

基于深度强化学习的主动配电网 实时调度与电压控制策略研究

黄佳秦

(国网江苏省电力有限公司南通市海门区供电公司, 江苏 南通 226100)

摘要 在主动配电网向高弹性、智能化转型背景下, 实时调度与电压控制面临高维、非线性及不确定性挑战。深度强化学习凭借从高维数据中直接学习最优决策的能力, 为解决该难题提供了新范式。本文首先构建了基于马尔可夫决策过程的主动配电网建模框架, 并选择适配混合动作空间的深度强化学习算法; 接着构建高保真仿真环境进行训练与离线评估; 最后探讨人机协作混合决策框架。同时, 分析了实际部署中测量通信、计算资源、系统兼容等关键问题, 并提出轻量化模型压缩、混合增强学习等应对策略, 以期为深度强化学习的进一步应用提供系统的技术路径参考。

关键词 深度强化学习; 主动配电网; 实时调度; 电压控制

中图分类号: TM761

文献标志码: A

DOI: 10.3969/j.issn.2097-3365.2025.35.005

0 引言

随着分布式光伏发电等可再生能源渗透率的提高, 由于其输出具有随机性、波动性的特点, 不利于配电网系统保持实时平衡和电压稳定。针对这种高维、非线性和不确定性的问题时, 传统的优化调度方法模型存在准确性不足、计算复杂方面的局限性。主动配电网可以协调分布式电源、储能、柔性负荷等可控资源来实现自主调度潮流电压, 但是如何实时高效、经济协同控制仍然是一个问题^[1]。深度强化学习作为一种能够直接从数据中获得最优决策的技术, 提供了解决该类复杂问题的新路径。

1 深度强化学习基本理论

深度强化学习是机器学习的重要组成部分, 它是使智能体能够基于其与环境交互试错过程中得到的经验来寻求获得长期目标最优解的一种算法。其核心要素是马尔可夫决策过程, 涵盖状态空间、动作空间、状态转移概率、奖励函数和折扣因子等多项参数。当智能体根据当前观测到的环境状态执行某个动作后, 相应的状态就会发生改变, 之后便会反馈回来一个标量奖励^[2]。

深度强化学习是利用深度神经网络函数强大的逼近能力, 在高维的状态 / 动作空间中做决策。主要分为两种: 一是以深度 Q 网络及其改进算法为代表的基于值函数的方法, 即用深度神经网络逼近 Q 值函数, 再借助经验回放、目标网络等方法稳定 DQN 的训练过程,

适合用于离散的动作空间; 二是以深度确定性策略梯度为代表的基于策略梯度的方法, 深度确定性策略梯度利用演员—评论家结构, 用演员网络直接产生对应的确定性连续动作, 评论家网络负责评价这个动作的价值并利用策略梯度不断迭代更新演员网络参数, 可以解决一些连续控制方面的问题, 比如配电系统调度等。

2 深度强化学习在主动配电网实时调度与电压控制中的应用策略

2.1 基于马尔可夫决策过程的建模框架

深度强化学习为解决主动配电网中高维、非线性和不确定性的实时调度与电压控制难题提供了革命性的范式转变。其核心在于将复杂的优化问题重新表述为一个序贯决策过程, 通过智能体与主动配电网环境的持续交互, 无需精确的物理模型和不确定性先验知识, 即可自主学习逼近最优的控制策略。应用深度强化学习的首要步骤是进行精准的马尔可夫决策过程建模: 状态空间需全面捕捉系统的动态特性, 通常包括所有节点的电压幅值、关键支路的有功与无功潮流、分布式光伏的实际出力与负荷需求、储能系统的实时荷电状态以及有载调压变压器分接头当前位置和电容器组的投切状态等, 这些数据共同构成了智能体感知环境的基础^[3]。动作空间的设计则直接对应了所有可控设备的调节指令。例如: 指令光伏逆变器调节其无功出力、控制储能的充放电功率、调整分接头档位以及执行投切操作, 对于连续控制变量, 动作空间是连

续的，而对于离散设备，则需设计为离散动作或进行连续量的离散化处理。奖励函数是引导智能体学习行为的“指挥棒”，其设计尤为关键，必须紧密围绕优化目标，通常为一个复合函数，包含运行成本的负项、对所有节点电压偏差的惩罚项以及对违反设备运行约束的严厉惩罚，通过精心设计各分量的权重，奖励函数能够有效平衡经济性与安全性。折扣因子则决定了智能体对远期回报的重视程度，在主动配电网调度中，一个小于1但接近1的折扣因子值有助于智能体在追求即时电压稳定的同时，也考虑长期的经济运行效益。

2.2 面向混合动作空间的深度强化学习算法选择与设计

主动配电网混合了连续控制和离散控制的动作特性，仅用DQN或者传统的深度学习方法无法满足任务需求，需要选用改进型或混合型算法，如将深度学习作为基础框架，采用深度学习方式解决连续控制的逆变器、储能器等控制，同时为离散装置配备单独的DQN网络或构建并行离散的动作输出头。另外，还可以选用近端策略优化这类更稳健的策略梯度算法，在这种算法下加入裁剪手段，避免其发生大的策略更新步长而导致性能崩溃。对于高危类型的电网系统而言，这种做法可以降低其训练风险。但需要注意的是，不论采取哪一种算法，都需要增加一些技术手段来提高学习效率与稳定性。

2.3 高保真仿真环境的构建与不确定性模拟

深度强化学习智能体在应用过程中，将其直接放到实际的物理电网中去进行探索性学习显然是非常危险的，因此应该提前建立一套高精度、高效率的仿真训练环境。基于电力系统的仿真软件或是自行编写的前推回代潮流计算程序搭建训练环境，并将这种仿真训练环境作为强化学习的行动者接受动作指令，求解配电网潮流，计算出新的系统状态，并反馈相应的奖励^[4]。为了让训练出来的策略能够应对所有可能出现的不确定性情况，就需要要求仿真环境必须能够再现主动配电网的动态性与随机性，要求对历史数据分析后能够生成具有典型日特征与随机波动性的光伏出力和负荷曲线，从而使智能体在训练过程中充分经历各种运行情况，避免学习出只适用于理想确定环境下执行的脆弱策略。此外，仿真环境的步长应符合实际运行需要的实时调度长度，通常为分钟级或15分钟级，这样才能满足及时控制的需要。

2.4 离线评估与基准测试

完成了仿真环境下的训练后，便能获得初步成熟度的深度强化学习智能体，但是在投入在线应用前还

需要对其进行性能上的评估和检验。评估过程指的是将该智能体部署到一系列未进行过训练的独立场景上进行运行测试，并将该智能体的控制性能与其他多种传统方法进行系统性对比，常见的基准方法包括基于完美预测的确定性优化、模型预测控制以及各种启发式规则控制。评价指标涵盖技术、经济等多个维度，如总运行成本、电压偏差指数、决策时间以及可再生能源消纳率等指标^[5]。通过对比分析，对中量化出深度强化学习策略相比其他方法在经济性、电压质量、计算效率以及对不确定性适应能力上的优势与不足，最重要的是要进行安全验证，检验智能体是否会在长时间运行过程中出现危害设备或者系统稳定的潜在的风险动作，在线评估中不可行的智能体不能被采纳。

2.5 人机协作的混合决策框架

为确保万无一失，将训练验证后的深度强化学习智能体应用于实际主动配电网时，必须采用一种人机协作、安全优先的混合决策框架，而非完全自主的“黑箱”控制。最主流和可靠的架构是“深度强化学习建议—模型校验—人工决策”的混合模式。在此模式下，深度强化学习智能体在每个调度时刻根据实时状态生成初步的控制动作建议集，但该建议不会直接被下发至现场设备。取而代之的是，一个快速的、基于精确潮流模型的安全校验模块会立即模拟执行该套动作，预测其执行后的系统状态，并严格校验所有安全约束是否会被违反。若校验通过，建议动作可被发送给调度人员作为强参考，或经授权后自动执行；若校验发现风险，则系统会触发警报，并自动切换至基于传统模型的保守安全控制策略，同时记录此次异常以供后续分析和对智能体进行微调。这种架构将深度强化学习的全局优化和快速响应能力与物理模型的精确性和安全保障相结合，极大地降低了直接应用人工智能技术的潜在风险，是当前阶段将深度强化学习研究成果转化为实际生产力的关键桥梁。

3 工程应用挑战与对策

3.1 实际部署关键问题

1. 测量装置精度与通信可靠性要求。深度强化学习控制策略的效能高度依赖于状态输入的准确性和动作执行的成功率。在实际配电网中，电压、电流互感器等测量装置的精度误差以及数据采集的不同步问题，会导致智能体感知的环境状态失真，形成“垃圾进，垃圾出”的困境，可能引发次优甚至不安全的控制决策。同时，控制指令的可靠下发依赖于通信网络，其延迟、丢包或中断会破坏控制的实时性与闭环稳定性，尤其对于频繁动作的设备，通信故障可能导致设备损坏或

系统振荡。因此，部署深度强化学习对量测系统的精度和通信网络的可靠性提出了远高于传统控制的要求。

2. 算法计算资源需求与嵌入式实现。主流的深度强化学习算法特别是基于深度神经网络的模型，不论是训练还是推理的过程都需要庞大的计算资源。在训练过程中需要用高配的 GPU 服务器完成长时间的数据模拟运算，持续数天甚至数周；而在部署时虽然推理过程所需计算量不大，但是若将其嵌入配电网现场的边缘计算装置中却依旧面临着较大挑战。由于这些终端装置上的 CPU 算力、内存及功耗都十分有限，并不能流畅地运行深度网络大模型，因此不可能满足毫秒级甚至是亚秒级的实时决策要求。只有将已经训练好的模型压缩到适合边缘端的规模，并能在条件极为有限的嵌入式平台中正常运行，才是实现落地应用的关键技术瓶颈。

3. 与现有调度自动化系统的兼容性设计。深度强化学习系统作为一套新兴的决策引擎，必须能够无缝集成到这一现有生态中。这涉及多方面的兼容性设计：首先，需要开发数据接口适配器，将调度自动化系统的实时数据流转换为深度强化学习智能体所需的状态向量；其次，深度强化学习输出的控制指令需要被封装成调度自动化系统能够识别和执行的标准化指令；最后，整个交互过程必须符合电力系统安全防护规定，确保不会对现有系统的稳定运行构成威胁。任何兼容性上的隔阂都将极大地增加深度强化学习系统的部署成本和复杂度。

3.2 应对策略

1. 轻量化模型压缩技术。为满足嵌入式平台的算力约束，模型压缩是必不可少的步骤。知识蒸馏是一种有效的轻量化技术，其核心思想是训练一个庞大而精确的“教师模型”，然后利用其输出的软标签来指导一个结构更小巧的“学生模型”进行训练。通过这种方式，学生模型能够继承教师模型的核心知识与泛化能力，同时参数量和计算量大幅减少，更易于在边缘设备上部署和快速推理。此外，还可以结合网络剪枝、量化等技术，进一步降低模型复杂度和存储开销，实现在有限资源下性能与效率的最佳平衡。

2. 混合增强学习框架。为了提升深度强化学习智能体的安全性和可解释性，并加速训练过程，构建混合增强学习框架是一种务实且高效的策略。该框架将深度强化学习智能体的决策与电力系统领域专家总结的规则库相结合。例如：当深度强化学习智能体输出的动作触发了某些预定义的安全边界时，规则系统可以优先执行保守的校正动作；在训练初期，通过专家规则提供示范性动作来引导智能体的探索方向。这种

“深度强化学习为主，规则为辅”的架构，既能发挥深度强化学习处理复杂问题的优势，又能利用专家知识规避高风险区域，增强了系统在实际应用中的鲁棒性和可信度。

3. 分布式协同训练机制。数据孤岛和隐私保护是跨区域协同训练深度强化学习模型时面临的主要障碍。联邦学习为此提供了理想的解决方案。在联邦学习框架下，多个配电网区域可以在不共享本地原始数据的前提下，协同训练一个全局深度强化学习模型。具体而言，各参与方利用本地数据进行模型训练，仅将模型参数的更新上传至中央服务器进行聚合，生成改进后的全局模型再分发给各方。这种方式充分保护了数据隐私，并能够利用更广泛的数据分布来提升全局模型的泛化能力和鲁棒性，为构建广域协同的智能配电网调度系统奠定了技术基础。

4 结束语

在主动配电网实时调度与电压控制过程中，通过合理的建模、算法设计与仿真训练，能够有效学习到协调多种可控资源、应对运行不确定性的智能策略，在保障电压安全的同时优化经济运行。然而，从实验室仿真到工程现场的跨越仍面临诸多挑战，包括对量测通信系统的高要求、嵌入式部署的计算资源限制以及与现有调度体系的融合问题。通过采用模型轻量化、构建人机协作的混合增强框架、利用联邦学习进行协同训练等策略，可以逐步克服这些障碍。随着算法效率、数据质量与硬件算力的持续提升，深度强化学习有望成为构建智能、灵活、坚韧新型配电系统的核心技术之一，为实现“双碳”目标下的能源转型提供关键支撑。

参考文献：

- [1] 吉兴全, 孙辰昊, 张玉敏, 等. 基于多智能体与改进目标级联法的输配协同优化调度[J]. 电力系统自动化, 2025, 49(02):165-174.
- [2] 李士丹, 李航, 李国杰, 等. 考虑分区与模仿学习的深度强化学习配电网电压优化策略[J]. 电力系统保护与控制, 2024, 52(22):1-11.
- [3] 朱东方, 朱丹丹, 周前, 等. 基于多智能体深度强化学习的新型配电系统多时间尺度无功电压分层分区优化研究[J]. 华北电力大学学报(自然科学版), 2025, 10(30):1-13.
- [4] 黄冬梅, 余京朋, 崔承刚, 等. 多时间尺度深度强化学习光储配电网电压优化[J]. 中国电机工程学报, 2025, 45(17): 6709-6723.
- [5] 耿天旭, 梁俊宇, 龚新勇, 等. 基于深度强化学习的配电网多主体协同电压控制方法[J]. 电网与清洁能源, 2024, 40(09):74-80, 91.