

基于深度强化学习的新型配电系统优化运行研究综述

王仁浚¹,高红均¹,罗龙波²,陈明辉²,徐杰¹,刘俊勇¹

(1. 四川大学 电气工程学院,四川 成都 610065;2. 广东电网有限责任公司广州供电局,广东 广州 510620)

摘要:高比例新能源接入背景下,新型配电系统运行优化问题不断凸显。考虑新型配电系统多元化分布式资源广泛接入、多类型主体互动协同与系统规模不断扩大的发展特征,综述由此带来的运行高度不确定性、复杂交互关系难以建模以及大规模系统求解困难等关键挑战。针对运行优化问题在集中/分散式决策、安全性约束及多主体互动性等方面的差异化特性,系统梳理并归纳了基于深度强化学习的关键运行优化技术,比较了不同算法的性能优劣及适用场景。总结得出了深度强化学习在新型配电系统优化运行中面临的主要技术瓶颈,展望了其未来在多时间尺度协同、模型泛化与智能交互等方向的发展前景。

关键词:新型配电系统;多元化分布式资源;多类型主体;规模化发展;深度强化学习

中图分类号:TM715;TP18

文献标志码:A

DOI:10.16081/j.epae.202505008

0 引言

为实现“双碳”目标,新型电力系统成为能源绿色低碳转型的关键支撑^[1]。2023年7月,中央全面深化改革委员会第二次会议审议通过《关于深化电力体制改革加快构建新型电力系统的指导意见》。该指导意见明确了科学合理规划电力系统建设,深化电力体制改革,建立清洁低碳、安全充裕、经济高效、供需协同、灵活智能的新型电力系统战略方向,并提到我国能源电力领域已经构建了强大国际竞争力的能源电力产业链,电力科技整体水平实现从跟跑向并行、领跑的战略性转变,但与建设高水平新型电力系统的要求仍存在一定差距。

与集中式新能源接入的输电系统不同,随着新能源比例逐步上升,新型配电系统将展现出多元化分布式资源广泛接入、多类型主体积极参与、规模化发展等显著特征。这些特征为能源的分布和利用带来了前所未有的机遇,也带来了一系列严峻的挑战:由于分布式资源的不确定性,如太阳能和风能的出力依赖于天气条件,使得配电系统的运行变得更加难以预测和管理^[2];系统的互动特性使得新型配电系统中,多种主体相互交织,形成复杂的相互作用网络,系统的稳定性和协调性成为挑战^[3];规模化系统使得优化和管理任务变得更加复杂和耗时,传统的集中式优化架构已经不适用。

为了更好地理解和解决这些挑战,本文基于深度强化学习的新型配电系统优化运行研究,深入分析了系统所面临的关键问题,探讨了如何应对不确

定性、多主体互动特性以及系统规模化带来的挑战。同时,总结了一系列基于深度强化学习的运行优化关键技术。这些技术的引入和应用,将为新型配电系统运行提供新的思路和方法,在提高系统的经济性、高效性和可持续性方面发挥关键作用。

1 新型配电系统典型特征

1.1 多元分布式资源接入

1) 分布式新能源。

传统配电系统是无源的,其演化基于中心化的发电方式,其中变电站将电能从主网传输到用户,潮流仅限于单向流动^[4]。然而,随着分布式新能源接入配电系统,系统的本质特征发生了改变。配电系统从无源转变为有源,而且潮流方向也变为双向流动^[5]。在新型配电系统中,存在着大量小容量的分布式光伏,而非集中式、大容量的风光电场,这使得系统运行变得更加复杂^[6]。

2022年5月,国务院办公厅转发了《关于促进新时代新能源高质量发展的实施方案》,该方案明确了2030年的目标,即风电和太阳能发电总装机容量要达到 1.2×10^9 kW以上,并强调了构建适应新能源占比逐渐提高的新型电力系统的重要性。根据国家能源局每年发表的光伏发电统计信息/建设情况,2016—2024年我国分布式和集中式光伏新增装机容量如附录A图A1所示。可见,2020—2024年我国分布式光伏新增装机容量逐年上升,尤其是2022—2024年容量增幅尤为明显,2024年容量已达到2021年的3倍之多。光伏并网规模从2016年的 10^7 kW增长至2024年的 3.74×10^8 kW,增长超37倍。从分布式光伏比例来看,我国分布式光伏在整体新增并网光伏中的比例从12.3%提升至40%~60%。随着我国形成完善的新能源全产业链,新能源开发成本不断降低^[7],中小型用户等投资的小容量的分

收稿日期:2024-09-14;修回日期:2025-04-07

在线出版日期:2025-05-23

基金项目:国家自然科学基金资助项目(52077146)

Project supported by the National Natural Science Foundation of China(52077146)

布式新能源将在配电系统中扮演日益重要的角色。

2) 电动汽车。

《进一步推进电能替代的指导意见》提出加快推进交通电气化。根据公安部每年发布的全国机动车保有量数据,2016—2024年我国新能源汽车总保有量见附录A图A2。由图可知,我国新能源汽车保有量逐年攀升,2022—2024年我国新能源汽车保有量分别为1310、2041、3140万辆,分别同比增长67.13%、55.8%、53.8%。同时,充、换电基础设施不断加强。根据《中国质量报》发布的充电基础设施统计数据,截至2024年年底,我国充电基础设施新增422.2万台,同比增长24.7%,累计达1281.8万台。其中:公共充电桩增量为85.3万台,同比下降8.1%;随车配建私人充电桩增量为336.8万台,同比增长37.0%。

随着电动汽车动力电池等关键技术的革新和基础设施技术的升级,以及超充标准的制定和双向式充电技术的突破,电动汽车正在迅速发展。未来,新型配电系统将面临双重挑战和机遇。一方面,大规模电动汽车和高功率快充站的普及可能会导致电力需求出现新的尖峰,给配电系统带来巨大的升级改造压力,需要应对电网负荷增加和能源供应的挑战;另一方面,电动汽车具有储能潜力和高度灵活性,可以参与新型配电系统的商用化市场交易和运行互动,这在缓解保供压力和提供灵活性资源方面具有巨大的潜力。在2020年11月,国务院发布了《新能源汽车产业发展规划(2021—2035)》,其中强调了需要鼓励新能源汽车充放电和电力调度需求的统筹规划,并加强新能源汽车与电网能量的互动。为了更好地管理电动汽车的充电需求和能源分配,新型配电系统需要引入更多智能化和自动化技术。这些技术可以帮助实时监测和调整电力分配,从而提高能源效率并降低运营成本。

3) 用户可调节负荷。

2022年1月,国家发展改革委和国家能源局发布《“十四五”现代能源体系规划》,力争到2025年,电力需求侧响应能力达到最大负荷的3%~5%。随后,国家能源局于2022年11月发布函复,强调将提升电力负荷侧灵活调节能力,确保能源电力安全。2022年11月,国家能源局还发布了《可调节负荷并网运行与控制技术规范》系列行业标准,这是全球首次建立可调节负荷参与电网运行的技术标准。这些措施旨在推动电力体系的灵活性和可持续性,以适应不断增长的电力需求,从而提高能源电力的效率与安全性。

目前,我国工业负荷容量高、可控性强,成为可调节负荷的重点开拓方向。在此背景下,一些试点项目已经展开,如2021年3月,国网冀北电力有限公

司成功完成了168 h的工业企业需求响应试点项目,让响应负荷参与华北调峰辅助服务市场;2022年6月,国网绍兴供电公司进行了空调控制试点工作,通过为每个空调安装负荷调控装置,并利用高频采集的负荷数据进行精准控制。由于负荷可调节潜力难以估计,激励非直接控制负荷结算容易产生偏差,目前的负荷调控手段主要采用直接控制方式,相对单一。但是随着信息化系统的发展和可控负荷技术的不断进步,未来新型配电系统将朝着面向多类型用户的可调节负荷多样化调控发展。

1.2 多类型利益主体并存

1) 产消用户。

在配电系统中,随着小规模分布式新能源的快速发展,传统的用能用户逐渐转变为兼具发电能力的产消用户^[8]。根据英国政府发布的《Solar photovoltaics deployment》统计数据,截至2024年11月,英国有140万户家庭拥有太阳能电池板。而根据澳大利亚政府的清洁能源监管机构发布的数据显示,2023年,约21%的家庭安装了小规模光伏发电系统。这一趋势还推动了储能系统的采用,分布式新能源与储能的结合允许用户储存多余的新能源发电来满足其余时段自身的用能,是典型的自消费模式^[9]。根据国家能源局数据,2024年全国户用光伏项目新增装机容量 1.45×10^8 kW,规模占全年新增光伏装机容量规模的25%。根据国家能源局发布的《户用光伏建设运行指南(2022年版)》,我国户用光伏上网分为全额上网和自发自用、余电上网2种模式,光伏系统根据容量的不同接入不同电压等级的公共电网。随着新能源与储能成本的不断下降,未来新型电力系统将面临高占比的产消用户。这种情况下,电力市场的发展为产消用户提供了更多交易方案的可能性,例如参与聚合商、虚拟电厂、点对点交易等。这些交易方案将促进电力市场的灵活性和效率,提高可再生能源的利用率,并推动电力系统向更可持续、智能化的方向发展。

2) 虚拟电厂。

虚拟电厂本质上是一种基于云的分布式能源管理系统,欧盟FENIX项目将其分为技术型与商业型^[10]。技术型虚拟电厂是指公用型企业通过系统集成与调度多种分布式资源,为新型配电系统提供供需平衡能力和灵活性调节能力,从而提高新能源消纳,并保障系统安全稳定运行。商业型虚拟电厂是指第三方商业实体通过整合资源参与市场交易,为缺乏容量优势的产消用户或者其他中小型分布式资源所有者提供更多的选择。

目前,我国的虚拟电厂虽然正处于初始发展阶段,但已受到国家的高度重视。2022年3月,国家发展改革委、工业和信息化部等十部委联合发布《关于

进一步推进电能替代的指导意见》,鼓励电动汽车电网互动(vehicle to grid, V2G)、大数据中心、5G 数据通信基站等利用虚拟电厂参与系统互动。《“十四五”现代能源体系规划》进一步提出,开展多种资源聚合的虚拟电厂示范,包括工业可调节负荷、楼宇空调负荷、大数据中心负荷、用户侧储能、新能源汽车与电网能量互动等。在2019年,华北电网有限公司进行了虚拟电厂试点项目,该虚拟电厂试点为技术型虚拟电厂,主要参与电力辅助服务市场和削峰填谷业务。在2021年,广州市虚拟电厂投入运行,通过与多个电力用户签约进行需求响应控制。在2023年8月,得克萨斯州的2家特斯拉商业虚拟电厂获得了运营资格,可以向得克萨斯电力可靠性委员会管理的电网提供可调度的电力,特斯拉在客户的住宅里安装了Powerwall储能系统,并将剩余电力出售到电力市场上。

随着信息化不断发展,相对于传统的发电厂,虚拟电厂作为数字化、智能化、物联化的集大成者,具有更高的响应能力与更快的响应速度,是未来新型配电系统的主要调控手段之一。同时,随着虚拟电厂容量不断提高,以及市场化的发展,未来商业模式也将更为丰富。

3) 聚合商。

欧盟将独立聚合商定义为“在任何组织化能源市场上出售、购买或拍卖多个用户负荷或发电的市场参与者,其与供应商或任何其他市场参与者没有隶属关系”。与虚拟电厂的主要区别在于,聚合商只可能是一个第三方实体或公司,并没有分散式资源的所有权与控制权,在本质上是一个电力市场的参与者和协调者。根据管理分布式资源的不同,聚合商可分为负荷聚合商、储能聚合商、电动汽车聚合商等。各种研究表明,用户通常对其资源的主动管理兴趣较低^[1],而聚合商能够以最低的交易成本将用户与市场连接起来,并通过优化和管理聚合的负荷获得一定的利益。聚合商可以提供响应服务,包括在高峰时段调控分布式资源,以换取经济激励或其他福利。2022年,《关于进一步推进电能替代的指导意见》鼓励通过培育负荷聚合商,促进用户积极主动参与需求侧响应。随着负荷的不断增长,新型电力系统中小型用户具有巨大的可调节潜力,作为整合这些资源的负荷聚合商将大有可为。

1.3 系统规模化发展

在“双碳”目标下,分布式新能源比例持续增长,用电负荷结构日益复杂,新型配电系统成为电网发展的核心方向,其显著特征是规模化发展。规模化发展是指在网架结构、可控设备、数据通信等方面形成大规模、高密度和高复杂度的系统格局。近年来,我国大力推动配电网建设,并显著提升了配电自动

化覆盖率。新型配电系统规模化发展主要体现在以下3个方面。

1) 网架结构规模化发展。

网架结构规模化是指随着配电系统不断扩展,其结构趋于复杂化和互联化。具体表现为馈线数量急剧增加,互联特性增强,导致广泛的弱耦合关系,形成大量大规模复杂馈线组。如附录A图A3所示,一个大规模复杂馈线组包含诸多馈线,结构复杂且分支众多。以广州配电网为例,截至2023年,馈线共计8862回,馈线总长度为 5.61×10^4 km,其中最大馈线组含742回馈线。可见,新型配电系统网架结构呈现规模化发展,导致网架分析难度与运行控制耗时大幅度上升。

2) 可控设备规模化发展。

可控设备规模化是指随着柔性化、智能化转型,新型配电系统中远程可控设备和智能融合终端呈现高密度部署趋势。国家发展改革委于2024年发布《关于新形势下配电网高质量发展的指导意见》提出到2030年,基本完成配电网柔性化、智能化、数字化转型,实现海量资源聚合互动、多元用户即插即用。根据中国南方电网统计数据,截至2023年年底,累计建设或改造智能台区、智能配电房 3.3×10^4 座,配电自动化有效覆盖率达91.98%。大量的终端设备也意味着更为复杂的控制优化需求,依赖更高效的智能化控制策略。

3) 数据通信规模化发展。

数据通信规模化指配电系统内终端设备的大量接入和数据传输需求的指数级增长。根据Verified Market Reports发布的数据,全球配电自动化终端市场规模在2024—2030年的预测复合年增长率为12.6%,展现出通信规模化发展的全球趋势。随着配电自动化的成熟,智能电表、风电、光伏、储能、电动汽车充电桩等终端设备广泛覆盖,接入通信网执行数据采集、指令传输等任务,为电网精细化控制提供支持。受新能源波动性影响,实时数据分析与决策需求激增,海量终端的同步数据传输加重通信网络负载,对配电系统的稳定性和处理能力提出更高要求。

2 新型配电系统运行挑战

2.1 不确定性带来的挑战

与传统配电系统不同,新型配电系统的主要特征之一是分布式新能源出力的高度不确定性。配电系统的调度中心依赖于日内或小时内的预测数据来进行调度。实际配电系统中,有功功率预测仍然具有挑战性,往往难以达到期望的准确度。这一挑战进一步受到天气条件与可再生能源产出之间密切关联的影响。提高天气预测的准确性仍然是一个全球

性难题。随着分布式可再生能源在新型配电系统中不断增加,传统发电比例不断下降,与可再生能源有功功率相关的高度不确定性对系统的安全运行构成了重大挑战。

对于日前与日内优化而言,针对新型配电系统高比例新能源的不确定性带来的挑战,目前国内外研究了多种不确定性优化技术来处理不确定性,以减少较长时间尺度上不确定性带来的影响,如基于场景的模拟优化^[12]、机会约束规划^[13]、模糊规划^[14]、鲁棒优化^[15]、分布鲁棒优化^[16]等。然而这些方法都有其缺陷,如基于场景的模拟优化方法很难获得不确定性参数的客观随机分布,机会约束规划计算复杂性太高,模糊规划可能会陷入局部最优,鲁棒优化对保守性与经济性难以权衡,而分布鲁棒优化中不确定性概率分布的波动范围在实际应用中难以确定^[17]。此外,计及不确定性的优化方法通常会导致求解难度进一步加大。因此,需要研究新的优化方法进行实时运行调控,从而更好地适应新能源出力不确定性带来的挑战。

2.2 互动特性带来的挑战

新型配电系统以容纳多元化分布式能源为特征,随着电力市场的发展,涌现出了多类型主体,包括产消用户、虚拟电厂和聚合商。这些主体之间的互动特性变得十分复杂,影响着整个电力生态系统的运作和协同。产消用户的双向能源流动,虚拟电厂的整合管理,以及聚合商的市场中介作用,都对电力市场的运行和可持续性产生深远影响。因此,了解和优化这些主体之间的互动关系,是确保新型配电系统高效运行和可持续发展的关键因素。

在市场环境下,多种类型的市场主体以松耦合的方式接入新型配电系统。这些市场主体有不同的利益诉求,涉及到能量交换、数据隐私等多方面的因素。管理这些多元市场主体的参与,确保公平竞争和有效运营,对系统的运行优化提出了更高的要求。这需要强大的市场机制和互动博弈技术来协调各方的行为,以实现系统的高效运行和资源分配。

2.3 系统规模化带来的挑战

随着网架结构不断复杂化,大规模分布式能源资源和智能设备不断增多,包括太阳能板、风力涡轮、储能系统、充电桩、智能电表等。

一方面,由于配电系统运行优化问题数学复杂度过高,传统的优化方法包括基于最优潮流的凸优化方法^[18]、启发式方法^[19]、元启发式方法^[20]等方法难以适应系统规模化带来的挑战。基于最优潮流的凸优化方法需要具有全局最优的理论保证,然而,其需要建立准确的物理模型,且面临变量维数灾的问题。此外,当优化模型的目标函数为非线性目标函数时,该方法难以求解。启发式方法适用于小规模

优化问题,且不受限于目标函数。然而,在规模化发展的新型配电系统中,该方法难以获得稳定的全局最优解。元启发式算法是启发式算法的改进,但其本身依然难以保证计算效率。

另一方面,海量设备的接入导致系统中的数据量急剧增加。处理这些数据需要强大的计算能力和高效的数据管理系统,从而确保系统运行的可靠性和响应速度。而传统配电系统多采用集中控制架构^[21],因此难以适应系统规模化带来的挑战,主要有以下3点。

1)实时性和响应速度:由于所有决策都必须经过中央处理单元,集中式架构通常会存在调控时延。这种时延可能会导致系统对快速变化作出迟缓响应,无法满足实时性要求。

2)通信压力增加:集中式架构要求各种设备将数据传输到中央处理单元,这可能导致网络通信的压力大幅增加。大量数据传输可能导致网络拥塞和通信延迟,从而影响系统的性能。

3)可拓展性:随着系统规模的增大,集中式架构可能会面临可拓展性问题。添加更多的设备和数据源需要大规模的硬件和软件升级,这可能导致高昂的成本。

综上所述,新型配电系统在面对分布式新能源出力不确定性、多类型市场主体复杂互动特性以及系统规模化快速增长的挑战时,传统优化方法和集中式架构已难以满足高效运行的需求。为此,亟需探索具有高度自适应性和实时决策能力的新型技术。深度强化学习作为一种结合感知与决策的智能算法,能够有效应对复杂场景中的动态不确定性和多主体交互,提供一种潜在的解决方案以实现新型配电系统的高效、灵活运行优化。

3 基于深度强化学习的新型配电系统优化运行关键技术

深度强化学习在规模化新型配电系统实时优化运行调控上具有良好的前景。由于深度强化学习模型的输出是从状态到动作的映射关系,因此,通过在不同的净负荷场景下进行训练,深度强化学习模型可以直接根据状态得到近优解,而无需重新优化,且具备一定的泛化能力,能适应变化的场景。目前,深度强化学习已经应用于配电系统运行优化的各个领域,包括网络重构、配电电压优化控制、调度优化运行等。

深度强化学习通过构建马尔可夫决策过程(Markov decision process, MDP)对配电系统运行优化问题进行建模,主要是对奖励函数、动作空间、状态空间的设计,分别对应传统优化方法的目标函数、控制变量、状态变量。无论构建何种MDP,深度强

化学习均需要根据智能体执行动作之后得到的反馈来学习更好的策略。根据反馈过程的不同,目前在配电系统运行优化领域,深度强化学习共有以下3种训练方式:①将智能体直接部署到实际物理配电系统中与系统进行在线交互学习;②构建配电系统的模拟模型,将其视为虚拟环境进行离线交互学习,再将在虚拟环境中训练好的智能体部署到实际系统中;③只通过配电系统历史运行数据集进行离线学习,而不与环境进行交互,然后在线执行。在深度强化学习应用到新型配电系统运行优化领域中,存在着设备动作空间复杂、运行约束安全性难以保证、训练策略鲁棒性低、多主体交互关系复杂、集中式架构适应性差等诸多问题。如图1所示,本文将针对新型配电系统优化运行问题,围绕差异化的MDP以及训练方式,对配电系统运行优化领域中多种深度强化学习算法进行讨论,包括单智能体算法、多智能体算法、安全学习算法、离线学习算法,在此基础上,对计及多主体交互的深度强化学习算法、云边智能优化进行进一步讨论。

3.1 单智能体深度强化算法

对于配电系统集中式优化问题,通常可以采用单智能体深度强化算法进行求解。单智能体可以控制单个设备,或者1个区域内的所有设备。根据配电系统优化问题控制设备不同属性(包括离散设备与连续可调节设备),单智能体算法的动作空间可分为离散动作空间、连续动作空间、混合动作空间,如附录B图B1所示。

经典的离散单智能体算法包括深度Q网络算法(deep Q-network algorithm, DQN)^[22]、双深度Q网络算法(double deep Q-network algorithm, DDQN)^[23]、竞争深度Q网络算法(dueling deep Q-network algorithm, Dueling DQN)^[24]。在离散动作空间中,智能体每次选择1个离散的动作集合中的某个动作。其

中,Q值函数用来估计所有状态-动作对(s, a)的价值,并通过学习来逐步提升Q值估计的准确性。对于每个状态 s ,根据Q值函数来评估并选择最优动作 a 。

对于连续动作空间的算法,通常基于演员-评论家结构,如近端策略优化算法(proximal policy optimization algorithm, PPO)^[25]、软演员-评论家算法(soft actor-critic algorithm, SAC)^[26]、深度确定性策略梯度算法(deep deterministic policy gradient algorithm, DDPG)等。其中,演员通过参数化策略 $\pi(a|s)$ 来输出在每个状态 s 下应采取的动作 a ,通常使用高斯分布来表示策略,输出均值和标准差,然后从中采样动作。评论家通过学习状态的价值 $V(s)$ 或状态-动作组合的价值 $Q(s, a)$,以帮助更新演员策略。

对于混合动作空间的处理,一些文献采用将原问题视为多时间尺度优化问题,其中离散设备调控对应长时间尺度,连续设备调控对应短时间尺度,分别采用不同算法进行优化,如采用传统数学优化模型与深度强化学习算法协同优化^[27]、连续深度强化学习算法与离散深度强化学习算法协同优化^[28]。一些文献将连续深度强化学习算法进行改进以适应混合动作空间,如基于重参数化技巧的DDPG^[29]、伯努利分布与高斯分布分别作为离散与连续神经网络的输出^[30]。然而,规模化新型配电系统中海量可控设备导致动作空间太大,这使得单智能体算法学习最优策略变得困难。

3.2 多智能体深度强化算法

针对涉及多个区域/主体的配电系统分散式优化问题,多智能体深度强化学习具有良好的适应性,如附录B图B2所示。此外,若单智能体深度强化学习算法涉及到过多的控制设备,其收敛性可能会下降,也可以采用多智能体算法将原动作空间分解为若干小动作空间进行求解,从而提升全局性能。

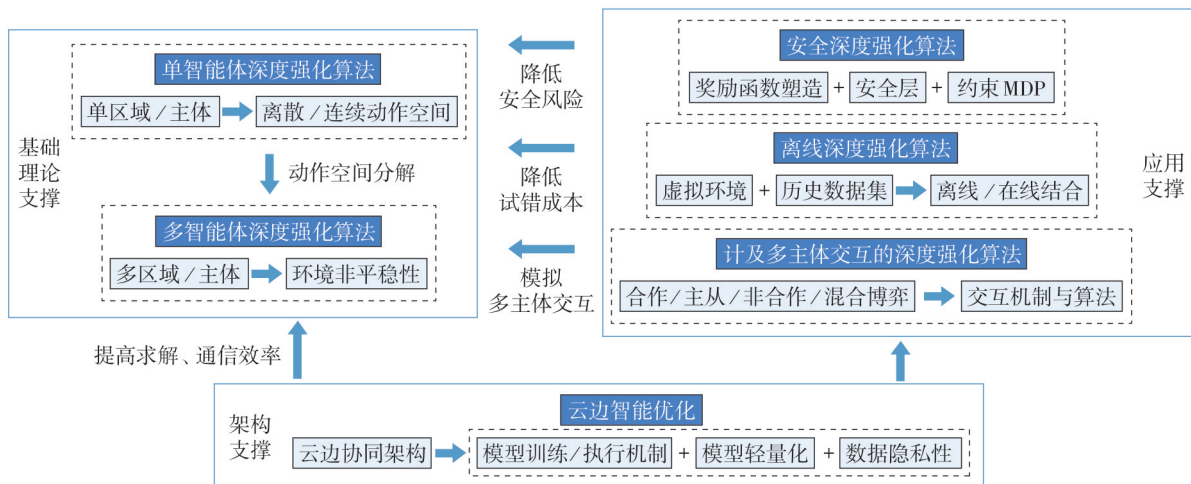


图1 基于深度强化学习的新型配电系统运行关键技术

Fig.1 Key technologies for new distribution system operation based on deep reinforcement learning

在新型配电系统运行优化任务中,每个智能体可以是一个带有控制中心的子区域,也可以是单独的一个可控设备。每个智能体根据局部观测选取动作,所有智能体通过学习全局最优策略协调优化配电系统。因此,多智能体算法通常解决的是1个完全合作的部分可观测优化问题。在多智能体环境中,当单个智能体采取动作时,由于其他智能体的差异化行为,可能会导致截然不同的状态变化,这被称为环境非平稳性。因此,对于完全合作的部分可观测优化问题,多智能体算法的关键点在于如何提升非平稳环境下所有智能体的协同策略,目前主要有值分解方法和策略梯度方法。这2种方法均利用系统全局信息提升智能体训练稳定性。

值分解方法(如QMIX算法^[31-32])源于DQN,其核心思想是将全局目标分解为个体目标,分解遵循“个体全局最大化”原则,即根据全局值函数 Q^{total} 选取的联合动作等价于根据每个智能体动作值函数 Q^{agent} 选取动作的组合,如图2所示。

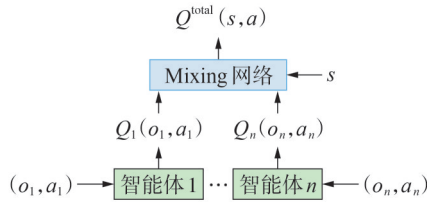


图2 多智能体值分解方法

Fig.2 Multi-agent value decomposition method

QMIX通过一个神经网络(称为Mixing混合网络)来学习将个体值函数 Q^{agent} 映射到全局值函数 Q^{total} 的关系,如式(1)所示。

$$Q^{\text{total}}(s, a) = f(Q_1(o_1, a_1), Q_2(o_2, a_2), \dots, Q_n(o_n, a_n)) \quad (1)$$

式中: $Q_i(o_i, a_i)$ ($i=1, 2, \dots, n$)为第 i 个智能体在局部观测 o_i 选择动作 a_i 的 Q 值。

该算法通过分解全局 Q 值为个体 Q 值,并通过混合网络组合这些 Q 值,来计算全局 Q 值,从而使得每个智能体的决策能够在保证全局最优化的同时,遵循局部最优化。

策略梯度方法(如多智能体DDPG)基于多智能体演员-评论家架构,可适用于连续/离散动作空间,核心思想是每个智能体共享全局奖励函数,根据系统全局信息训练评论家,再通过评论家帮助基于局部信息的演员家改进策略,其训练方式如图3所示。

目前,多智能体DDPG得到了广泛的应用^[33],也有一些文献采用了更先进的方法,如多智能体延迟深度确定性策略梯度算法(twin delayed deep deterministic policy gradient algorithm, TD3)^[34],经验增

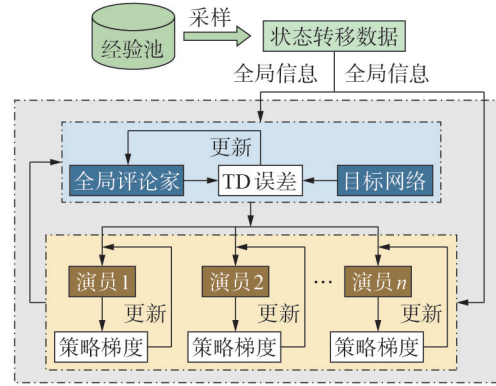


图3 多智能体策略梯度方法

Fig.3 Multi-agent policy gradient method

强多演员注意力评论家算法^[35]、多智能体SAC^[36-37],基于参数共享框架的多智能体深度确定性策略梯度算法^[38],基于Shapley值分配法的多智能体深度确定性策略梯度算法^[39]。然而,目前的多智能体算法通常解决的是少量智能体的任务,随着海量异构设备的接入,智能体数量增长,环境的非平稳性不断增大,多智能体的全局性能可能会进一步下降,目前的方法可能无法适应海量智能体的优化任务。

3.3 安全深度强化算法

单/多智能体深度强化学习在应用到实际新型配电系统运行优化任务上存在安全性挑战。由于配电系统运行优化问题本质上是一个约束优化问题,若智能体直接与物理配电系统进行交互学习,很容易生成违反物理约束的动作,导致系统的安全性受到影响。安全学习基于单/多智能体深度强化学习架构,通过奖励函数塑造方法、策略网络附加安全层方法、基于约束MDP的安全学习算法等方法,以保证智能体安全地探索动作。

奖励函数塑造方法将约束越限视为加权惩罚项添加到原奖励函数中,得到含约束惩罚的奖励函数 R ,如式(2)所示。

$$R = \sum_{t=1}^T \gamma \left(r_t - \sum_{m=1}^M \omega_{t,m} c_{t,m} \right) \quad (2)$$

式中: $\omega_{t,m}$ 为第 t 时间步第 m 个惩罚性因子; $c_{t,m}$ 为第 t 时间步第 m 个惩罚项成本; r_t 为第 t 时间步的奖励; T 为总时间步; M 为约束的总数量; γ 为折扣因子。

奖励函数塑造方法能够简单有效地将多约束变为惩罚项,使得智能体学习到安全的策略。然而,奖励函数 R 中的惩罚权重因子很难设置。如果惩罚因子过小,智能体可能学习到不安全的动作。相反,如果惩罚因子过大,智能体将无法学习到高奖励的策略。更进一步的方式是当满足约束条件,奖励函数对应系统运行优化的目标函数;当不满足约束条件时,奖励对应惩罚项^[40]。尽管该方法可以提升惩罚因子的可调整空间,但仍需要调整惩罚因子以缓解

神经网络估计的不准确性问题。策略网络附加安全层方法可以实现安全动作的采样,如图4所示。图中: $Q_i(s)$ 为第*i*个动作在状态*s*下对应的*Q*值。

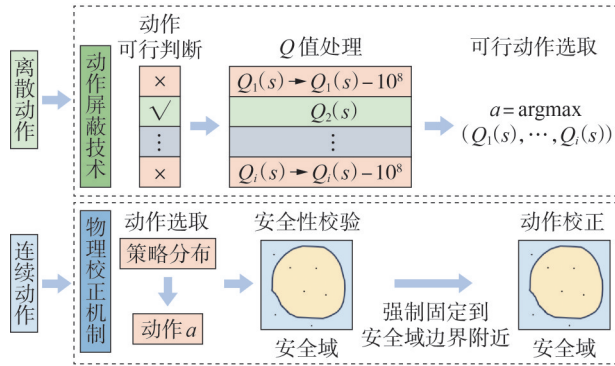


图4 策略网络附加安全层方法

Fig.4 Method for adding security layers to policy networks

对于离散动作空间,可以采用动作Masking掩码技术约束^[41],即将策略网络的输出中不安全行为对应的值减去一个大数,从而使所得的离散动作概率分布中不安全行为的概率趋近于0。对于连续动作空间而言,可以采用物理校正机制来校正不安全行为^[42],即处于安全域附近(如电池容量快达到上、下限)时,强制动作的正负号。安全层通过提前判断动作是否安全的方式来进行动作屏蔽,可以很好地满足显式约束条件(如动作次数约束)。然而,对于隐式约束条件(如系统电压安全约束等),由于在执行阶段无法提前判断系统状态,该方法无效。另外,安全层可能会导致算法训练过程不稳定,增大了算法选取与调参过程的难度。

基于约束MDP的安全学习算法是一种更好的解决方式。约束MDP在标准马尔可夫框架中增加约束条件,当前求解算法主要包括原对偶优化算法(primal-dual optimization algorithm, PDO)^[43-45]和约束策略优化算法(constrained policy optimization algorithm, CPO)^[46-47]。PDO基于拉格朗日松弛技术,将约束MDP转化为min-max问题,max问题对应以神经网络参数为变量的原域问题,min问题对应以拉格朗日乘子为变量的对偶域问题。

$$\mathcal{L}(\pi, \lambda) = R(\pi) - \lambda(C(\pi) - d) \quad (3)$$

$$(\pi^*, \lambda^*) = \arg \min_{\lambda \geq 0} \max_{\pi \in \Pi_\phi} \mathcal{L}(\pi, \lambda) \quad (4)$$

式中: $\mathcal{L}(\pi, \lambda)$ 为拉格朗日函数, π 为智能体策略, λ 为拉格朗日乘子; $C(\pi)$ 为约束成本函数; d 为约束成本上限; Π_ϕ 为神经网络参数为 ϕ 的策略; π^* 与 λ^* 分别为最优策略与拉格朗日乘子。

如式(5)所示,PDO通过对原域与对偶域分别进行梯度上升,实现神经网络的训练与拉格朗日乘子的自适应调节,从而更好地权衡奖励和成本。

$$\phi_{k+1} = \phi_k + \eta \nabla(\phi) (\mathcal{L}(\pi(\phi), \lambda_k)) \Big|_{\phi=\phi_k} \quad (5)$$

式中: λ_k 为第*k*步的拉格朗日乘子; ϕ_{k+1} 为第*k*步时将 λ_k 作为定值,对策略网络参数 ϕ_k 进行更新后,得到的第*k+1*步的神经网络参数; $\nabla(\phi)$ 表示对神经网络参数 ϕ 进行梯度下降; η 为 ϕ_k 对应更新梯步。

$$\lambda_{k+1} = [\lambda_k + \beta(C(\pi) - d)]^+ \quad (6)$$

式中: λ_{k+1} 为第*k*步将策略网络参数 ϕ_k 作为定值,并对拉格朗日乘子 λ_k 进行更新后,得到的第*k+1*步的拉格朗日乘子; β 为 λ 对应更新梯步; $[x]^+$ 表示只取非负值,当 $[x]$ 为负值时, $[x]^+$ 取0。

CPO和PPO之间的主要差异在于,CPO中对偶域参数的更新方式为求解一个优化问题。然而,目前约束MDP的安全学习算法仍然存在诸多限制,如PDO对偶变量的更新具有延迟,导致训练过程中仍然可能违反约束,CPO可以确保在训练过程中始终满足约束条件,但只能应用于信赖域策略优化(trust region policy optimization, TRPO)框架,且对多约束条件的处理效率较低。此外,CPO需要计算二阶导数,实现复杂。在此基础上,内点策略优化方法可以更高效地进行导数计算^[48]。然而,该方法仍会受到可扩展性与计算复杂度的限制。

3.4 离线深度强化算法

对于深度强化学习的第1种训练方式,由于直接与配电系统交互学习训练时间长,试错成本高,且即使采用安全学习算法也可能会导致训练前期安全性不足,一些学者采用离线学习算法进行学习。离线学习算法可以分为基于虚拟环境的离线学习(对应第2种训练方式)与基于历史数据集的离线学习(对应第3种训练方式)。

目前,配电系统虚拟环境可分为潮流计算模型、最优潮流模型、数据驱动模型。绝大多数离线学习研究通过构建潮流模型模拟配电系统的状态变化,智能体与虚拟环境进行交互学习。最优潮流模型通常可以与模仿学习结合加快智能体的训练过程^[48],即将最优解视为专家系统进行模仿学习。相比潮流计算模型,最优潮流模型的求解难度更大,导致训练时间更长。数据驱动模型通过深度学习构建潮流模型的替代模型,可以快速得到系统的状态变化量。总体而言,若构建配电系统虚拟环境进行交互学习,则存在信息偏差,主要有以下2点:配电系统可观性不足且多样化分布式资源动态特性复杂导致无法搭建准确的虚拟系统环境/模型;未来新型配电系统面临结构不断变化且广泛分布式资源自适应接入与退出,导致虚拟环境搭建具有滞后性。

基于历史数据集的离线学习是指仅从固定的历史数据集中学习策略,而不与环境进行交互,因此重点在于如何最大限度地利用静态的离线数据集来训

练智能体。由于历史数据集是有限的,智能体生成的动作未必能够处于历史数据集中,因此会导致估计偏差,现在标准深度强化学习算法学习效果往往非常差,批量学习用来解决这个问题^[49]。目前,批量学习在配电系统运行优化领域应用仍然较少。该方法无需准确的虚拟环境,并可以减少大量的模拟计算开销。然而,该方法需要足够多样化的历史数据才能训练到足够好的策略。

一些学者针对以上问题进行了探索。如将离线学习与在线学习进行结合,提出了新的学习架构与算法;将基于历史数据的离线安全学习与基于实际系统交互的在线安全学习进行结合,既可减少智能体的试错成本,也实现了安全探索下智能体的自趋优学习^[50]。同样,基于虚拟环境的离线学习方法也可以与在线学习算法结合,可以进一步提升策略的性能。一些文献对虚拟环境的构建与对应的学习算法进行改进,如采用对抗性SAC来训练对模型不匹配具有鲁棒性的离线智能体^[51]、构建基于图形学习的功率流代理模型模拟虚拟环境^[52]、构建考虑分布式新能源动态特性的虚拟环境并采用基于清晰度感知最小化的SAC方法^[53],从而减少对精确配电系统参数的依赖,更好地适用变化的系统参数。可见,考虑虚拟环境不准确性的深度强化学习方法架构是未来的重要研究方向之一。

3.5 涉及多主体交互的深度强化学习算法

在涉及多主体交互的新型配电系统运行优化问题中,根据参与博弈的主体之间的关系以及特定问题建模的区别,互动博弈优化可分为合作博弈、主从博弈、非合作博弈(竞争博弈)以及混合博弈。博弈模型通常涉及到多个层面,目前,通常采用深度强化学习算法与数学模型结合的方式进行求解,以解决数学模型难以快速求解的问题。目前涉及多主体交互的深度强化学习算法通常是将其其他主体通过优化物理模型来进行建模,并与该算法对应的决策主体进行交互学习,从而学习到最优调度策略^[45]或者是定价策略^[54]。

在新型配电系统中海量复杂主体的接入下,多主体之间并非简单的合作以及非合作关系,通常由混合博弈建模。混合博弈用来描述配电系统中多主体复杂的博弈关系,通常是合作博弈、主从博弈、非合作博弈之间的两两结合,可以分为主从与非合作博弈的结合、主从博弈与合作博弈的结合、非合作博弈与合作博弈的结合。这些结合模型通常涉及多层次的博弈,其中不同层次的主体在不同方面进行竞争和博弈。混合博弈可以更好地理解实际配电系统中多主体之间的复杂关系,同时也导致建模非常复杂,并且难以兼顾各主体的隐私性。目前,采用混合多智能体算法用以解决混合博弈问题仍处于探索阶

段。基于混合设置的多智能体深度强化学习算法有潜力为各种智能体之间的复杂互动提供更全面的建模方法。

3.6 云边智能优化

传统基于垂直结构的集中式优化方法存在可拓展性、高延迟、带宽消耗显著增加等诸多问题。云边协同优化作为新兴的弱中心化技术,结合新一代无线通信技术(5G技术标准),扩展了传统的智能电网基础设施,可以有效解决传统集中式运行优化架构的巨大通信压力等。云服务器具有高计算能力与储存能力,但其通信带宽较低、通信时延较高。边缘计算可以视为云的延伸,其缩短了数据源与计算资源之间的距离,提供了超低延迟、高带宽和实时访问的特定服务。因此,可以将高计算复杂度、低时延要求的任务交由云中心进行处理,将低计算复杂度、高时延要求的任务交由与控制终端设备更近的边缘服务器进行处理,通过融合云边的优势实现高效、稳定的运行优化控制。

云边协同范式与深度强化学习的结合具有良好的前景。深度强化学习为云边计算提供了技术与方法,云边计算为深度强化学习提供了场景和平台,且深度强化学习可以通过边缘计算扩展其适用性。一般而言,深度强化学习模型可以在云端进行训练,然后部署在边缘执行。这种组合的优势体现在以下3个方面。

1)由于云的规模效应成本,深度强化学习训练成本较低。而对于存储和计算资源有限的边缘节点,单独训练一个深度强化学习模型是不现实的。

2)云中心可以提供全局最优决策服务来训练深度强化学习模型,再将模型部署到边缘端,从而提升模型的整体效果。

3)由于边缘智能终端的计算资源有限,在边缘侧部署智能算法能拓宽边缘计算的应用场景。由于与终端设备的地理位置接近,边缘服务器可以支持深度强化学习模型的实时执行,可以显著加快决策时间。通过云边协同范式与深度强化学习的结合,可以更加充分地发挥两者的优势,实现具备低计算成本、全局优化能力、低延迟控制等诸多特征的新型配电系统运行优化。

一些学者研究了云边智能在配电系统中的应用,云边智能在配电系统中的应用如附录C图C1所示,其中云中心负责模型的训练,训练好的模型被部署到边缘端进行执行^[55-58]。云边协同与多智能体深度强化技术十分契合^[50],通过在云中心利用全局视角进行集中式训练,并将训练好的多智能体模型部署在边缘端中通过局部观测实现模型的局部智能控制,可以更好地发挥两者的优势。由于模型自适应更新的需求以及可能涉及的数据隐私问题,云边智

能逐步向边缘智能发展。然而,传统的人工智能框架是为云设计的,没有考虑硬件性能和能耗,不适合边缘。因此,轻量级深度强化学习模型的设计成为未来的发展趋势,这涉及到许多因素的限制,包括计算能力、数据特征和属性以及应用需求。目前,一些专家通过迁移学习技术实现云中心辅助边缘端进行训练^[59-60]。此外,在边缘服务器上部署一个浅层模型,在云服务器上部署一个深层模型,让深层模型协助培训浅层模型,以提高后者的性能,也是一种行之有效的方^[61]。针对云边协同结构设计高效的轻量级深度学习模型的研究尚处于起步阶段。

此外,在采用深度强化学习技术学习第三方主体的互动交易策略时,由于单个第三方主体的数据量限制,难以训练出性能优越的模型。联邦学习与云边协同的结合是一种很好的解决方案,原始私人数据保存在用户的本地设备上,仅共享中间模型参数,从而在保护用户隐私的前提下提升模型的性能^[62-63]。然而,多元异构主体拥有的训练数据通常是非独立同分布的,这对联邦学习的性能造成了很大的影响。

最后,对深度强化学习各类算法的具体优劣与适应情况总结如表1所示。

表1 各类算法的优劣和适用情形

Table 1 Advantages, disadvantages and applicable circumstances of various algorithms

算法	优点	缺点	适用情形
单智能体深度强化算法	算法简单,易于实现;可控制单个区域内的所有设备;可处理离散、连续及混合动作空间	只适用于小规模设备;动作空间过大会导致学习困难;随着配电系统复杂度增加,性能下降	配电系统小规模优化问题
多智能体深度强化算法	可以多个区域或设备协同控制;可以通过分解动作空间提升全局性能	环境非平稳性可能影响收敛性;智能体数量过多时,可能导致全局性能下降	配电系统大规模优化问题,分散式优化需求
安全深度强化算法	保证智能体不会违反物理约束	奖励函数塑造方法惩罚因子调节难度大,可能影响学习效果;策略网络附加安全层可能增加算法不稳定性;基于约束MDP的算法存在计算复杂度和扩展性限制	需要确保安全性的优化问题
离线深度强化算法	可以减少与配电系统的直接交互,降低试错成本与风险	需要大量多样化的历史数据;虚拟环境可能无法准确模拟真实系统特性	历史数据丰富;对虚拟环境准确性要求不高
涉及多主体交互的深度强化学习算法	可以模拟智能体之间的交互与协调关系	多主体间博弈模型复杂,导致训练时间长	适用于多主体互动优化问题
云边智能算法	实时性高,低延迟;计算资源协同节省带宽与计算资源	系统复杂度增加;边缘端计算资源有限;存在安全性和隐私问题	低延迟要求场景,多层次优化需求

4 结论与展望

本文围绕新型配电系统多元化分布式资源、多类型主体、规模化发展等特征进行分析,讨论了新型配电系统运行的关键挑战,如高度不确定性、互动特征难以建模、规模化系统问题难以求解。并对基于深度强化学习的新型配电系统运行优化关键技术进行了总结与探讨。基于对目前新型配电系统优化运行各方面的分析,未来亟需研究以下技术。

1)基于分层强化学习技术的多时间尺度新型配电系统运行优化技术。由于在配电系统存在不同操作特性的设备,在采用深度强化学习进行多时间尺度上的优化往往将离散设备与连续设备分成2个单独时间尺度上的任务分别进行优化,或者将2种设备放在同一时间尺度上进行优化,这不利于多时间尺度上的运行优化协同。因此,如何设计分层强化学习架构,将多时间尺度任务进行分层及协同优化,是未来的关键技术之一。

2)适应海量智能体的多智能体深度强化学习算法。与部分单智能体算法不同,目前多智能算法通常基于直觉进行设计改进,从理论上缺乏对于多智能体联合策略单调不下降的理论证明。因此,传统

方法通常解决的是少量智能体的任务,随着海量异构设备的接入,智能体数量增长,环境的非平稳性不断增大,可能无法适应海量智能体的优化任务。亟需研究适应海量智能体的多智能体深度强化学习算法以及联合策略单调提升的理论验证。

3)考虑基于模型的虚拟环境不准确性的深度强化学习方法架构。深度强化学习通常需要构建基于模型的虚拟环境进行交互学习,并放置于在线环境中进行执行,模型的不准确性与滞后性极大地影响了深度强化学习的实际应用。因此,亟需深入分析物理模型的不准确性,研究新型的深度强化学习方法与神经网络架构,如域随机化方法等。

4)基于混合作 / 竞争设置的多智能体强化学习互动博弈优化技术。在多智能体深度强化学习混合设置中,既考虑合作元素也考虑竞争元素。在新型配电系统中,多类型主体之间不仅仅是存在单一的合作或者竞争关系,计及系统安全运行的混合博弈问题数据建模复杂,基于混合设置的多智能体强化学习算法有潜力为各种智能体之间的复杂互动提供更全面的建模方法。

5)基于联邦学习的新型配电系统第三方主体智能互动交易技术。在采用深度强化学习技术学习第

三方主体的互动交易策略时,由于单个第三方主体的数据量限制,难以训练出性能优越的模型。可以采用联邦学习,将原始私人数据保存在用户的本地设备上,仅共享中间模型参数,从而在保护用户隐私的前提下提升模型的性能,但是需要考虑多元异构用户拥有的训练数据通常是非独立同分布的,亟需研究新型联邦学习技术。

6)适用于边缘端的轻量级深度学习模型。传统的人工智能框架是为云计算设计的,没有考虑硬件性能和能耗,不适合边缘。有2个优化方向:一是自上向下优化,即对在云中训练好的人工智能模型进行压缩放置在边缘端中,然而,过高的压缩比将不可避免地导致模型性能损失;二是自下而上的优化,即设计一套适用于边缘环境的轻量级深度学习模型。轻量级深度学习模型的设计需要考虑许多因素的限制,包括计算能力、数据特征和属性以及应用需求。针对云边协同结构设计高效的轻量级深度学习模型的研究尚处于起步阶段。

此外,尽管基于深度强化学习的新型配电系统优化运行已取得了很大的进展,但在数据准备、可扩展性、场景迁移等方面仍面临较大挑战。

1)数据质量低。电力系统数据可能存在噪声,数据不完整性或出现异常值,直接影响模型的训练效果。

2)可扩展性差。新型配电系统一、二次网架变化,可控负荷、分布式电源、储能等海量设备/资源投切变化均会引发深度强化学习的状态空间、动作空间改变,进而导致原模型无法适用,需重新训练新的模型。

3)场景迁移能力不足。不同区域的拓扑结构、不同时间段的负荷特性、设备参数可能变化较大,导致训练模型的泛化能力下降,使得原模型难以适应新场景。同时,数据驱动方法在极端条件(如极端天气、网络故障)下的表现可能难以保障,无法可靠应对未知场景。模型泛化能力提高的关键在于对外部环境变化的精准感知与准确值的估计,涉及到训练方式、奖励函数(如归一化奖励函数)、以及神经网络架构的进一步优化。

这些局限与挑战表明,数据驱动方法需要与物理模型、专家知识等传统方法结合,通过物理-数据融合、模型简化与迁移学习等技术途径,进一步提升其实用性与可靠性。

附录见本刊网络版(<http://www.epae.cn>)。

参考文献:

- [1] 张智刚,康重庆. 碳中和目标下构建新型电力系统的挑战与展望[J]. 中国电机工程学报, 2022, 42(8): 2806-2819.
ZHANG Zhigang, KANG Chongqing. Challenges and prospects

for constructing the new-type power system towards a carbon neutrality future[J]. Proceedings of the CSEE, 2022, 42(8): 2806-2819.

- [2] 李承周,王宁玲,窦潇潇,等. 多能源互补分布式能源系统集成研究综述及展望[J]. 中国电机工程学报, 2023, 43(18): 7127-7150.
LI Chengzhou, WANG Ningling, DOU Xiaoxiao, et al. Review and prospect on the system integration of distributed energy system with the complementation of multiple energy sources[J]. Proceedings of the CSEE, 2023, 43(18): 7127-7150.
- [3] 杨楠,董邦天,黄禹,等. 考虑不确定性和多主体博弈的增量配电网源网荷协同规划方法[J]. 中国电机工程学报, 2019, 39(9): 2689-2702.
YANG Nan, DONG Bangtian, HUANG Yu, et al. Incremental distribution network source-load collaborative planning method considering uncertainty and multi-agent game[J]. Proceedings of the CSEE, 2019, 39(9): 2689-2702.
- [4] 董旭柱,华祝虎,尚磊,等. 新型配电系统形态特征与技术展望[J]. 高电压技术, 2021, 47(9): 3021-3035.
DONG Xuzhu, HUA Zhuhu, SHANG Lei, et al. Morphological characteristics and technology prospect of new distribution system[J]. High Voltage Engineering, 2021, 47(9): 3021-3035.
- [5] 王育飞,陈强,郑云平,等. 基于动态无功电压灵敏度的有源配电网移动式储能优化调度[J]. 电力自动化设备, 2023, 43(2): 29-35, 59.
WANG Yufei, CHEN Qiang, ZHENG Yunping, et al. Optimal scheduling of mobile energy storage in active distribution network based on dynamic reactive power and voltage sensitivity[J]. Electric Power Automation Equipment, 2023, 43(2): 29-35, 59.
- [6] 白泽,刘灏,毕天姝. 基于图注意力网络的配电网分布式光伏实时消纳能力评估方法[J]. 电力自动化设备, 2024, 44(12): 92-99.
BAI Ze, LIU Hao, BI Tianshu. Real-time distributed PV consumption capacity assessment method for distribution network based on graph attention network[J]. Electric Power Automation Equipment, 2024, 44(12): 92-99.
- [7] 邹才能,熊波,薛华庆,等. 新能源在碳中和中的地位与作用[J]. 石油勘探与开发, 2021, 48(2): 411-420.
ZOU Caineng, XIONG Bo, XUE Huaqing, et al. The role of new energy in carbon neutral[J]. Petroleum Exploration & Development, 2021, 48(2): 411-420.
- [8] ESPE E, POTDAR V, CHANG E. Prosumer communities and relationships in smart grids: a literature review, evolution and future directions[J]. Energies, 2018, 11(10): 2528.
- [9] LUTHANDER R, WIDÉN J, NILSSON D, et al. Photovoltaic self-consumption in buildings: a review[J]. Applied energy, 2015, 142: 80-94.
- [10] 严兴煜,高赐威,陈涛,等. 数字孪生虚拟电厂系统框架设计及其实践展望[J]. 中国电机工程学报, 2023, 43(2): 604-618.
YAN Xingyu, GAO Ciwei, CHEN Tao, et al. Framework design and application prospect for digital twin virtual power plant system[J]. Proceedings of the CSEE, 2023, 43(2): 604-618.
- [11] POPLAVSKAYA K, DE VRIES L. Behind and beyond the meter[M]. Amsterdam, Netherland: Elsevier, 2020: 105-135.
- [12] 谢冬清,李兵,丁炆,等. 含低风速分散式风电的配电网动态鲁棒重构[J]. 太阳能学报, 2021, 42(11): 226-233.
XIE Dongqing, LI Bing, DING Yang, et al. Dynamic robust re-configuration of distribution network involving low-wind-speed dispersed wind power[J]. Acta Energiæ Solaris Sinica, 2021, 42(11): 226-233.

- [13] 黄伟,刘斯亮,羿应棋,等. 基于光伏并网点电压优化的配电网多时间尺度趋优控制[J]. 电力系统自动化,2019,43(3):92-100.
HUANG Wei, LIU Siliang, YI Yingqi, et al. Multi-time-scale slack optimal control in distribution network based on voltage optimization for point of common coupling of PV [J]. Automation of Electric Power Systems, 2019, 43(3): 92-100.
- [14] 金国彬,潘狄,陈庆,等. 考虑源荷不确定性的直流配电网模糊随机日前优化调度[J]. 电工技术学报,2021,36(21):4517-4528.
JIN Guobin, PAN Di, CHEN Qing, et al. Fuzzy random day-ahead optimal dispatch of DC distribution network considering the uncertainty of source-load [J]. Transactions of China Electrotechnical Society, 2021, 36(21): 4517-4528.
- [15] 马望,高红均,李海波,等. 考虑智能软开关的配电网灵活性评估及优化调度模型[J]. 电网技术,2019,43(11):3935-3943.
MA Wang, GAO Hongjun, LI Haibo, et al. Flexibility evaluation and optimal dispatch model of distribution network considering soft open point [J]. Power System Technology, 2019, 43(11): 3935-3943.
- [16] YANG Y, WU W C. A distributionally robust optimization model for real-time power dispatch in distribution networks [J]. IEEE Transactions on Smart Grid, 2019, 10(4): 3743-3752.
- [17] 贺帅佳,阮贺彬,高红均,等. 分布鲁棒优化方法在电力系统中的理论分析与应用综述[J]. 电力系统自动化,2020,44(14):179-191.
HE Shuaijia, RUAN Hebin, GAO Hongjun, et al. Overview on theory analysis and application of distributionally robust optimization method in power system [J]. Automation of Electric Power Systems, 2020, 44(14): 179-191.
- [18] YANG T S, GUO Y, DENG L R, et al. A linear branch flow model for radial distribution networks and its application to reactive power optimization and network reconfiguration [J]. IEEE Transactions on Smart Grid, 2021, 12(3): 2027-2036.
- [19] CHEN S H, HU W H, CHEN Z. Comprehensive cost minimization in distribution networks using segmented-time feeder reconfiguration and reactive power control of distributed generators [J]. IEEE Transactions on Power Systems, 2016, 31(2): 983-993.
- [20] HELMI A M, CARLI R, DOTOLI M, et al. Efficient and sustainable reconfiguration of distribution networks via metaheuristic optimization [J]. IEEE Transactions on Automation Science and Engineering, 2022, 19(1): 82-98.
- [21] 杨志淳,闵怀东,杨帆,等. 新型配电系统层级式数字化架构体系发展范式[J]. 电力自动化设备,2024,44(12):42-51.
YANG Zhichun, MIN Huaidong, YANG Fan, et al. Development paradigm of hierarchical digital architecture system for novel power distribution system [J]. Electric Power Automation Equipment, 2024, 44(12): 42-51.
- [22] BUI V H, SU W C. Real-time operation of distribution network: a deep reinforcement learning-based reconfiguration approach [J]. Sustainable Energy Technologies and Assessments, 2022, 50: 101841.
- [23] 王子晗,高红均,高艺文,等. 基于深度强化学习的城市配电网多级动态重构优化运行方法[J]. 电力系统保护与控制,2022,50(24):60-70.
WANG Zihan, GAO Hongjun, GAO Yiwen, et al. Multi-level dynamic reconfiguration and operation optimization method for an urban distribution network based on deep reinforcement learning [J]. Power System Protection and Control, 2022, 50(24): 60-70.
- [24] 王光华,李晓影,宋秉睿,等. 基于深度强化学习的配电网负荷转供控制方法[J]. 电力自动化设备,2022,42(7):253-260.
WANG Guanghua, LI Xiaoying, SONG Bingrui, et al. Load transfer control method of distribution network based on deep reinforcement learning [J]. Electric Power Automation Equipment, 2022, 42(7): 253-260.
- [25] 张沛,朱驻军,谢桦. 基于深度强化学习近端策略优化的电网无功优化方法[J]. 电网技术,2023,47(2):562-572.
ZHANG Pei, ZHU Zhujun, XIE Hua. Reactive power optimization based on proximal policy optimization of deep reinforcement learning [J]. Power System Technology, 2023, 47(2): 562-572.
- [26] JIN J L, XU Y J. Optimal policy characterization enhanced actor-critic approach for electric vehicle charging scheduling in a power distribution network [J]. IEEE Transactions on Smart Grid, 2021, 12(2): 1416-1428.
- [27] SUN X Z, QIU J. Two-stage Volt/VAR control in active distribution networks with multi-agent deep reinforcement learning method [J]. IEEE Transactions on Smart Grid, 2021, 12(4): 2903-2912.
- [28] 倪爽,崔承刚,杨宁,等. 基于深度强化学习的配电网多时间尺度在线无功优化[J]. 电力系统自动化,2021,45(10):77-85.
NI Shuang, CUI Chenggang, YANG Ning, et al. Multi-time-scale online optimization for reactive power of distribution network based on deep reinforcement learning [J]. Automation of Electric Power Systems, 2021, 45(10): 77-85.
- [29] 张剑,崔明建,姚潇毅,等. 基于数据驱动与物理模型的主动配电网双时间尺度协调优化[J]. 电力系统自动化,2023,47(20):64-71.
ZHANG Jian, CUI Mingjian, YAO Xiaoyi, et al. Dual-time-scale active and reactive power coordinated optimization for active distribution network based on data-driven and physical model [J]. Automation of Electric Power Systems, 2023, 47(20): 64-71.
- [30] LI H P, WAN Z Q, HE H B. Real-time residential demand response [J]. IEEE Transactions on Smart Grid, 2020, 11(5): 4144-4154.
- [31] RASHID T, SAMVELYAN M, DE WITT C S, et al. Monotonic value function factorisation for deep multi-agent reinforcement learning [J]. The Journal of Machine Learning Research, 2020, 21(1): 7234-7284.
- [32] 刘硕,郭创新,冯斌,等. 基于价值分解深度强化学习的分布式光伏主动电压控制方法[J]. 电力自动化设备,2023,43(10):152-159.
LIU Shuo, GUO Chuangxin, FENG Bin, et al. Active voltage control method of distributed photovoltaic based on value decomposition deep reinforcement learning [J]. Electric Power Automation Equipment, 2023, 43(10): 152-159.
- [33] 徐博涵,向月,潘力,等. 基于深度强化学习的含高比例可再生能源配电网就地分散式电压管控方法[J]. 电力系统保护与控制,2022,50(22):100-109.
XU Bohan, XIANG Yue, PAN Li, et al. Local decentralized voltage management of a distribution network with a high proportion of renewable energy based on deep reinforcement learning [J]. Power System Protection and Control, 2022, 50(22): 100-109.
- [34] 全歆,彭显刚,刘涵予,等. 基于深度强化学习的配电网实时电压优化控制方法[J]. 电网技术,2023,47(5):2029-2039.
QUAN Huan, PENG Xiangang, LIU Hanyu, et al. Voltage optimal control of distribution network based on deep reinforcement learning [J]. Power System Technology, 2023, 47(5): 2029-2038.

- [35] 蒲天骄,杜帅,李烨,等. 面向隐私保护基于联邦强化学习的分布式电源协同优化策略[J]. 电力系统自动化,2023,47(8):62-70.
PU Tianjiao, DU Shuai, LI Ye, et al. Collaborative optimization strategy of distributed generators based on federated reinforcement learning for privacy preservation[J]. Automation of Electric Power Systems, 2023, 47(8): 62-70.
- [36] WU T, WANG J H, LU X N, et al. AC/DC hybrid distribution network reconfiguration with microgrid formation using multi-agent soft actor-critic[J]. Applied Energy, 2022, 307: 118189.
- [37] 谢黎龙,李勇汇,范培潇,等. 基于深度强化学习的孤立多微电网系统频率和电压综合控制[J]. 电力自动化设备,2024,44(6):118-126.
XIE Lilong, LI Yonghui, FAN Peixiao, et al. Deep reinforcement learning-based integrated frequency and voltage control for isolated multi-microgrid system[J]. Electric Power Automation Equipment, 2024, 44(6): 118-126.
- [38] WANG Y, QIU D W, STRBAC G, et al. Coordinated electric vehicle active and reactive power control for active distribution networks[J]. IEEE Transactions on Industrial Informatics, 2023, 19(2): 1611-1622.
- [39] QIU D W, WANG Y, WANG J H, et al. Resilience-oriented coordination of networked microgrids: a shapley Q-value learning approach[J]. IEEE Transactions on Power Systems, 2023, 39(2): 3401-3416.
- [40] 李琦,乔颖,张宇精. 配电网持续无功优化的深度强化学习方法[J]. 电网技术,2020,44(4):1473-1480.
LI Qi, QIAO Ying, ZHANG Yujing. Continuous reactive power optimization of distribution network using deep reinforcement learning[J]. Power System Technology, 2020, 44(4): 1473-1480.
- [41] 巨云涛,陈希. 基于双层多智能体强化学习的微网群分布式有功无功协调优化调度[J]. 中国电机工程学报,2022,42(23):8534-8548.
JU Yuntao, CHEN Xi. Distributed active and reactive power coordinated optimal scheduling of networked microgrids based on two-layer multi-agent reinforcement learning[J]. Proceedings of the CSEE, 2022, 42(23): 8534-8548.
- [42] CHEN P C, LIU S C, WANG X Z, et al. Physics-shielded multi-agent deep reinforcement learning for safe active voltage control with photovoltaic/battery energy storage systems[J]. IEEE Transactions on Smart Grid, 2023, 14(4): 2656-2667.
- [43] LIU H T, WU W C. Online multi-agent reinforcement learning for decentralized inverter-based Volt-VAR control[J]. IEEE Transactions on Smart Grid, 2021, 12(4): 2980-2990.
- [44] YAN R D, XING Q, XU Y. Multi-agent safe graph reinforcement learning for PV inverters-based real-time decentralized Volt/VAR control in zoned distribution networks[J]. IEEE Transactions on Smart Grid, 2024, 15(1): 299-311.
- [45] SHI X Y, XU Y L, CHEN G B, et al. An augmented Lagrangian-based safe reinforcement learning algorithm for carbon-oriented optimal scheduling of EV aggregators[J]. IEEE Transactions on Smart Grid, 2024, 15(1): 795-809.
- [46] LI H P, WAN Z Q, HE H B. Constrained EV charging scheduling based on safe deep reinforcement learning[J]. IEEE Transactions on Smart Grid, 2020, 11(3): 2427-2439.
- [47] LI H P, HE H B. Learning to operate distribution networks with safe deep reinforcement learning[J]. IEEE Transactions on Smart Grid, 2022, 13(3): 1860-1872.
- [48] AL ZISHAN A, HAJI M M, ARDAKANIAN O. Adaptive congestion control for electric vehicle charging in the smart grid[J]. IEEE Transactions on Smart Grid, 2021, 12(3): 2439-2449.
- [49] GAO Y Q, WANG W, SHI J, et al. Batch-constrained reinforcement learning for dynamic distribution network reconfiguration[J]. IEEE Transactions on Smart Grid, 2020, 11(6): 5357-5369.
- [50] GAO H J, WANG R J, HE S J, et al. A cloud-edge collaboration solution for distribution network reconfiguration using multi-agent deep reinforcement learning[J]. IEEE Transactions on Power Systems, 2024, 39(2): 3867-3879.
- [51] LIU H T, WU W C. Two-stage deep reinforcement learning for inverter-based volt-var control in active distribution networks[J]. IEEE Transactions on Smart Grid, 2021, 12(3): 2037-2047.
- [52] CAO D, ZHAO J B, HU J X, et al. Physics-informed graphical representation-enabled deep reinforcement learning for robust distribution system voltage control[J]. IEEE Transactions on Smart Grid, 2024, 15(1): 233-246.
- [53] YI Z K, XU Y L, WANG X, et al. An improved two-stage deep reinforcement learning approach for regulation service disaggregation in a virtual power plant[J]. IEEE Transactions on Smart Grid, 2022, 13(4): 2844-2858.
- [54] QIAN T, SHAO C C, LI X L, et al. Enhanced coordinated operations of electric power and transportation networks via EV charging services[J]. IEEE Transactions on Smart Grid, 2020, 11(4): 3019-3030.
- [55] JIA Q G, CHEN S J, YAN Z, et al. Optimal incentive strategy in cloud-edge integrated demand response framework for residential air conditioning loads[J]. IEEE Transactions on Cloud Computing, 2022, 10(1): 31-42.
- [56] ZHANG X Y, BIAGIONI D, CAI M M, et al. An edge-cloud integrated solution for buildings demand response using reinforcement learning[J]. IEEE Transactions on Smart Grid, 2021, 12(1): 420-431.
- [57] ROSERO D G, DÍAZ N L, TRUJILLO C L. Cloud and machine learning experiments applied to the energy management in a microgrid cluster[J]. Applied Energy, 2021, 304: 117770.
- [58] LI S H, SUN Y, RAMEZANI M, et al. Artificial neural networks for Volt/VAR control of DER inverters at the grid edge[J]. IEEE Transactions on Smart Grid, 2019, 10(5): 5564-5573.
- [59] 陈思光,陈佳民,赵传信. 基于深度强化学习的云边协同计算迁移研究[J]. 电子学报,2021,49(1):157-166.
CHEN Siguang, CHEN Jiamin, ZHAO Chuanxin. Deep reinforcement learning-based cloud-edge collaborative computation offloading mechanism[J]. Acta Electronica Sinica, 2021, 49(1): 157-166.
- [60] TAO Y C, QIU J, LAI S Y. A hybrid cloud and edge control strategy for demand responses using deep reinforcement learning and transfer learning[J]. IEEE Transactions on Cloud Computing, 2022, 10(1): 56-71.
- [61] DING C T, ZHOU A, LIU Y X, et al. A cloud-edge collaboration framework for cognitive service[J]. IEEE Transactions on Cloud Computing, 2022, 10(3): 1489-1499.
- [62] LI Y, LI J Z, WANG Y. Privacy-preserving spatiotemporal scenario generation of renewable energies: a federated deep generative learning approach[J]. IEEE Transactions on Industrial Informatics, 2022, 18(4): 2310-2320.
- [63] SU Z, WANG Y T, LUAN T H, et al. Secure and efficient federated learning for smart grid with edge-cloud collaboration[J]. IEEE Transactions on Industrial Informatics, 2022, 18(2): 1333-1344.

作者简介:

王仁浚(1997—),男,博士研究生,通信作者,主要研究方向为配电网运行优化、人工智能(**E-mail**: 1328860755@qq.com);

高红均(1989—),男,副教授,博士研究生导师,博士,主要研究方向为高比例可再生能源电力系统与配电网规划运

行、综合能源系统优化与市场交易等(**E-mail**: gaohongjun@scu.edu.cn);

罗龙波(1987—),男,高级工程师,硕士,主要研究方向为电力调度控制运行分析和运行管理(**E-mail**: 104267600@qq.com)。

(实习编辑 丁欣欣)

Review of research on new distribution system optimization operation based on deep reinforcement learning

WANG Renjun¹, GAO Hongjun¹, LUO Longbo², CHEN Minghui², XU Jie¹, LIU Junyong¹

(1. School of Electrical Engineering, Sichuan University, Sichuan 610065, China;

2. Guangzhou Power Supply Bureau of Guangdong Power Grid Co., Ltd., Guangzhou 510620, China)

Abstract: On the background of high penetration of renewable energy, the problems of operation optimization for new distribution system are emerging. Considering the development characteristics of new distribution system, such as extensively access of diversified distributed resources, the interactive collaboration of multi-type entities, and the continuous expansion of system scale, significant challenges such as the high uncertainty of operation, difficulty in modelling complex interaction and difficulty in solving large-scale systems. Aiming at the differentiated characteristics of the operation optimization problem in centralized / decentralized decision-making, safety constraints and multi-agent interaction, key operation optimization technologies based on deep reinforcement learning are systematically sorted out and summarized, the performance advantages and disadvantages of different algorithms and their applicable scenarios are compared. The major technical bottlenecks faced by deep reinforcement learning in optimization operation of new distribution system are summarized, and its future development prospects in multi-time scale coordination, model generalization and intelligent interaction are prospected.

Key words: new distribution system; diversified distributed resources; multi-type entities; large-scale development; deep reinforcement learning

(上接第151页 continued from page 151)

Review of research on traction emergency power supply and emergency rescue technology of electric railway

DENG Wenli¹, JIANG Lin¹, JIANG Yu², DAI Chaohua³

(1. School of Electrical Engineering and Information, Southwest Petroleum University, Chengdu 610500, China;

2. School of Electronic Information and Electrical Engineering, Chengdu University, Chengdu 610106, China;

3. School of Electrical Engineering, Southwest Jiaotong University, Chengdu 611756, China)

Abstract: The traction emergency power supply and emergency rescue technology of electric railway play an important role in improving the independent emergency support and self-rescue ability of traditional traction power supply system in fault scenarios, and reducing the potential risk loss under extreme events. By summarizing the research status and latest progress of traction emergency power supply and emergency rescue technology at home and abroad, four typical schemes are classified, including technologies of traction system self-healing reconstruction, self-traction with onboard energy storage, onboard-ground cooperative traction, emergency of rescue train. The application forms, technical features, operation characteristics and other aspects of different schemes are sorted out, and the advantages and constraints of existing researches are also compared and reviewed. The research difficulties and key problems need to be improved urgently in different schemes from multiple perspectives are analyzed, and the future prospect and corresponding suggestions are indicated.

Key words: electric railway; traction power supply; emergency support; fault rescue; self-traction