 4273-4282.
- [73] JUNHONG Y E, ZHANG Y J. DRAG: deep reinforcement learning based base station activation in heterogeneous networks[J]. IEEE Transactions on Mobile Computing, 2019, doi: 10.1109/TMC.2019.2922602.
- [74] WU J, YU P, FENG L, et al. 3D aerial base station position planning based on deep Q-network for capacity enhancement[C]//2019 IFIP/IEEE Symposium on Integrated Network and Service Management (IM). Piscataway: IEEE Press, 2019: 482-487.
- [75] MISMAR F B, EVANS B L. Deep Q-learning for self-organizing networks fault management and radio performance improvement[C]//2018 52nd Asilomar Conference on Signals, Systems, and

- Computers. Piscataway: IEEE Press, 2018: 1457-1461.
- [76] NGUYEN T D, MARCHAL S, MIETTINEN M, et al. D²IoT: a federated self-learning anomaly detection system for IoT[C]//2019 IEEE 39th International Conference on Distributed Computing Systems (ICDCS). Piscataway: IEEE Press, 2019: 756-767.
- [77] QING C, YU W, CAI B, et al. ELM-based frame synchronization in burst-mode communication systems with nonlinear distortion[J]. IEEE Wireless Communications Letters, 2020, 9(6): 915-919.
- [78] HUTTER F, KOTTHOFF L, VANSCHOREN J. Automated machine learning[M]. Berlin: Springer, 2019.
- [79] NICHOL A, SCHULMAN J. Reptile: a scalable meta-learning algorithm[J]. arXiv Preprint, arXiv:1803.02999, 2018.
- [80] SILVER D, SCHRITTWIESER J, SIMONYAN K, et al. Mastering the game of go without human knowledge[J]. Nature, 2017, 550(7676): 354-359.
- [81] 王元卓, 林闯, 程学旗, 等. 基于随机博弈模型的网络攻防量化分析方法[J]. 计算机学报, 2010, 33(9): 1748-1762.
WANG Y Z, LIN C, CHENG X Q, et al. Analysis for network attack-defense based on stochastic game model[J]. Chinese Journal of Computers, 2010, 33(9): 1748-1762.
- [82] LIN Y, HAN S, MAO H, et al. Deep gradient compression: reducing the communication bandwidth for distributed training[J]. arXiv Preprint, arXiv:1712.01887, 2017.
- [83] KONEČNÝ J, MCMAHAN H B, YU F X, et al. Federated learning: strategies for improving communication efficiency[J]. arXiv Preprint, arXiv:1610.05492, 2016.
- [84] HARDY S, HENECKA W, IVEY-LAW H, et al. Private federated learning on vertically partitioned data via entity resolution and additively homomorphic encryption[J]. arXiv Preprint, arXiv:1711.10677, 2017.
- [85] GEYER R C, KLEIN T, NABI M. Differentially private federated learning: a client level perspective[J]. arXiv Preprint, arXiv:1712.07557, 2017.
- [86] LI L, XU W, CHEN T, et al. RSA: Byzantine-robust stochastic aggregation methods for distributed learning from heterogeneous datasets[C]//Proceedings of the AAAI Conference on Artificial Intelligence. Palo Alto: AAAI Press, 2019: 1544-1551.

[作者简介]



梁应敞 (1968-), 男, 江西赣州人, 博士, 电子科技大学教授、博士生导师, 主要研究方向为智能无线通信、认知无线电、智慧物联网等。



谭俊杰 (1994-), 男, 广东肇庆人, 电子科技大学博士生, 主要研究方向为动态频谱共享、认知无线电、智能无线通信。



Dusit Niyato (1978-), 男, 泰国北柳府人, 博士, 南洋理工大学教授、博士生导师, 主要研究方向为无线能量采集、物联网和传感器网络等。