

控制与决策

Control and Decision

考虑需风量不确定性的矿井通风网络风量深度强化学习优化

吴亮红, 张艳云, 左词立, 王汐, 谭伟华

引用本文:

吴亮红, 张艳云, 左词立, 等. 考虑需风量不确定性的矿井通风网络风量深度强化学习优化[J]. *控制与决策*, 2026, 41(4): 955-964.

在线阅读 View online: <https://doi.org/10.13195/j.kzyjc.2025.0165>

您可能感兴趣的其他文章

Articles you may be interested in

[输入约束不确定系统的点对点迭代学习控制与优化](#)

Point-to-point iterative learning control and optimization for uncertain systems with constrained input
控制与决策. 2021, 36(6): 1435-1441 <https://doi.org/10.13195/j.kzyjc.2019.0908>

[基于MCPDDPG的智能车辆路径规划方法及应用](#)

The method and application of intelligent vehicle path planning based on MCPDDPG
控制与决策. 2021, 36(4): 835-846 <https://doi.org/10.13195/j.kzyjc.2019.0460>

[基于多班教学优化的多目标分布式混合流水车间调度](#)

Multi-class teaching-learning-based optimization for multi-objective distributed hybrid flow shop scheduling
控制与决策. 2021, 36(2): 303-313 <https://doi.org/10.13195/j.kzyjc.2020.0549>

[基于生成对抗网络的大规模路网交通流预测算法](#)

Traffic flow forecasting algorithm for large-scale road network based on GAN
控制与决策. 2021, 36(12): 2937-2945 <https://doi.org/10.13195/j.kzyjc.2020.0333>

[基于深度强化学习与迭代贪婪的流水车间调度优化](#)

Scheduling optimization for flow-shop based on deep reinforcement learning and iterative greedy method
控制与决策. 2021, 36(11): 2609-2617 <https://doi.org/10.13195/j.kzyjc.2020.0608>

考虑需风量不确定性的矿井通风网络风量深度强化学习优化

吴亮红^{1,2†}, 张艳云¹, 左词立¹, 王 汐¹, 谭伟华¹

(1. 湖南科技大学 信息与电气工程学院, 湖南 湘潭 411201;

2. 湖南工业职业技术学院 电气工程学院, 长沙 410217)

摘要: 矿井通风网络优化调节是矿井通风系统安全、稳定、经济运行的重要保障. 通风网络结构和状态参数随机动态变化给矿井通风网络优化求解和决策带来了极大的挑战. 充分考虑矿井通风系统的随机不确定性, 提出一种基于深度强化学习的矿井通风网络鲁棒优化调控方法. 首先, 对矿井通风网络风量优化问题进行数学描述, 将该风量优化问题建模为马尔可夫决策过程模型, 无需对系统不确定性进行建模和预测; 然后, 采用一种改进分布式近端策略优化算法对连续风量优化问题进行动态优化和决策, 能够直接得到不同需风量的优化调控方案. 实验结果表明, 所提出方法能够有效应对通风系统的多重不确定性, 降低矿井通风系统的风机能耗.

关键词: 矿井通风网络; 风量优化调控; 不确定性; 深度强化学习; 分布式近端策略优化

中图分类号: TD725 文献标志码: A

DOI: 10.13195/j.kzyjc.2025.0165

引用格式: 吴亮红, 张艳云, 左词立, 等. 考虑需风量不确定性的矿井通风网络风量深度强化学习优化 [J]. 控制与决策, 2026, 41(4): 955-964.

Deep reinforcement learning optimization of air volume in mine ventilation network considering uncertainty of air demand

WU Liang-hong^{1,2†}, ZHANG Yan-yun¹, ZUO Ci-li¹, WANG Xi¹, TAN Wei-hua¹

(1. School of Information and Electrical Engineering, Hunan University of Science and Technology, Xiangtan 411201, China; 2. School of Electrical Engineering, Hunan Industry Polytechnic, Changsha 410217, China)

Abstract: The optimization and regulation of mine ventilation networks are essential for the safe, stable, and economical operation of these systems. The random and dynamic changes in the structure and state parameters of the ventilation network present significant challenges for optimization and decision-making. This paper addresses the stochastic uncertainties inherent in mine ventilation systems and proposes a robust optimization and control method based on deep reinforcement learning. Initially, the airflow optimization problem is mathematically formulated as a Markov decision process, eliminating the need to model and predict system uncertainties. Subsequently, an improved distributed proximal policy optimization algorithm is employed to dynamically optimize and make decisions regarding continuous airflow, directly yielding optimized control solutions for varying airflow demands. Experimental results indicate that the proposed method effectively mitigates multiple uncertainties in the ventilation network and reduces the energy consumption of mine ventilation fans.

Keywords: mine ventilation network; air volume optimization and control; uncertainty; deep reinforcement learning; distributed proximal policy optimization

0 引言

矿井通风是保障矿井安全生产的重要手段^[1], 合理控制风流、满足各用风地点的按需供风, 才能保障井下生产安全. 矿井通风系统借助通风机向井下输送新鲜空气, 由于主通风机必须全年连续运行, 通风系统也是煤矿企业的主要耗能耗资设施, 通常矿井

通风能耗占采矿总能耗的 25% ~ 60%^[2], 占煤矿总成本的 15% ~ 20%^[3]. 因此, 实现通风网络风量优化对矿井通风系统平稳、经济、节能运行具有重要意义, 也是矿井智能化通风发展的需要^[4].

近年来, 众多学者应用智能优化算法来求解通风网络风量优化问题. 库向阳等^[5]将通风网络风量

收稿日期: 2025-02-18; 录用日期: 2025-04-11.

基金项目: 国家自然科学基金项目 (62373146); 湖南省科技人才托举工程项目中青年学者培养计划项目 (2022TJ-Q03); 国家自然科学基金青年科学基金项目 (52104192).

责任编辑: 王凌.

†通信作者. E-mail: lhwu@hnust.edu.cn.

优化问题分为风量最优分配和风流最优调控两个问题,并引入遗传算法进行求解;郭一楠等^[6]提出一种递阶文化算法求解加阻值总和最优的矿井通风网络优化问题,获得了最优的调风成本.吴新忠等^[7]提出一种收敛速度快、精度高的多种群粒子群算法求解通风网络风量优化问题,降低了通风系统的总功率;邵良杉等^[8]建立总功率最小的按需通风网络风量优化模型,在改进粒子群算法求解过程中陷入局部最优时,引入模拟退火算法引导粒子重新进行搜索.Sjöström等^[9]采用按需通风的优化控制策略,有效降低了风机能耗;宋佳林^[10]以实际矿井为对象,提出一种改进粒子群禁忌搜索算法(IPSO-TS)求解按需通风的风量优化问题,降低总能耗43.27%.这些算法在通风网络优化上取得了较大的成功,但都是基于确定性优化模型,没有考虑通风网络参数变化带来的不确定性.

矿井通风系统是一个复杂的非线性动态系统,随着环境、通风设施状态参数的改变,通风网络结构和状态参数随机动态变化^[2,11].这给通风网络优化求解和决策带来了极大的挑战,亟需发展智能化调控理论和方法解决不确定性问题.

近年来深度强化学习(DRL)^[12]方法迅速发展,并广泛应用于优化调度问题的求解^[13].Zhou等^[14]提出一种改进近端策略优化算法求解考虑源荷不确定性的热电联产系统经济调度问题,获得大量不同场景的调度策略,可以直接使用于有相同网络拓扑的经济调度;Zhang等^[15]采用深度强化学习求解考虑负荷需求不确定性的混合能源系统最优能源管理问题,提供了更好的策略,降低了14.17%的成本;杨潇等^[16]在遗传算法上引入近端策略优化对群智能感知系统的多目标任务指派问题进行优化求解;朱家政等^[17]针对模糊作业车间调度发生变化很难扩展的问题,提出一种LSTM-PPO(proximal policy optimization with long short-term memory)算法进行求解,实验验证能有效应对车间调度变化问题;张梅等^[18]提出改进双深度Q网络(DDQN)求解多人协同、多工种的混流装配平衡优化问题;彭自然等^[19]改进双延迟深度确定性策略梯度算法(TD3)提高电动汽车的能量回收效率,实现制动能量回收最优;王艳红等^[20]设计一种结合图网络框架和深度强化学习的作业车间调度算法.

为了解决通风网络结构和状态参数随机动态变化对通风网络优化调控的影响,本文考虑矿井通风系统的随机不确定性,提出一种基于深度强化学习的矿井通风网络鲁棒优化调控方法.同时,针对近端

策略优化算法探索环境有限的不足,引入多进程分布式策略,提高深度强化学习方法求解矿井通风网络风量动态优化问题的有效性和可靠性.

1 通风网络风量优化调控模型

1.1 目标函数

矿井通风网络风量优化的目标是实现按需分风,以保证系统平稳、经济运行.因此,建立以通风机功率最小化为目标的通风网络风量优化模型.假设通风网络表示为 $G = (V, E)$, V 和 E 分别是图 G 的节点集和分支集,并记 $|V| = m$, $|E| = n$,分别为网络 G 中的节点和分支数.

本文研究中,矿井通风网络的优化目标是通风机功率

$$W = \min H_f Q_f. \quad (1)$$

其中: H_f 是通风的风压, Q_f 是通风机的风量.

1.2 约束条件

通风网络三大定律是矿井通风网络解算的理论基础^[21],也是矿井通风网络优化问题的约束条件.

1) 风量平衡约束

风流达到稳定状态时,任意节点流入的风量与流出的风量代数和为0,即

$$\sum_{j=1}^n a_{lj} Q_j = 0, \quad l = 1, 2, \dots, m-1. \quad (2)$$

其中: Q_j 是分支 j 的风量; a_{lj} 是关联矩阵,有

$$a_{lj} = \begin{cases} 1, & \text{分支 } j \text{ 的风量流入节点 } l; \\ -1, & \text{分支 } j \text{ 的风量流出节点 } l; \\ 0, & \text{节点 } l \text{ 与分支 } j \text{ 不关联.} \end{cases} \quad (3)$$

2) 通风阻力约束

通风网络中,任意分支的风压等于该分支风阻与风量平方的乘积,即

$$H_j = R_j Q_j |Q_j|. \quad (4)$$

其中: H_j 是分支 j 的风压, R_j 是分支 j 的风阻.

3) 回路风压平衡约束

通风网络中,任意闭合回路中所有分支风压代数和等于风机风压与自然风压代数和,即

$$\sum_{j=1}^n c_{kj} R_j Q_j |Q_j| - P_k - F_k(Q_j) = 0, \quad k = 1, 2, \dots, n - m + 1. \quad (5)$$

其中: $F_k(Q_j) = c_1 Q_j^2 + c_2 Q_j + c_3$ 是通风机的特性曲线; c_1, c_2, c_3 是风机曲线拟合系数; $F_k(Q_j)$ 是回路 k 中的风机风压; P_k 是回路 k 中的自然风压; c_{jk} 是回路矩阵,有

$$c_{kj} = \begin{cases} 1, & j \in k, \text{且分支 } j \text{ 与回路 } k \text{ 同向;} \\ 0, & j \notin k; \\ -1, & j \in k, \text{且分支 } j \text{ 与回路 } k \text{ 反向.} \end{cases} \quad (6)$$

一般分风网络是部分分支风量已知, 需要求解其余未知分支风量. t 表示已知风量的分支数量, 调整回路矩阵 c_{sj} , 则通风网络中任意一条分支风量都可以表示为

$$Q_j = \sum_{s=1}^t c_{sj} Q_s + \sum_{s=t+1}^k c_{sj} Q_s. \quad (7)$$

其中: Q_s 表示风量已知分支风量, 部分已知分支风量为固定风量. 固定分支需风量不确定性受人员流动、采掘进度、瓦斯涌出等因素的影响, 即

$$Q_u = q \times k \times n. \quad (8)$$

其中: q 为基本需风量; n 为不确定性因子; $k = 1.1$ 表示风量备用系数. 在一定范围内随机改变不确定性因子 n 以体现需风量的不确定性.

考虑到通风机安全方面, 其工况点需运行在一定范围内, 否则通风机将发生喘振^[22]. 根据矿井通风机安全规程, 通风机的风压 $H_f \leq 0.9H_{f \max}$, 分支风量、风压应满足风量、风压上下限. 考虑优化目标和约束条件, 矿井通风网络优化模型如下所示:

$$\begin{aligned} W &= \min H_f Q_f; \\ \sum_{j=1}^n R_j \left(\sum_{s=1}^t c_{sj} Q_s + \sum_{s=t+1}^k c_{sj} Q_s \right)^2 - F_k(Q_j) &= 0, \\ k &= 1, 2, \dots, n - m + 1; \\ \sum_{j=1}^n a_{lj} \left(\sum_{s=1}^t c_{sj} Q_s + \sum_{s=t+1}^k c_{sj} Q_s \right) &= 0, \\ l &= 1, 2, \dots, m - 1; \\ H_f &\leq 0.9H_{f \max}; \\ Q_{j \min} &\leq Q_j \leq Q_{j \max}, \quad j = 1, 2, \dots, n; \\ H_{j \min} &\leq H_j \leq H_{j \max}, \quad j = 1, 2, \dots, n. \end{aligned} \quad (9)$$

其中: $Q_{j \min}$ 和 $Q_{j \max}$ 是分支 j 的风量调节上下限, $H_{j \min}$ 和 $H_{j \max}$ 是分支 j 的风压调节上下限. 若通风网络回路 k 中无通风机时, 则 $F_k(Q_j) = 0$.

2 基于 DPPO 的通风网络风量优化问题求解

2.1 马尔可夫决策过程

深度强化学习的基础是马尔可夫决策过程, 包括状态空间 (S)、动作空间 (A) 和奖励函数 (R). 矿井通风网络结构和状态参数的随机动态变化使得风量优化的求解具有较高的复杂性. 实际上, 矿井通风网络优化问题本质上是一个随机序贯决策问题, 而深度强化学习也适用于序贯决策场景. 因此, 利用深度强化学习方法来解决考虑需风量不确定性的矿井

通风网络风量优化动态调控问题具有显著的优势.

本文将矿井通风网络视为深度强化学习的环境, 智能体通过矿井通风网络的状态对分支风量进行最优调节决策, 状态空间包括通风网络的分支风量和回路余数分支的风压. 因此, 其状态空间可以定义为

$$S = [q_1, q_2, \dots, q_n, h_1, h_2, \dots, h_k]. \quad (10)$$

其中: n 为通风网络的分支数, k 为通风网络独立回路数.

动作空间为通风网络可调分支的风量变化量, 由通风网络独立回路数确定可调分支, 可调分支风量变化量为动作空间

$$A = [\Delta q_1, \Delta q_2, \dots, \Delta q_k], \quad (11)$$

其中 k 表示通风网络独立回路数.

矿井通风网络风量优化动态调控目标为通风机功率最小, 为此将该目标转化为深度强化学习奖励最大化问题. 此外, 智能体在作出决策时需确保各可调分支的风量符合需求. 为了引导智能体对环境进行有效探索, 对奖励函数进行改进, 将风量供需之间的不平衡作为惩罚项加入奖励函数. 通风网络风量优化的奖励函数为

$$r = -H_f Q_f - \|q\|^2 + I[\|q\|^2 < \varepsilon]. \quad (12)$$

其中: $I = I[\cdot]$ 是指标函数; ε 是最大累积失配; q 为需风量与实际风量的差值, $\|q\|^2$ 为风量供需失配惩罚量, 当通风网络达到一个风量平衡时, 将增加额外的奖励, 鼓励智能体将风量失配最小化.

2.2 深度强化学习方法

DRL 是基于智能体与环境之间的交互模型, 如图 1 所示. 智能体执行动作与环境进行交互, 旨在通过最大化累积奖励来学习优化策略^[23].

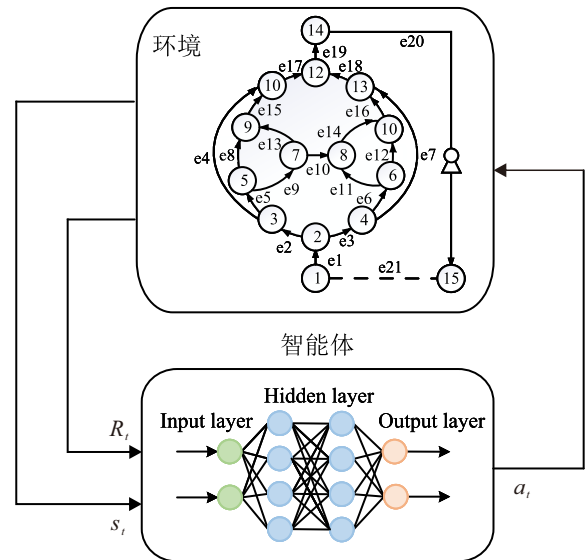


图1 深度强化学习

强化学习可以建模为一个马尔可夫决策过程,通常定义为四元组 (S, A, R, π) . 其中: S 为状态空间, A 为动作空间, R 为奖励函数, π 为策略. 在时间步 t , 智能体在状态 $s_t \in S$ 下, 根据策略 π 选择动作 $a_t \in A$, 并与环境交互, 随后得到奖励函数 r_t 和下一个状态 $s_{t+1} \in S$. 其中策略值 $\pi(a|s)$ 表示在状态 s 下选择动作 a 的概率. DRL 的目标是通过最大化未来奖励的期望 $Q(a|s)$ 来寻找最优策略

$$\pi^* = \arg \max Q_\pi(s, a), \quad (13)$$

其中

$$Q_\pi(s, a) = \mathbb{E} \left(\sum_{t=0}^K \gamma r_t | s_0 = s, a_0 = a, \pi \right). \quad (14)$$

$\gamma \in [0, 1]$ 为折扣因子, 表示未来奖励对当前时刻的重要程度; K 表示智能体与环境交互的总步数.

近端策略优化 (PPO)^[24] 算法是 DRL 的一种经典算法, 适用于求解连续空间优化问题. PPO 作为一种基于策略的方法, 采用 actor-critic 框架, 包含 actor 和 critic 两个网络. 其中: critic 网络用于估计值函数, 即在给定状态下采取某个动作获得的累积回报; actor 网络是策略网络, 基于 critic 网络提供的价值函数估计来评估动作的优劣, 并以此优化策略网络.

价值函数是在状态 s 下, 智能体根据策略 π 选择动作 a , 获得长期累积回报的期望值. 值函数定义如下:

$$V(s_t, \omega) = \sum_{a_t} \pi(a_t | s_t) R(s_t, a_t), \quad (15)$$

其中 $R(s_t, a_t)$ 表示长期累积回报的奖励函数. 对式(14)做蒙特卡洛近似, 可得

$$R(s_t, a_t) = r_t + \gamma \sum_{s_{t+1} \in S} V(s_{t+1}, \omega). \quad (16)$$

其中: r_t 为即时奖励, ω 为 critic 网络权重.

优势函数是奖励函数与值函数的差值, 通过广义优势估计计算优势函数

$$\hat{A}_t^{\text{GAE}} = \sum_{l=0}^{\infty} (\gamma \lambda)^l \delta_{t+l}^V. \quad (17)$$

其中: $\lambda \in [0, 1]$ 为衰减因子, 控制偏差与方差之间的权衡; δ_t^V 为时间差分误差, 有

$$\delta_t^V = r_t + \gamma V(s_{t+1}, \omega) - V(s_t, \omega). \quad (18)$$

在 PPO 算法中, actor 网络根据 \hat{A}_t^{GAE} 的大小对策略进行修正. 为防止新策略更新过大而引起剧烈变化, 引入概率比 $r_t(\theta)$ 来反应新旧策略的差异程度^[25], 并将该比率限制在预定范围内. 该方法不仅确保策略的更新方向是正确的, 而且提高了算法的稳定性和收敛性.

概率比 $r_t(\theta) = \pi_\theta(a_t | s_t) / \pi_{\theta_{\text{old}}}(a_t | s_t)$, actor 网络的损失函数为

$$L^{\text{clip}}(\theta) = \mathbb{E}[\min(r_t(\theta) \hat{A}_t^{\text{GAE}}, \text{clip}(r_t(\theta), 1 - \epsilon, 1 + \epsilon) \hat{A}_t^{\text{GAE}})]. \quad (19)$$

其中: θ 为 actor 网络的权重; ϵ 为超参数, 通过裁剪概率比将 actor 损失函数控制在范围内. 此外, 为了增强算法的探索能力, 引入策略商 $H(\pi(\cdot | s_t))$. 因此, 损失函数

$$L_{\text{ppo}}(\theta) = L^{\text{clip}}(\theta) - H(\pi(\cdot | s_t)) = L^{\text{clip}}(\theta) - \mathbb{E}_{a_t \sim \pi}[-\log(\pi(a_t | s_t))]. \quad (20)$$

critic 网络的损失函数为值函数与奖励函数的均方差, 即

$$L_{\text{cri}}(\omega) = \mathbb{E}(R(s, \omega) - V(s, \omega))^2. \quad (21)$$

actor 网络和 critic 网络的参数通过梯度方法进行更新. actor 网络通过梯度上升更新网络权重^[26], 即

$$\theta \leftarrow \theta + \alpha \nabla_{\theta} L_{\text{ppo}}(\theta), \quad (22)$$

其中 α 为 actor 网络的学习率.

critic 通过梯度下降更新网络权重^[26], 即

$$\omega \leftarrow \omega - \beta \nabla_{\omega} L_{\text{cri}}(\omega), \quad (23)$$

其中 β 为 critic 网络的学习率.

矿井通风网络的场景复杂多变, 为确保智能体在各种随机场景中获得良好的性能, 必须保证其能够探索更多的环境. 因此, 本文在 PPO 算法中引入多进程分布式设计, 同时执行多个进程以探索不同的环境, 并将采集的数据传递给主进程进行训练. 然而, 分布式近端策略优化 (DPPO) 算法容易陷入局部最优, 且表现出较大的波动. 为了防止 DPPO 算法陷入局部最优, 提高收敛速度并增强模型的稳定性, 本文对 DPPO 算法的学习率进行调整, 引入动态学习衰减策略优化模型训练过程, 即

$$\lambda_t = \lambda_0 \exp \left(k \ln \left(1 - \frac{t}{\tau + 1} \right) t \right). \quad (24)$$

其中: λ_t 为第 t 次权值更新时的学习率; λ_0 为初始学习率; $k = 0.001$, 决定衰减速率; τ 为总训练次数. 学习率在训练前期阶段保持一个较大的值, 下降速率较慢, 以保证算法的全局探索能力; 在训练的后期阶段, 学习率下降为 0, 以提高算法的局部开发能力. 通过动态调整学习率, 使 DPPO 算法达到全局探索和局部开发的平衡.

2.3 算法流程

在训练过程中, actor 和 critic 网络的输入均为状

态 s_t , actor 网络输出为动作 a_t , critic 网络输出为值函数 $V(s_t, \omega)$. 智能体与环境按时段顺序交互, 智能体获取环境状态 s_t 根据策略选择动作 a_t , 然后返回奖励 r_t 和下一个状态 s_{t+1} . 在此过程中, 经验池保存 $\{s_t, a_t, r_t, done, p_t\}$. 其中: done 为布尔值, 提示当前交互是否已结束, 若当前时刻交互完成, 则 done 为 True, 否则为 False; p_t 为当前时刻的输出动作的概率. 训

练后期保存期望奖励最大的 actor 和 critic 网络参数. 训练完成后, 利用保存的网络参数生成优化方案.

利用深度强化学习方法求解矿井通风网络风量优化动态调控问题的过程如图 2 所示. 引入分布式设置, 每一个进程的环境初始状态均不同, 以提高模型在应对需风量不确定性时的适用性. 训练过程中, 每轮训练的环境初始状态通过随机方式产生.

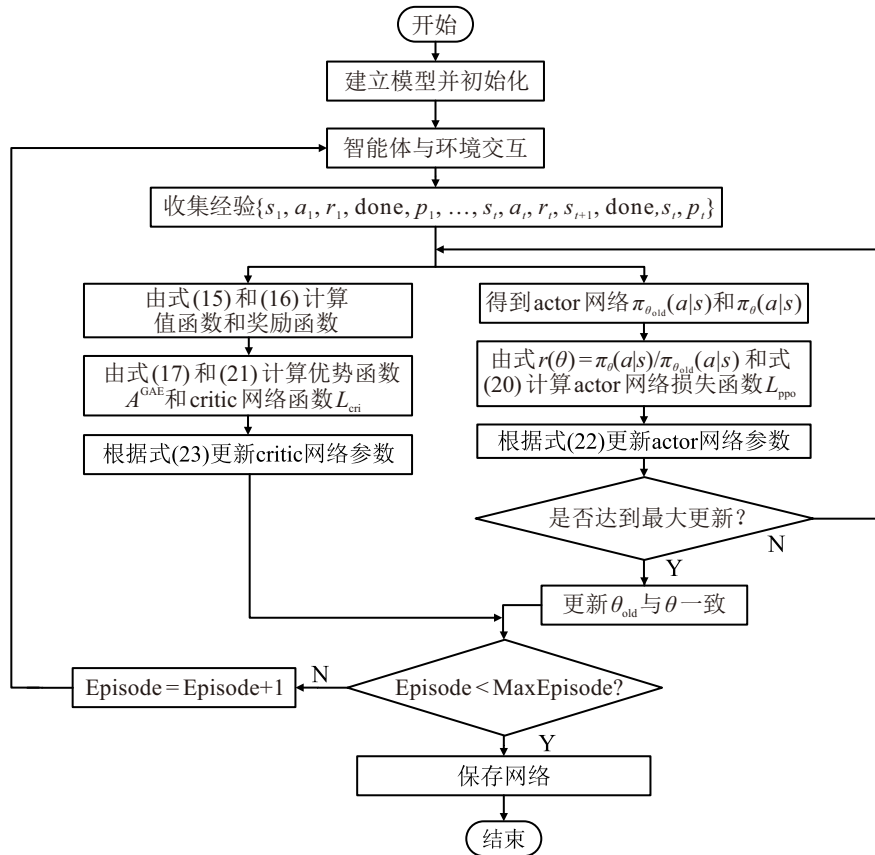


图2 DPPO 算法训练流程

本文所提出算法的流程如图 2 所示, 详细步骤如下.

step 1: 建立矿井通风网络模型, 并初始化训练环境和 actor-critic 网络参数.

step 2: 让智能体与环境交互, 并收集经验和训练数据.

step 3: 根据式 (15) 和 (16) 计算值函数和奖励函数, 获得 actor 网络旧参数 θ_{old} 和新参数 θ .

step 4: 根据式 (17) 计算优势函数, 根据式 (21) 计算 critic 网络的损失函数, 并完成自己的更新; actor 网络使用数据来计算变化率, 并在一定范围内对其进行剪裁.

step 5: 使用 critic 网络获得的剪裁概率比和式 (20) 计算 actor 网络的损失函数.

step 6: 根据式 (23) 更新 critic 网络参数.

step 7: 根据式 (22) 更新 actor 网络参数.

step 8: 判断是否达到最大更新, 如果达到, 更新 actor 网络参数; 否则转至 step 4.

step 9: 判断是否达到最大训练步数 Episode, 如果是, 则保存网络, 结束训练; 否则 Episode + 1 并转至 step 2.

3 算例研究

为验证本文所提出方法在随机环境下风量优化调控问题中的适用性与有效性, 在如图 3 所示的矿井通风网络上进行测试, 并与其他方法进行性能比较. MDP 与 DPPO 算法的建模使用了 Python 编程语言, 并利用 Pytorch 深度学习框架实现. 所有实验均在配备 Intel Core i5 第 12 代处理器和 RTX 3050 显卡的计算机上进行.

3.1 算例仿真设置

通风网络的数据来源是基于实际矿井通风系统

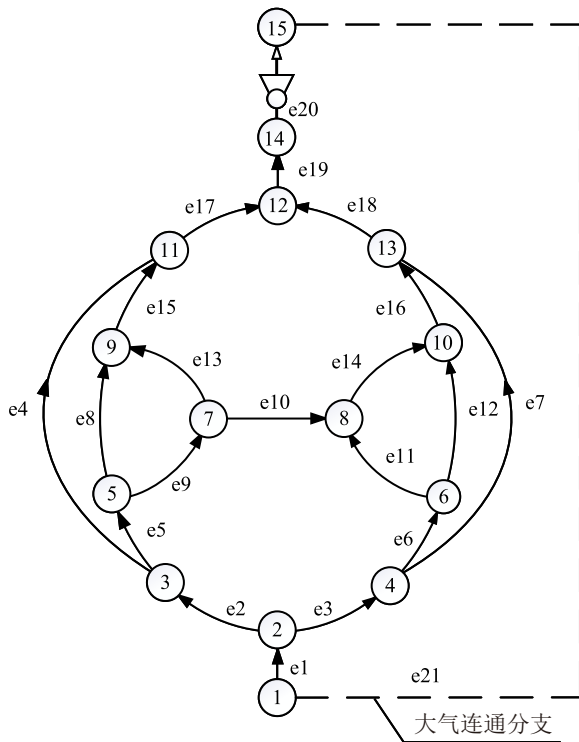


图3 矿井通风网络

拓扑构建的实验模型^[27],如图3所示.该通风网络包含15个节点,21条分支和7个独立回路.其中:分支e1为通风网络进风口,风机位于分支e20,分支e8和e12表示采煤工作面.表1为通风网络初始参数,包含初始风阻、风量.

分支e8和分支e12是固定风量分支,其风量需

表1 通风网络初始参数

分支编号	始节点	末节点	风阻/(N·s ² ·m ⁻⁸)	风量/(m ³ ·s ⁻¹)	分支类型
e1	1	2	0.46	46.00	一般分支
e2	2	3	0.12	30.17	一般分支
e3	2	4	0.21	15.83	一般分支
e4	3	11	1.08	8.00	一般分支
e5	3	5	0.04	22.17	一般分支
e6	4	6	0.20	9.18	一般分支
e7	4	13	1.16	6.65	一般分支
e8	5	9	1.38	7.44	固定风量分支
e9	5	7	0.35	14.73	一般分支
e10	7	8	0.65	13.59	一般分支
e11	8	6	0.33	6.33	一般分支
e12	6	10	0.32	15.52	固定风量分支
e13	7	9	0.34	1.14	一般分支
e14	8	10	1.21	7.26	一般分支
e15	9	11	0.14	8.58	一般分支
e16	10	13	0.21	22.78	一般分支
e17	11	12	0.30	16.58	一般分支
e18	13	12	0.07	29.43	一般分支
e19	12	14	0.13	46.00	一般分支
e20	14	15	0.73	46.00	一般分支
e21	15	1	0	46.00	大气联通分支

求会因采掘与运输活动和瓦斯气体涌出等因素而变化.通风网络^[28]中优化分支的风量和风压上下限如表2所示.

表2 优化分支风量和风压上下限

优化分支	风量/(m ³ ·s ⁻¹)		风压/Pa	
	下限	上限	下限	上限
e4	6.75	8.25	40.55	110.94
e7	6.55	8.03	31.75	111.55
e13	1.08	1.31	0.29	0.72
e14	7.02	8.58	36.47	100.12

在得到通风网络优化方案之前,对改进DPPO算法进行训练.为提高改进DPPO算法在处理不确定性方面的能力,使用多种场景进行环境探索训练.基于经验和文献^[28]选择改进DPPO的超参数配置,其中actor和critic网络均包含两个隐含层,每层含256个神经元,激活函数均为Tanh函数.最大训练次数为5000,由于引入多进程分布式设计,每个进程的训练集次数为5000,总训练集次数为25000.超参数设置详见表3.

表3 DPPO超参数

超参数	值
折扣系数 γ	0.95
初始学习率 α, β	3e-4
裁剪参数 ϵ	0.2
衰减因子 λ	0.95
mini batch	128
经验池容量 F_B	2 ¹⁴ /5

3.2 基于DPPO的优化结果分析

图4给出了模型训练过程中智能体奖励值随训练集次数的变化曲线.可以看出,PPO算法、DPPO算法以及改进DPPO算法在矿井通风网络风量优化中均表现出较好的有效性.随着训练集次数的增加,奖励值整体上升并趋于稳定,最终达到收敛状态.然而,在矿井通风网络环境中,PPO算法的奖励值明显低于DPPO算法,且收敛速度较慢.而DPPO算法的奖励值表现出较大波动,约在16000个Episode之

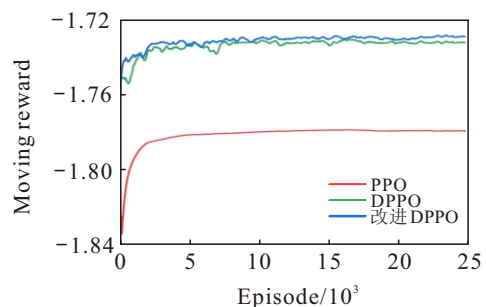


图4 训练期间奖励曲线

后开始收敛.相比之下,改进 DPPO 算法展现出较为平稳的奖励波动,并在约 11 000 个 Episode 之后开始逐渐收敛.

算法的收敛意味着智能体学会了通风网络风量优化策略.由于智能体初期对环境不够熟悉,最初获得的奖励值较低.此外,每个迭代过程的初始状态存在差异,导致了训练过程中奖励值的波动,这种波动反映了智能体在学习过程中对环境的逐步适应和优化决策能力的提高.训练完成后的 actor 和 critic 网络即可用于通风网络风量优化.

为验证本文方法的有效性和优越性,选择一个基于真实矿井通风网络构建的优化模型进行测试,并与粒子群 (PSO)^[6] 算法、反向增强型烟花算法 (OBEFWA)^[27]、PPO 算法、DPPO 算法以及改进 DPPO 算法的优化结果进行比较.所有实验均使用图 3 所示的通风网络,优化结果对比如表 4 所示.可以看出,通过对风量进行优化,本文算法在满足分支风量要求的基础上,尽可能减少分支 e20 通风量的大小,对比 PSO、OBEFWA、PPO、DPPO 算法,本文算法中分支 e20 风量最优,达到了通风机能耗优化的目的.

表4 不同方法的优化结果对比

分支	风量/(m ³ ·s ⁻¹)					
	自然分风	PSO	OBEFWA	PPO	DPPO	本文算法
e1	46.00	42.59	42.13	42.89	42.33	42.09
e2	30.17	28.17	28.46	28.71	28.29	28.30
e3	15.83	14.42	13.67	14.18	14.03	13.80
e4	8.00	6.67	6.76	6.75	6.75	6.75
e5	22.17	21.40	21.70	21.96	21.54	21.55
e6	9.18	7.78	7.11	7.63	7.48	7.25
e7	6.65	6.58	6.56	6.55	6.55	6.55
e8	7.44	6.89	6.87	6.87	6.87	6.87
e9	14.73	14.51	14.83	15.09	14.67	14.68
e10	13.59	13.33	13.74	14.01	13.59	13.60
e11	6.33	6.29	6.71	6.77	6.57	6.32
e12	15.52	14.13	13.82	14.39	14.06	13.57
e13	1.14	1.18	1.09	1.08	1.08	1.08
e14	7.26	7.04	7.03	7.24	7.02	7.28
e15	8.58	8.07	7.96	7.95	7.95	7.95
e16	22.78	21.17	20.85	21.64	21.08	20.85
e17	16.58	14.84	14.72	14.70	14.70	14.70
e18	29.43	27.75	27.41	28.19	27.63	27.40
e19	46.00	42.59	42.13	42.89	42.33	42.09
e20	46.00	42.59	42.13	42.89	42.33	42.09
e21	46.00	42.59	42.13	42.89	42.33	42.09

表 5 给出了不同算法优化方案的通风机能耗情况.本文方法在通风机功率方面表现最佳,具有最低的功率消耗和最高的节能率.具体而言,本文所提出优化方案的通风机功率最低值为 54.43 kW,相较于

PSO 的优化方案低了 1.97 kW,节能率提高 2.77%;与 OBEFWA 算法获得的方案相比,功率降低了 0.16 kW,节能率提高 0.22%;与 PPO 算法获得的方案相比,功率降低了 3.17 kW,节能率提高 4.46%;与 DPPO 算法获得的方案相比,功率降低了 0.94 kW,节能率提高 1.32%.显然,本文提出的改进 DPPO 算法所获得的优化方案优于其他方法,表明本文方法在提升通风系统经济性方面具有优势.

表5 通风能耗比较

优化方法	通风机输出风量/(m ³ ·s ⁻¹)	通风机输出风压/Pa	通风机总功率/kW	节能率/%
自然分风	46.00	1 544.68	71.06	
PSO	42.59	1 324.15	56.40	20.63
OBEFWA	42.13	1 259.70	54.59	23.18
PPO	42.89	1 342.87	57.60	18.94
DPPO	42.33	1 308.04	55.37	22.08
本文算法	42.09	1 293.24	54.43	23.40

表 6 给出了不同算法的计算效率.虽然深度强化学习算法训练时间较长,但是训练完成后,网络对任意需风量变化的响应时间为 15 ms 左右,而智能优化算法需要对每个新场景进行重新计算,难以满足在线鲁棒优化要求.

表6 不同算法的计算效率

算法	离线训练时间/h	在线推理时间/ms	适应新场景能力
PSO	—	2.4×10 ³	需重新计算
OBEFWA	—	2.4×10 ³	需重新计算
PPO	8.2	15	直接适用
DPPO	6.6	15	直接适用
本文算法	6.5	15	直接适用

按需分风是根据固定风量分支的风量需求,合理优化分支风量,实现风量的按需分配.本文以分支 e8 和 e12 为固定风量分支,假设其需风量均为 8 m³·s⁻¹,不同方法的按需分风的优化结果如表 7 所示.

表 7 为固定风量的矿井通风网络优化结果.可以看出,PPO、DPPO 以及改进 DPPO 均能实现固定风量优化调控,且风量均维持在 8 m³·s⁻¹.能耗结果如表 8 所示,改进 DPPO 算法的能耗对比固定风量自然分风节能率提高了 20.91%,比 PPO 算法节能率提高了 6.77%,比 DPPO 算法节能率提高了 4.71%.尽管 DRL 方法需要较长的训练时间,但其能够有效处理通风网络中的不确定性,且训练完成即可根据需风量的变化在线直接输出供风优化方案,无须重新对网络进行训练.

为验证该方法处理不确定性的能力,假设 e8 的需风量保持 8 m³·s⁻¹ 不变,改变 e12 的需风量.假设

表7 不同方法固定风量优化结果

分支	风量/(m ³ ·s ⁻¹)				最小需风量
	自然分风	PPO	DPPO	本文算法	
e1	46.00	43.72	43.37	42.54	41.93
e2	24.35	22.52	22.49	22.05	16.04
e3	21.65	21.94	20.88	20.49	14.26
e4	8.08	7.16	7.63	6.88	5.33
e5	16.27	15.26	14.87	15.17	10.71
e6	13.87	14.14	13.72	13.10	9.14
e7	7.78	7.05	7.16	7.39	5.11
e8	8.32	8.1	8.0	8.0	7.83
e9	7.95	7.26	6.87	7.17	5.23
e10	2.05	2.32	1.98	2.12	1.35
e11	5.86	6.14	5.72	5.10	3.86
e12	8.01	8.0	8.0	8.0	7.54
e13	5.90	4.94	4.89	5.05	3.88
e14	7.91	8.46	7.70	7.21	5.21
e15	14.22	13.04	12.89	13.05	9.36
e16	15.92	16.46	15.70	15.21	10.49
e17	22.30	20.21	20.52	19.94	14.70
e18	23.70	23.51	22.86	22.61	15.61
e19	46.00	43.72	43.37	42.54	41.93
e20	46.00	43.72	43.37	42.54	41.93
e21	46.00	43.72	43.37	42.54	41.93

表8 固定风量优化的能耗比较

优化方法	通风机输出风量/(m ³ ·s ⁻¹)	通风机输出风压/Pa	通风机总功率/kW	节能率/%
自然分风	46.00	1544.68	71.06	
PPO	43.72	1395.35	61.01	14.14
DPPO	42.56	1373.10	59.55	16.20
本文算法	42.54	1321.05	56.20	20.91

表9 e12 需风量随机波动的优化结果

分支	风量/(m ³ ·s ⁻¹)		
	Q _j = 8	Q _j = 8.5	Q _j = 9
e1	42.54	42.81	43.02
e2	22.05	22.25	21.98
e3	20.49	20.56	21.04
e4	6.88	7.10	6.88
e5	15.17	15.15	15.10
e6	13.10	13.51	13.93
e7	7.39	7.05	7.11
e8	8.00	8.00	8.10
e9	7.17	7.15	7.00
e10	2.12	2.15	2.09
e11	5.10	5.01	4.93
e12	8.00	8.50	9.00
e13	5.05	5.0	4.91
e14	7.21	7.17	7.02
e15	13.05	13.00	13.01
e16	15.21	15.67	16.02
e17	19.94	20.10	19.89
e18	22.61	22.72	23.13
e19	42.54	42.81	43.02
e20	42.54	42.81	43.02
e21	42.54	42.81	43.02

e12 的需风量不确定, 需风量在 [8, 9.05] m³·s⁻¹ 内波动. 在训练过程中采用随机数生成的方法产生该范围的大量随机变量, 以模拟各种不确定性场景供网络学习. 表 9 给出了 e12 需风量分别为 8 m³·s⁻¹、8.5 m³·s⁻¹、9 m³·s⁻¹ 三种情况的网络输出结果.

由表 9 可以看出, 随着分支 e12 需风量的变化, 深度强化学习网络能够在线输出对应的优化方案, 有效满足分支 e12 和通风网络中其他分支风量要求. 与传统模型预测方法不同, 深度强化学习方法通过大量数据训练, 使智能体在各种不同环境中进行交互, 从而生成一个适应众多环境的优化决策网络. 这种方法不仅提高了模型的泛化能力, 还增强了预测的准确性, 能够有效应对通风系统中的不确定性问题. 因此, 基于深度强化学习的方法能够即时响应需风量的动态变化, 并提供经济可行的最优调控方案.

4 结论

为了解决通风网络状态参数随机动态变化给矿井通风网络优化求解带来的挑战, 本文考虑了矿井通风系统的随机不确定性, 建立了矿井通风网络动态优化调控模型, 并引入深度强化学习方法求解通风网络动态调控问题. 本文所提出方法无需对需风量不确定性进行预测. 针对通风网络动态优化调控问题, 在 PPO 算法基础上引入分布式设计和衰减学习率进行改进, 改进 DPPO 算法提高了深度强化学习网络的收敛性和稳定性. 针对通风网络不同需风量的场景, 在满足需风量的同时, 实现通风机节能降耗. 通过通风网络历史数据进行仿真实验, 结果表明, 相较于启发式算法和 PPO 算法, 使用改进 DPPO 算法的优化调控方案能够实现按需分风, 同时降低主通风机的功耗, 提高了通风系统的节能率.

参考文献 (References)

- [1] 周福宝, 辛海会, 魏连江, 等. 矿井智能通风理论与技术研究进展[J]. 煤炭科学技术, 2023, 51(1): 313-328. (Zhou F B, Xin H H, Wei L J, et al. Research progress of mine intelligent ventilation theory and technology[J]. Coal Science and Technology, 2023, 51(1): 313-328.)
- [2] Li J Q, Li Y C, Zhang J, et al. Accurate and real-time network calculation for mine ventilation without wind resistance measurement[J]. *Journal of Wind Engineering and Industrial Aerodynamics*, 2022, 230: 105183.
- [3] Nel A J H, Arndt D C, Vosloo J C, et al. Achieving energy efficiency with medium voltage variable speed drives for ventilation-on-demand in South African mines[J]. *Journal of Cleaner Production*, 2019, 232: 379-390.
- [4] 张庆华, 姚亚虎, 赵吉玉. 我国矿井通风技术现状及智能化发展展望[J]. 煤炭科学技术, 2020, 48(2): 97-103.

- (Zhang Q H, Yao Y H, Zhao J Y. Status of mine ventilation technology in China and prospects for intelligent development[J]. *Coal Science and Technology*, 2020, 48(2): 97-103.)
- [5] 库向阳,常新坦,孙艺珍.基于遗传算法的通风网络两步法风流调节优化算法[J].*中南大学学报:自然科学版*, 2011, 42(9): 2729-2736.
(She X Y, Chang X T, Sun Y Z. Optimization algorithm adjusting ventilation flow about Min ventilation networks based on two step way and genetic algorithm[J]. *Journal of Central South University: Science and Technology*, 2011, 42(9): 2729-2736.)
- [6] 郭一楠,程健,陈美蓉,等.矿井通风网络的递阶文化优化方法[J].*控制与决策*, 2012, 27(10): 1542-1546.
(Guo Y N, Cheng J, Chen M R, et al. Mine ventilation network optimization adopting hierarchy cultural algorithm[J]. *Control and Decision*, 2012, 27(10): 1542-1546.)
- [7] 吴新忠,张兆龙,程健维,等.矿井通风网络的多种群自适应粒子群算法优化研究[J].*煤炭工程*, 2019, 51(2): 75-81.
(Wu X Z, Zhang Z L, Cheng J W, et al. Optimization of multi-group self-adaptive particle swarm algorithm for mine ventilation network[J]. *Coal Engineering*, 2019, 51(2): 75-81.)
- [8] 邵良杉,王振,李昌明.基于模拟退火与改进粒子群的矿井通风优化算法[J].*系统仿真学报*, 2021, 33(9): 2085-2094.
(Shao L S, Wang Z, Li C M. Optimization algorithm of mine ventilation based on SA-IPSO[J]. *Journal of System Simulation*, 2021, 33(9): 2085-2094.)
- [9] Sjöström S, Klintenäs E, Johansson P, et al. Optimized model-based control of main mine ventilation air flows with minimized energy consumption[J]. *International Journal of Mining Science and Technology*, 2020, 30(4): 533-539.
- [10] 宋佳林. PSO-TS 算法在矿井通风网络风量优化中的应用研究[J]. *矿业安全与环保*, 2022, 9(2): 8-82.
(Song J L. Application of IPSO-TS algorithm in air volume optimization of mine ventilation network[J]. *Mining Safety & Environmental Protection*, 2022, 9(2): 8-82.)
- [11] Walentek A, Janoszek T, Prusek S, et al. Influence of longwall gateroad convergence on the process of mine ventilation network-model tests[J]. *International Journal of Mining Science and Technology*, 2019, 29(4): 585-590.
- [12] Li Y X. Deep reinforcement learning: An overview[J]. 2017, arXiv: 1701.07274.
- [13] Waubert de Puiseau C, Meyes R, Meisen T. On reliability of reinforcement learning based production scheduling systems: A comparative survey[J]. *Journal of Intelligent Manufacturing*, 2022, 33(4): 911-927.
- [14] Zhou S Y, Hu Z J, Gu W, et al. Combined heat and power system intelligent economic dispatch: A deep reinforcement learning approach[J]. *International Journal of Electrical Power & Energy Systems*, 2020, 120: 106016.
- [15] Zhang G Z, Hu W H, Cao D, et al. Data-driven optimal energy management for a wind-solar-diesel-battery-reverse osmosis hybrid energy system using a deep reinforcement learning approach[J]. *Energy Conversion and Management*, 2021, 227: 113608.
- [16] 杨潇,郭一楠,吉建娇,等.异构群智感知 PPO 多目标任务指派方法[J].*控制理论与应用*, 2024, 41(6): 1056-1066.
(Yang X, Guo Y N, Ji J J, et al. PPO multi-objective task allocation method for heterogeneous crowd sensing[J]. *Control Theory & Applications*, 2024, 41(6): 1056-1066.)
- [17] 朱家政,张宏立,王聪,等.基于深度强化学习的模糊作业车间调度问题[J].*控制与决策*, 2024, 39(2): 595-603.
(Zhu J Z, Zhang H L, Wang C, et al. Fuzzy job shop scheduling problem based on deep reinforcement learning[J]. *Control and Decision*, 2024, 39(2): 595-603.)
- [18] 张梅,田镇遇,朱金辉,等.基于深度强化学习的多人协同混流装配线平衡优化研究[J].*控制与决策*, 2024, 39(10): 3395-3404.
(Zhang M, Tian Z Y, Zhu J H, et al. Multi-manned collaborative mixed-model assembly line balancing optimization based on deep reinforcement learning[J]. *Control and Decision*, 2024, 39(10): 3395-3404.)
- [19] 彭自然,贺振宇,肖仲平,等.基于深度强化学习模型 TD3 优化和改进的电动汽车制动能量回收策略研究[J].*控制与决策*, 2025, 40(8): 2361-2372.
(Peng Z R, He Z Y, Xiao S P, et al. Research on electric vehicle brake energy recovery strategy based on deep reinforcement learning model TD3 optimization and improvement[J]. *Control and Decision*, 2025, 40(8): 2361-2372.)
- [20] 王艳红,付威通,张俊,等.基于改进近端策略优化算法的柔性作业车间调度[J].*控制与决策*, 2025, 40(6): 1883-1891.
(Wang Y H, Fu W T, Zhang J, et al. Flexible job-shop scheduling based on improved proximal policy optimization algorithm[J]. *Control and Decision*, 2025, 40(6): 1883-1891.)
- [21] 宿敬肖,江春然,吕晓华,等.基于 GA-BP 神经网络的矿井节能通风控制系统设计[J].*煤炭技术*, 2022, 41(12): 232-234.
(Su J X, Jiang C R, Lyu X H, et al. Design of mine energy-saving ventilation control system based on GA-BP neural network[J]. *Coal Technology*, 2022, 41(12): 232-234.)
- [22] 周敏.煤矿主通风机运行方式的研究与应用探讨[J].*机械管理开发*, 2018, 33(3): 176-177.
(Zhou M. Research and Application of Main Fan Operation Mode in Coal Mine[J]. *Mechanical*

- Management and Development, 2018, 33(3): 176-177.)
- [23] Yang T, Zhao L Y, Li W, et al. Dynamic energy dispatch strategy for integrated energy system based on improved deep reinforcement learning[J]. *Energy*, 2021, 235: 121377.
- [24] Schulman J, Wolski F, Dhariwal P, et al. Proximal policy optimization algorithms[J/OL]. 2017, arXiv: 1707.06347.
- [25] Li Y, Bu F J, Li Y Z, et al. Optimal scheduling of island integrated energy systems considering multi-uncertainties and hydrothermal simultaneous transmission: A deep reinforcement learning approach[J]. *Applied Energy*, 2023, 333: 120540.
- [26] 蔺伟山, 王小君, 孙庆凯, 等. 不确定性环境下基于深度强化学习的综合能源系统动态调度[J]. *电力系统保护与控制*, 2022, 50(18): 50-60.
(Lin W S, Wang X J, Sun Q K, et al. Dynamic dispatch of an integrated energy system based on deep reinforcement learning in an uncertain environment[J]. *Power System Protection and Control*, 2022, 50(18): 50-60.)
- [27] 吴新忠, 胡建豪, 魏连江, 等. 矿井通风网络的反向增强型烟花算法优化研究[J]. *工矿自动化*, 2019, 45(10): 17-22.
(Wu X Z, Hu J H, Wei L J, et al. Research on opposition-based enhanced fireworks algorithm optimization for mine ventilation network[J]. *Industry and Mine Automation*, 2019, 45(10): 17-22.)
- [28] Engstrom L, Ilyas A, Santurkar S, et al. Implementation matters in deep policy gradients: A case study on PPO and TRPO[J/OL]. 2020, arXiv: 2005.12729.

作者简介

吴亮红 (1977-), 男, 教授, 博士, 主要研究方向为智能优化算法、能源系统优化调度, E-mail: lhwu@hnust.edu.cn;

张艳云 (2000-), 女, 硕士生, 主要研究方向为先进控制理论与方法, E-mail: 2056326557@qq.com;

左词立 (1989-), 男, 博士, 主要研究方向为深度学习和智能优化, E-mail: cilizuo@163.com;

王汐 (1989-), 男, 博士, 主要研究方向为无人机群智能协同控制策略、无人机相关行业应用, E-mail: wangxi@hnust.edu.cn;

谭伟华 (1994-), 男, 博士, 主要研究方向为离散制造过程建模分析、工业智能优化, E-mail: tanwh@hnust.edu.cn.