

Research on Optimization of Nuclear Power Plant Primary Circuit Pressure Control Based on Reinforcement Learning

Wei Wang

Daya Bay Nuclear Power Operation Management Co., Ltd., Shenzhen, Guangdong, 518000, China

Abstract

The primary circuit pressure in nuclear power plants serves as a critical parameter for ensuring reactor safety and stable operation. The precision of control directly impacts both operational efficiency and nuclear safety margins. Traditional control methods struggle to balance dynamic response speed and steady-state accuracy when dealing with nonlinear systems, time-varying parameters, and complex disturbances in primary circuit systems. This paper proposes a reinforcement learning-based optimization scheme for primary circuit pressure control. The study first analyzes the control mechanisms of nuclear power plant primary circuit pressure and the core theories of reinforcement learning, then establishes a reinforcement learning model tailored for pressure control scenarios. By improving traditional algorithms to enhance control robustness and designing comprehensive optimization strategies, the proposed approach demonstrates advantages in disturbance adaptability and steady-state error control through simulation platform-based comparative experiments. This work aims to provide a more reliable technical pathway for critical parameter control in nuclear power plants, contributing to safer and more efficient reactor operations.

Keywords

Nuclear power plant; Primary circuit pressure control; Reinforcement learning; Control optimization; Robustness

基于强化学习的核电站一回路压力控制优化研究

王伟

大亚湾核电运营管理有限责任公司, 中国·广东深圳 518000

摘要

核电站一回路压力是保证反应堆安全稳定运行的重要参数, 控制精度的好坏直接影响到机组的运行效率以及核安全边界。传统的控制方法对于一回路系统非线性、参数时变以及复杂扰动来说, 很难在动态响应速度和稳态精度之间取得平衡。本文提出基于强化学习的一回路压力控制优化方案, 先分析核电站一回路压力控制机理和强化学习的核心理论, 建立适合于压力控制场景的强化学习模型, 改进传统算法提高控制鲁棒性, 设计完整的优化策略实现流程。通过搭建仿真平台来开展对比实验, 验证所提方案在扰动适应性、稳态误差控制等方面的优势。希望给核电站关键参数控制提供更可靠的技术途径, 有利于机组安全高效运行。

关键词

核电站; 一回路压力控制; 强化学习; 控制优化; 鲁棒性

1 引言

核电站是清洁能源供应的重要支柱, 运行安全、稳定一直为核工业领域的重要研究课题。一回路是核电站反应堆冷却系统的核心组成部分, 它承担导出堆芯热量、维持反应堆正常运行的重要任务, 压力参数的稳定控制对于防止堆芯过热、避免系统超压破损具有极其重要的作用。传统一回路压力控制多用 PID 控制、模糊控制等方法, 虽然结构简单、易于实现, 但是面对负荷剧烈波动、设备老化造成的参数漂移、外部复杂扰动时, 控制参数不能实时自适应调整, 容易

出现超调量大、稳态误差超标等问题。强化学习因为没有模型自适应、在线迭代优化的特点, 在复杂非线性系统控制方面有独特的优势。本文根据强化学习理论设计回路压力控制优化方案, 解决传统控制方法存在的不足, 为核电站一回路压力控制提供更高效、可靠的技术支持。

2 相关理论基础

2.1 核电站一回路压力控制基础

核电站一回路系统由反应堆压力容器、蒸汽发生器、主循环泵、稳压器以及连接管道等组成, 它的主要功能是把反应堆堆芯产生的热量传递到蒸汽发生器的二次侧, 给二回路发电提供动力。一回路压力的稳定直接影响堆芯冷却效率、蒸汽发生器传热效果、系统设备寿命, 其控制目标就是

【作者简介】王伟(1987—), 男, 中国安徽池州人, 硕士, 工程师, 从事核电运行研究。

将压力保持在设定的阈值范围（一般为 15 ~ 16MPa）内，抑制负荷变化、设备扰动等因素引起的压力波动。

传统的一回路压力控制是以稳压器为执行设备，通过调节稳压器电加热器功率、喷淋流量和安全阀开度来实现压力的控制^[1]。当一回路压力低于设定值时，启动电加热器加热稳压器内水体，使水体膨胀从而提高系统压力；当压力高于设定值时，开启喷淋装置向稳压器内喷洒低温水，或者通过安全阀泄放一部分高压蒸汽来降低压力。但是，一回路系统具有明显的非线性特性，主循环泵流量与压力响应的非线性关系、蒸汽发生器传热系数的时变性，受到反应堆负荷调整、设备故障前兆等扰动的影响，传统控制方法依靠固定的控制参数，不能实现全工况下的精准控制，容易出现压力波动超过允许范围的情况，给机组运行带来安全隐患。

2.2 强化学习核心理论

强化学习是机器学习中的一种重要的分支，它的思想就是智能体和环境不断交互，根据从环境中得到的信息，对行动策略做出调整来改善目标函数，最终达成最佳决策。强化学习与监督学习、无监督学习不一样，不需要先验标签数据，只需“交互、反馈、迭代”的方式进行自适应优化，可以应用于难以建立精确数学模型的复杂系统。强化学习的基本框架是由智能体（Agent）、环境（Environment）、状态（State）、动作（Action）、奖励（Reward）这五个主要部分组成的。智能体是执行控制决策的主体，对应本文中的压力控制模块；环境是被控对象，对应核电站一回路压力系统；状态是环境在某一时刻的特征描述，对应一回路压力、温度、流量等参数组合；动作是智能体对环境的干预行为，对应稳压器加热功率、喷淋流量等调节指令；奖励是环境对智能体动作的反馈信号，用于评价动作的好坏，对应压力控制精度、能耗等目标。对于核电站一回路压力控制场景，强化学习的无模型特性可以规避系统非线性、参数时变造成的建模难题，在线迭代优化能力可以实现控制参数的实时自适应调整。

3 基于强化学习的一回路压力控制优化方案设计

3.1 控制模型构建

以智能体与环境的交互模式为基本，按照一回路压力系统的运行特点，进行状态空间、动作空间和奖励函数的设定，保证该模型能满足实际控制需要。状态空间设计要全面反映一回路压力系统运行状态，不能维度过高造成算法复杂度增大。本文选取一回路实时压力、压力变化率、稳压器水位、主循环泵流量、蒸汽发生器二次侧蒸汽压力五个主要参数作为状态变量，建立高维状态空间。其中一回路实时压力是核心状态变量，直接决定控制动作的方向，压力变化率反映压力波动的趋势，用来预判压力变化的方向，提前调整控制策略，其余参数用来辅助判断系统运行工况，避免单一参

数反馈造成控制偏差^[2]。为了降低算法复杂度，把各状态变量都归一化到 0 到 1 的区间内，消除了由于参数量级的差异带来的影响。动作空间的设计要对应一回路压力控制的实际调节手段，结合稳压器的的工作原理，将动作空间定义为连续动作空间，包含三项主要动作，即稳压器电加热器功率调节、喷淋装置流量调节、备用安全阀开度调节。其中电加热器功率调节范围为 0 至 100% 额定功率，喷淋流量调节范围为 0 至 80% 额定流量，安全阀开度调节范围为 0 至 20%（只在压力超阈值时使用）。连续动作空间的设计可以实现控制量的平滑调节，可以防止因为离散动作造成的压力突变，从而提高控制的稳定性。

3.2 强化学习算法改进与优化

为了解决传统强化学习算法在控制一回路压力时收敛慢、鲁棒性差的问题，本文在 DQN 算法的基础上提出自适应经验回放深度 Q 网络（AER-DQN）算法，提高算法对于复杂非线性系统控制的性能。传统的 DQN 算法使用随机经验回放机制，把智能体与环境交互得到的经验数据存放在经验池中，在训练的时候随机抽取数据更新网络参数，虽然可以打破经验数据的相关性，防止网络过拟合，但是没有区分经验数据的重要性，大量的无效经验数据会降低训练效率，使收敛速度变慢。用优先级经验回放来提高训练数据的有效性，加快算法收敛的速度。由于一回路压力系统时变特性，传统的 DQN 算法的固定学习率不能同时满足收敛速度和稳态性能的要求，学习率太高容易造成训练震荡，学习率太低则收敛速度慢。本文设计了自适应学习率调整机制，根据网络损失函数变化情况动态调节学习率：当损失函数下降快的时候适当降低学习率防止训练震荡；当损失函数趋于平稳或者上升时提高学习率加快参数更新。用衰减因子，训练轮次增加时减小学习率的改变程度，使算法后期稳定收敛。

3.3 优化控制策略实现流程

根据改进过的 AER-DQN 算法设计出一回路压力控制优化策略完整实现流程，分三个阶段，初始化、在线训练、实时控制，保证策略从训练到部署连贯性、可靠性。初始化阶段主要是对网络参数、经验池、控制参数进行设置。构建 AER-DQN 算法神经网络模型，输入层维度为状态空间维度（5 维），输出层维度为动作空间维度（3 维）；用 ReLU 作为激活函数，增加了非线性拟合能力；经验池容量初始设为 10000，优先级初始值为平均；初始学习率为 0.001，衰减因子为 0.99，折扣因子为 0.95（考虑即时奖励和未来奖励）；训练轮次设定为 5000 次。同时初始化一回路压力系统的初始状态，把压力设定值设为 15.5MPa，各个设备运行参数调整到额定工况。在线训练阶段用交互、存储、训练的循环模式来达到算法迭代优化的目的。智能体根据当前系统状态，根据网络得到的最优动作，对一回路压力系统（环境）施加影响；然后系统返回新的状态以及奖励值，将当前状态、动作、奖励、下一状态这四个要素保存到经验池里，更新经验

优先级；当经验池容量达到设定的阈值（2000条）的时候，开始网络训练，从经验池里按照优先级抽取批量数据（批量大小为32），用评估网络计算当前Q值，用目标网络计算目标Q值，构建损失函数，反向传播更新评估网络参数；最后每训练100轮更新一次目标网络参数，重复上述过程直到训练轮次结束，得到最优控制策略模型。对压力控制精度和算法收敛情况进行实时检测，如果出现训练震荡现象就调节学习率、折扣因子。实时控制阶段把训练好的最优策略模型部署到一回路压力控制系统上，达到动态自适应控制的目的。系统实时采集一回路压力、流量、温度等状态参数，经过归一化处理之后送入评价网络，评价网络输出最佳控制动作指令，驱动稳压器、喷淋装置等设备进行调节^[3]。

4 仿真实验与结果分析

4.1 仿真平台搭建

为了验证 AER-DQN 算法一回路压力控制优化方案的有效性，搭建仿真平台模拟核电站一回路压力系统运行特性，给对比实验提供一个可靠的测试环境。仿真平台以 MATLAB/Simulink 为工具，结合核电站一回路系统的数学模型，构建反应堆堆芯、稳压器、蒸汽发生器、主循环泵等核心设备的仿真模型，还原一回路压力系统的非线性特性、参数时变及扰动响应特性。仿真模型参数参照国内某压水堆核电站一回路实际运行参数，主循环泵额定流量为 1800m³/h，稳压器额定容积为 50m³，电加热器额定功率为 20MW，喷淋装置额定流量为 150m³/h，压力控制阈值设为 15.5±0.2MPa。在仿真平台上设置了三种典型的扰动，负荷扰动，反应堆负荷在 50%~100% 额定负荷范围内随机波动，设备参数扰动，主循环泵流量偏差 ±10%、蒸汽发生器传热系数偏差 ±8%，外部扰动，环境温度变化 ±5℃。

4.2 对比实验设计与实施

为了突出 AER-DQN 控制优化方案的优势，设计对比实验，选取传统的 PID 控制算法、常规 DQN 控制算法为对照组，本文提出的 AER-DQN 控制算法为实验组，从压力控制精度、动态响应性能、鲁棒性、能耗四个方面做量化比较。实验参数设置统一，三种算法的控制目标均为将一回路压力保持在 15.5±0.2MPa，仿真时间设为 1000s，前 200s 为稳态运行阶段，200s-800s 为扰动阶段（依次加入负荷扰动、设备参数扰动、外部扰动），800s-1000s 为恢复稳态阶

段。实验重复三次，取平均值作为最终的实验结果，防止随机误差影响。压力控制精度用稳态误差、超调量两个指标来衡量。从实验结果可以看出，实验组的稳态误差均值是 ±0.05MPa，超调量均值是 2.1%；常规 DQN 对照组的稳态误差均值是 ±0.12MPa，超调量均值是 4.3%；PID 对照组的稳态误差均值是 ±0.18MPa，超调量均值是 6.5%。可见实验组压力控制精度比两组对照组好得多，可以更好地控制一回路压力在设定范围内。动态响应性能用上升时间、调节时间这两个指标来衡量。上升时间即压力在受扰动以后，从最低值或者最高值恢复到设定值范围的时间，调节时间指压力稳定在设定值范围之内，波动幅度小于 0.03MPa 的时间。实验结果表明，实验组的平均上升时间是 8.2s，平均调节时间是 15.3s；常规 DQN 对照组的平均上升时间是 12.6s，平均调节时间是 22.8s；PID 对照组的平均上升时间是 18.9s，平均调节时间是 31.5s。说明实验组的动态响应速度更快，扰动发生之后能够很快地使压力稳定下来，减少压力波动给系统带来的影响。鲁棒性用扰动下压力最大偏差和恢复稳态时间这两个指标来度量。在三种扰动同时作用的时候，实验组的压力最大偏差是 0.18MPa，恢复稳态时间是 20.5s，常规 DQN 对照组的压力最大偏差是 0.32MPa，恢复稳态时间是 35.7s，PID 对照组的压力最大偏差是 0.45MPa，恢复稳态时间是 48.2s。说明实验组对于复杂的扰动有更强的鲁棒性，可以抑制扰动对压力的影响，保证系统的稳定运行。

5 结语

本文针对核电站一回路压力控制优化问题，在强化学习理论的基础上提出基于 AER-DQN 算法的控制方案，经过理论分析和仿真实验验证了方案的有效性。研究确定了强化学习在复杂非线性系统控制中的适用性，改进的算法克服了传统控制方法在扰动适应性、参数自适应方面的不足，从而能够给一回路压力控制提供了一条新的技术途径。

参考文献

- [1] 韩立杰,张智源,詹华斌,等.核电厂一回路计算机压力控制系统的研发与应用[J].自动化应用,2024,65(01):59-61+65.
- [2] 刘光伟,刘文杰,张庆国,等.“华龙一号”一回路水压试验压力系数的选取[J].核科学与工程,2021,41(01):174-180.
- [3] 吴迪,吴义兵,黄新明,等.核电站大修一回路不间断净化和树脂减容技术研究[J].工业水处理,2025,45(12):216-222.