基于强化学习的冶金热动系统多工况自适应控制策略

刘　凯
宝武集团武钢有限公司　湖北　武汉　430080

摘　要：文章聚焦基于强化学习的冶金热动系统多工况自适应控制策略。先分析冶金热动系统特性并构建混合模型，接着阐述强化学习理论基础，优选适用于该系统的改进Actor-Critic算法。随后设计控制策略，涵盖总体架构、状态与动作空间、奖励函数及多工况自适应机制。通过“感知-决策-执行-反馈”闭环架构，实现多工况精准控制与自适应优化，兼顾能效、参数稳定与设备安全，提升系统运行效率与稳定性，为冶金热动系统控制提供新思路。

关键词：强化学习；冶金热动系统；多工况；自适应控制

引言：冶金热动系统作为冶金生产核心能量供给单元，其稳定高效运行对冶金生产至关重要。然而，该系统具有多工况、非线性、高维耦合及动态变化等特性，传统控制方法难以满足复杂需求。强化学习凭借无需先验标签数据、动态适应环境变化的优势，为解决这一问题提供了新途径。本文深入分析冶金热动系统特性并建模，选择合适强化学习算法，设计多工况自适应控制策略，旨在实现系统在不同工况下的精准控制与自适应优化，提升整体运行性能。

1 冶金热动系统特性分析与建模

1.1 冶金热动系统组成与工作原理

冶金热动系统作为冶金生产流程中不可或缺的核心能量供给单元，在整个冶金工业体系中占据着举足轻重的地位。它由燃烧系统、传热系统、动力传输系统以及调控单元等关键部分紧密构成，涵盖了锅炉、换热器、风机、泵体等众多关键设备。这些组件相互配合、协同运作，共同实现了热能从产生到传递，再到高效利用的全过程。其工作原理以能量转化为根本核心。在系统中，燃料充分燃烧，释放出大量的热能，这些热能通过传热介质，像水、蒸汽等，被精准地传递至冶金工艺的各个环节，为冶金生产提供必需的热量支持。与此同时，动力设备持续运转，维持着系统内介质的稳定循环，确保温度、压力、流量等参数严格符合冶金生产的严苛要求。该系统具备多设备耦合、能量流动复杂等显著特点，各组件的运行状态紧密相连、相互影响。任一环节的参数出现波动，都极有可能像多米诺骨牌效应一样，影响整个系统的运行效率与稳定性[1]。所以，深入明确系统的组成结构以及详细的工作机制，是开展特性分析、进行科学控制设计的重要前提和坚实基础。

1.2 多工况特性分析

冶金热动系统的工况随冶金生产负荷、原料成分、工艺要求等因素动态变化，常见工况包括额定负荷工况、低负荷启停工况、变负荷调节工况及故障扰动工况等，不同工况下系统的运行参数与特性存在显著差异。额定负荷工况下，系统参数稳定，能量转化效率处于最优区间，各设备按设计参数协同运行；低负荷启停工况中，系统温度、压力变化剧烈，设备启停冲击大，易出现能量损耗增加、部件磨损加剧等问题；变负荷调节工况需快速响应负荷变化，参数动态调整过程中易产生超调与振荡；故障扰动工况则伴随设备故障、介质泄漏等异常，系统稳定性遭到破坏，需具备快速容错能力。通过多工况特性分析，可明确各工况下系统的参数变化规律、耦合关系及关键影响因素，为后续建模与控制策略设计提供数据支撑。

1.3 系统建模

冶金热动系统建模是实现精准控制与性能优化的关键，核心目标是构建能够准确反映系统输入、输出及内部状态变化规律的数学模型。考虑到系统多耦合、非线性、时变的特性，需结合机理分析与数据驱动方法开展建模工作。机理建模基于热力学、流体力学等基础理论，通过推导能量平衡方程、动量平衡方程等建立模型，能够体现系统内在运行机制，鲁棒性较强，但难以完全表征复杂耦合关系与非线性特性。数据驱动建模依托系统运行数据，采用回归分析、神经网络等算法构建模型，无需深入掌握机理，能快速适配工况变化，但对数据质量与数量要求较高。实际建模中需融合两种方法的优势，修正机理模型的偏差，弥补数据驱动模型的机理缺失，构建兼顾准确性与适应性的混合模型，为强化学习控制策略的设计与验证提供可靠载体。

2 强化学习理论基础与算法选择

2.1 强化学习基本概念与原理

强化学习作为一种极具创新性的机器学习方法，其

本质是基于试错学习来达成目标。其核心思想在于，智能体作为决策的主体，持续不断地与环境展开交互。在这个过程中，智能体首先感知环境当前所处的状态，然后依据自身的策略执行特定的动作。环境在接收到智能体的动作后，会做出相应的反馈，并给予智能体一个奖励信号，这个奖励信号可能是正奖励，也可能是负奖励[2]。智能体根据所获得的奖励信号，对自身的行为策略进行调整和优化。通过这样不断地循环迭代，智能体逐步优化目标函数，最终实现特定任务的最优决策。与监督学习和无监督学习不同，强化学习有着独特的优势。它无需依赖先验的标签数据，而是通过与环境的实时交互，在探索中不断寻找最优策略，具备强大的动态适应环境变化的能力。其核心原理严格遵循马尔可夫决策过程，该过程做出一个重要假设，即环境下一个状态的出现仅仅依赖于当前状态和所执行的动作，而无需考虑历史状态。这一假设为强化学习策略的优化提供了坚实的数学框架，使得强化学习特别适用于冶金热动系统这种多工况动态变化的复杂控制场景，能够有效地应对系统运行过程中的各种不确定性。

2.2 常用强化学习算法分析

常用强化学习算法主要可分为值函数类算法、策略梯度类算法以及Actor-Critic混合算法这三大类。这三类算法在收敛速度、稳定性、复杂度等关键方面各自展现出独特的优势，适用场景也存在明显差异。值函数类算法以Q-learning、SARSA为典型代表。这类算法的核心思路是通过学习状态-动作对的价值评估函数，以此来确定最优动作。其算法逻辑相对简单，易于实现，对于状态空间较小的简单任务，能够发挥出较好的效果。然而，当面对高维状态空间时，这类算法容易出现维度灾难问题，导致计算量急剧增加，收敛速度变得十分缓慢。策略梯度类算法，例如REINFORCE算法，直接对策略函数进行参数化建模，并采用梯度上升的方法进行优化。这类算法的优势在于能够处理连续动作空间问题，在高维场景中往往有着更出色的表现。但不足之处在于其方差较大，收敛过程不够稳定，容易受到系统参数波动的影响。Actor-Critic算法则巧妙地融合了前两类算法的优势。其中，Actor负责生成动作策略，Critic负责评估动作价值并指导Actor进行更新。这种协同机制使得该算法兼顾了收敛速度与稳定性，有效降低了方差，成为复杂系统控制中极具潜力的算法类型。

2.3 适用于冶金热动系统的强化学习算法选择

冶金热动系统具有多工况、非线性、高维耦合以及动态变化等显著特性，这就要求在选择强化学习算法时，必须综合考虑算法的适应性、稳定性、收敛速度以及处理连续动作空间的能力等多方面因素。值函数类算法由于存在维度灾难问题，在面对冶金热动系统高维状态空间时，难以进行有效的适配，仅仅适用于经过简化后的低维场景，无法满足实际复杂的控制需求。策略梯度类算法虽然具备处理连续动作空间的能力，但其在收敛稳定性方面表现欠佳，容易受到系统参数波动的影响，在应对多工况切换时的复杂动态变化时显得力不从心[3]。而Actor-Critic算法凭借Actor与Critic的协同工作机制，展现出了强大的优势。它既能够快速响应系统工况的变化，又能通过价值评估有效降低优化方差，显著提升收敛稳定性，同时还可以很好地处理温度、压力等连续动作空间的控制问题。进一步优选基于深度神经网络的DDPG、TD3等改进Actor-Critic算法，这些算法通过深度网络对值函数与策略函数进行拟合，极大地增强了对复杂系统的表征能力，能够精准适配冶金热动系统的多工况特性，为自适应控制策略提供了可靠且坚实的算法支撑。

3 基于强化学习的多工况自适应控制策略设计

3.1 控制策略总体架构

基于强化学习的冶金热动系统多工况自适应控制策略总体架构，采用了“感知-决策-执行-反馈”的闭环设计理念，精心划分为状态感知层、强化学习决策层、执行层以及反馈调节层这四大核心模块。这四大模块紧密协作，共同实现多工况下的精准控制与自适应优化。状态感知层作为整个系统的“眼睛”，通过各类高精度传感器，实时采集系统温度、压力、流量、燃料消耗等关键运行参数。这些原始数据经过预处理与特征提取后，转化为强化学习算法能够准确识别的状态信息。强化学习决策层以优化后的Actor-Critic算法为核心，结合多工况特征库，依据当前状态迅速生成最优控制动作指令。执行层则如同系统的“手脚”，通过调节阀、变频器等执行机构，将控制指令精准转化为实际操作，进而调整系统运行参数。反馈调节层实时采集执行后的系统状态，计算奖励信号并反馈至决策层，为策略的迭代更新提供有力指导。另外，架构还增设了工况识别模块，能够实时判定并切换工况，确保不同工况下控制策略的无缝适配，显著提升系统整体运行效率与稳定性。

3.2 状态空间与动作空间设计

状态空间设计对于全面表征冶金热动系统的运行状态至关重要，需兼顾完整性与简洁性，避免出现维度冗余或信息缺失的问题。为此，精心选取系统关键运行参数作为状态变量，涵盖燃烧温度、蒸汽压力、介质流

量、燃料供给量、风机转速以及工况标识等。为了消除量纲差异对算法的影响，将这些状态变量通过归一化处理映射至[0,1]区间，有效提升算法收敛速度。针对多工况特性，在状态空间中巧妙融入工况特征参数，实现不同工况下状态信息的差异化表征。动作空间设计则需紧密对应系统可调节变量，在冶金热动系统中，可调节动作主要包括燃料供给量调节、风机转速调节、调节阀开度调节等，均属于连续动作类型。因此，采用连续动作空间设计，并设置动作边界限制调节范围，防止超出设备安全运行阈值。动作空间的维度与状态空间相匹配，确保每个动作都能精准作用于系统状态。通过动作平滑处理，有效减少工况切换时的参数波动，进一步提升系统运行稳定性。

3.3 奖励函数设计

奖励函数作为强化学习算法的核心导向，必须紧密贴合冶金热动系统多工况运行目标，兼顾能效优化、参数稳定、设备安全等多维度需求。为此，设计了一种加权复合型奖励函数，确保策略优化方向与实际运行需求高度一致。该奖励函数由基础奖励、惩罚项及工况适配奖励三部分构成。基础奖励与系统能量转化效率紧密挂钩，根据实际能效与最优能效的偏差进行计算，偏差越小，奖励值越高，以此激励算法不断优化能效。惩罚项则针对参数超调、设备故障、能耗超标等异常情况设置，当参数超出安全范围或出现故障时，给予负奖励，有效约束控制动作的合理性。工况适配奖励则是针对多工况切换场景专门设计，当系统平稳完成工况切换、参数快速趋于稳定时，给予额外正奖励，显著提升策略的工况自适应能力。通过动态调整各部分权重，在不同工况下侧重不同优化目标，如低负荷工况侧重能耗控制，额定工况侧重参数稳定，确保奖励函数具备强大的适应性与导向性。

3.4 多工况自适应机制设计

多工况自适应机制的核心在于实现工况的精准识别、策略的动态切换与参数的实时优化，确保系统在不同工况下均能保持最优运行状态。工况识别模块基于支持向量机算法，结合系统关键参数特征，精心构建工况分类模型，能够实时判定当前运行工况，并将识别结果迅速反馈至强化学习决策层。策略切换机制采用多策略存储与自适应调用模式，针对不同工况预训练最优控制策略，并存储于策略库中[4]。当工况切换时，能够快速调用对应策略，同时启动在线微调流程，结合当前状态反馈优化策略参数，有效避免策略切换导致的参数波动。参数自适应优化机制通过动态调整奖励函数权重与算法超参数，精准适配不同工况的优化需求。同时，引入经验回放机制，存储不同工况下的交互数据，加速策略在线学习效率。另外，该机制还具备强大的容错能力，当出现工况识别误差或参数异常时，能够自动启动备用策略，切实保障系统稳定运行。

结束语

本文围绕基于强化学习的冶金热动系统多工况自适应控制策略展开研究，通过系统特性分析、算法选型以及控制策略的精心设计，构建了一套完整的解决方案。该策略有效解决了冶金热动系统多工况下的控制难题，实现了能效优化、参数稳定和设备安全的多目标协同。未来研究可进一步探索更先进的强化学习算法，优化控制策略细节，以更好地适应冶金热动系统日益复杂的运行需求，推动冶金行业向智能化、高效化方向发展。

参考文献

[1]戚鑫.冶金工业加热炉自动燃烧控制系统优化设计策略[J].商品与质量，2021(42)：114-115.

[2]胡江涛.智能电气自动化系统在冶金工业节能降耗中的实践路径[J].冶金与材料,2025,45(11):16-18.DOI:10.3969/j.issn.2096-4854.2025.11.007.

[3]祁永平,孙文君,汪全儒,等.冶金工业加热炉先进自动燃烧控制系统设计与应用研究[J].工业加热,2025,54(5):12-15.

[4]林建峰.冶金工业加热炉自动燃烧控制系统优化设计探究[J].自动化应用，2020(2)：34-35.