

基于近端策略优化算法的微网优化调度

马跃锋¹, 招苏硕¹, 王凯²

(1. 国网内蒙古东部电力有限公司赤峰供电公司, 内蒙古 赤峰 024000;

2. 华南理工大学 电力学院, 广东 广州 510640)

摘要:微网是实现大量分布式电源集成到电网中的有效手段。针对微网优化调度问题,提出了一种基于近端策略优化算法的优化调度方法。首先,综合考虑微网运行成本和各类设备运行约束构建微网优化调度模型;其次,将该问题表述为强化学习框架,设计了强化学习状态、动作和奖励函数等要素;最后,设计了基于近端策略优化算法的求解流程,通过仿真验证了所提方法的有效性。

关键词:微电网;优化调度;人工智能;强化学习

中图分类号:TM28 **文献标识码:**A **DOI:**10.19457/j.1001-2095.dqed24592

Optimal Dispatch of Microgrid Based on Proximal Policy Optimization

MA Yuefeng¹, ZHAO Suqi¹, WANG Kai²

(1. Chifeng Power Supply Company of State Grid East Inner Mongolia Power Co., Ltd., Chifeng 024000,

Nei Monggol, China; 2. School of Electric Power Engineering, South China

University of Technology, Guangzhou 510640, Guangdong, China)

Abstract: Microgrid is an effective method to integrate a large number of distributed generators into the power grid. Aiming at the optimal dispatch problem of microgrid, an optimal dispatch method based on proximal policy optimization algorithm was proposed. Firstly, the optimal dispatch model of microgrid was constructed by considering the operation cost of microgrid and operation constraints of various equipment. Secondly, the problem was formulated as a reinforcement learning framework, and the elements of reinforcement learning such as state, action and reward function were designed. Finally, the solution flow based on the proximal policy optimization algorithm was designed, and the effectiveness of the proposed method was verified by simulation.

Key words: microgrid; optimal dispatch; artificial intelligence; reinforcement learning

分布式发电具有明显的随机性、波动性,近年来大规模分布式电源接入电网,给电网的调度带来了极大的挑战^[1-2]。微电网是各种分布式电源、储能装置、负荷和能量转换设备整合而成的小型发配电系统,既可独立运行,也可并网运行作为大电网的有益补充^[3]。微网因其环境友好性、运行方式灵活等优点得到了广泛的研究。文献[4]以多微网系统为领导者、各微网负荷聚合商为跟随者建立一主多从的主从博弈优化调度模型,采用混沌粒子群算法嵌套gurobi求解器求解以上模型;文献[5]以微网总运行成本最小为目标建立优化调度模型,采用基于天牛群搜索算法的改进智能算法对模型进行求解;文献[6]采用协同

进化算法求解微网经济调度模型,有效降低了微网运行成本。上述方法均为启发式算法,易陷入局部最优解。

随着人工智能技术的快速发展,基于数据驱动的强化学习方法为微网优化调度提供了新的求解思路^[7]。强化学习中智能体以试错的方式对环境不断探索,以整个训练回合奖励最大化为目标逐渐学习得到最优策略,这与微网优化调度以整个调度周期运行成本最低为目标是契合的^[8]。文献[9-11]采用Q学习方法求解优化调度策略,但需将各电源功率离散化,离散化不仅会带来调度误差,而且随着变量的增多易出现“维数灾难”。文献[12-13]均采用深度强化学习方法进行

基金项目:国家自然科学基金(52077083)

作者简介:马跃锋(1985—),男,本科,工程师,研究方向为电力系统继电保护,Email:82765829@qq.com

优化调度,采用神经网络进行决策,智能体通过更新神经网络参数的形式探索得到最优调度决策。近端策略优化算法(proximal policy optimization, PPO)^[14]是一种典型的深度强化学习方法,具有收敛速度快、数据利用率高等优点。文献[15]提出基于PPO算法的家庭实时能量管理策略,在最优居民日用电成本的同时提升了对不确定性光伏发电的适应性;文献[16]利用PPO算法有效降低了储能系统运行成本。

本文针对包含光伏、电储能、燃气轮机、负荷的典型微网系统的日前优化调度问题,首先以日运行成本最低为目标构建调度模型;然后采用PPO算法进行日前优化调度策略求解,并对仿真结果进行了分析;最后通过与粒子群算法进行对比验证了所提方法的有效性。

1 微网优化调度模型

1.1 目标函数

本文所考虑的微网系统包括燃气轮机、电储能、光伏、负荷,微网优化目标为日运行成本最低,如下式所示:

$$\min \sum_{t=1}^T [C_{MT}(t) + C_{ES}(t) + C_{buy}(t)] \quad (1)$$

式中: T 为调度周期时段数; $C_{MT}(t)$, $C_{ES}(t)$ 分别为微燃气轮机运行成本、电储能运行成本; $C_{buy}(t)$ 为外网购电成本。

微燃气轮机运行成本与发电功率呈线性关系:

$$C_{MT}(t) = \alpha P_{MT}(t) \quad (2)$$

式中: α 为成本系数; $P_{MT}(t)$ 为 t 时刻发电功率。

对于电储能考虑其度电成本,度电成本表达式为

$$C_{ES}(t) = \rho |P_{ES}(t)| \quad (3)$$

式中: ρ 为度电成本系数; $P_{ES}(t)$ 为 t 时刻储能充/放电功率,为正时处于放电状态,为负时处于充电状态。

微网系统从主网的购电成本如下式所示:

$$C_{buy}(t) = \begin{cases} \lambda_{buy}(t) \cdot P_{buy}(t) & P_{buy}(t) > 0 \\ \lambda_{sell}(t) \cdot P_{buy}(t) & P_{buy}(t) < 0 \end{cases} \quad (4)$$

式中: $P_{buy}(t)$ 为 t 时刻从主网的购电成本; $\lambda_{buy}(t)$, $\lambda_{sell}(t)$ 分别为 t 时刻主网购电、售电价格。

1.2 约束条件

微网的约束包括功率平衡约束和设备运行约束。

功率平衡约束。在时段 t 微网内电能供应与

负荷须达到平衡,如下式所示:

$$P_{PV}(t) + P_{MT}(t) + P_{ES}(t) + P_{buy}(t) = P_{load}(t) \quad (5)$$

式中: $P_{PV}(t)$, $P_{load}(t)$ 分别为 t 时段微网内光伏、总负荷功率。

设备运行约束。微网中各设备均有运行上、下限范围,燃气轮机运行约束如下式:

$$P_{MT}^{\min} \leq P_{MT}(t) \leq P_{MT}^{\max} \quad (6)$$

$$-R_{down} \leq P_{MT}(t) - P_{MT}(t-1) \leq R_{up} \quad (7)$$

式中: P_{MT}^{\max} , P_{MT}^{\min} 分别为燃气轮机输出功率的上、下限; R_{up} , R_{down} 分别为燃气轮机最大向上、向下爬坡功率。

电储能运行约束如下式所示:

$$-P_{dis}^{\max} \leq P_{ES}(t) \leq P_{ch}^{\max} \quad (8)$$

式中: P_{ch}^{\max} , P_{dis}^{\max} 分别为最大充、放电功率。

为避免过度充放电对电储能设备造成的损害,需将电储能荷电状态 $SOC(t)$ 限制在一定范围内:

$$SOC^{\min} \leq SOC(t) \leq SOC^{\max} \quad (9)$$

其中 $SOC(t+1) = SOC(t) + \int P_{ES}(t)dt \times \eta/E$

式中: SOC^{\max} , SOC^{\min} 分别为最大、最小荷电状态; E 为电储能容量; η 为电储能的充放电效率。

2 微网优化调度强化学习框架

微网优化调度的强化学习框架如图1所示,其中智能体由深度神经网络构成,具有调度决策能力。在调度时刻 t ,具有决策能力的智能体根据微网系统状态 s ,探索性做出调度决策 a_t ,并获得下一时刻微网系统给出的奖励 r_t ,智能体进入下一时刻状态。智能体不断对微网系统探索,根据经验调整调度决策,使一个调度周期累积奖励之和最大化。

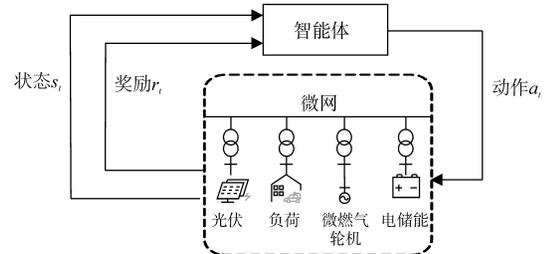


图1 微网优化调度的强化学习框架

Fig.1 Reinforcement learning framework of microgrid optimal dispatch

对图1中所示状态、动作和奖励函数等强化学习要素设计如下:

1)状态 s_t 。 s_t 为 t 时刻智能体接收到的微网系统运行状态,包括光伏发电功率、负荷功率、上一时刻微燃气轮机功率和电储能荷电状态,状态 s_t 可表示如下:

$$s_t = \{P_{PV}(t), P_{load}(t), P_{MT}(t-1), SOC(t)\} \quad (10)$$

2)动作 a_t 。微网优化调度问题的目标是最优地确定其微燃气轮机和电储能功率,动作 a_t 即为决策变量,可表示为

$$a_t = \{P_{MT}(t), P_{ES}(t)\} \quad (11)$$

3)奖励函数 r_t 。微网优化问题优化目标为最小化整个调度周期运行成本,而强化学习以整个调度周期获得的奖励最大化为更新方向。因此,智能体在时刻 t 获得的奖励可表示为该时刻负的运行成本:

$$r_t = -[C_{MT}(t) + C_{ES}(t) + C_{buy}(t)] + r_0 \quad (12)$$

式中: r_0 为偏置常数,使奖励函数由负转正,加快收敛速度。

智能体的最终目标为寻找一策略 π^* ,使其获得的累积奖励最大:

$$\pi^* = \arg \max_{\pi} E\left(\sum_{t=0}^{\infty} \gamma^t r_t\right) \quad (13)$$

其中, γ 为折扣因子,表示当前所选动作对未来收益的影响程度, $\gamma \in [0,1]$ 。由于本文的优化目标为最小化整个调度周期的运行成本,因此 γ 取值为1。

3 强化学习求解流程

针对上述优化调度模型,设计其基于PPO算法的求解流程如图2所示。该方法中智能体不断对微网优化调度策略进行探索,以调度成本最优为目标指导神经网络参数更新,直到找到最优策略。具体流程如下:

1)随机初始化智能体神经网络参数 θ ,包括如图3所示的神经网络中每个神经元的权重系数和偏置项;

2)初始化环境,即确定式(10)中各变量的初始值;

3)在调度时刻 t ,智能体神经网络输入 s_t ,输出调度决策 a_t ,并获得下一时刻状态 s_{t+1} 和奖励 r_t 。判断 t 是否达到调度周期时段数,如果没有则继续该步骤,否则进入神经网络更新;

4)更新智能体神经网络参数 θ 。PPO算法除采用参数为 θ 的策略网络进行调度决策外,还采用价值网络对策略网络所采取策略进行评估,以

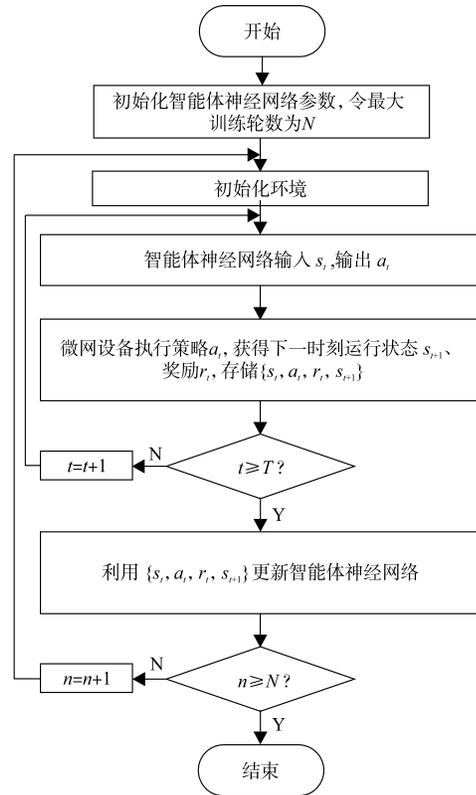


图2 基于PPO算法的求解流程

Fig.2 Solution flow based on PPO

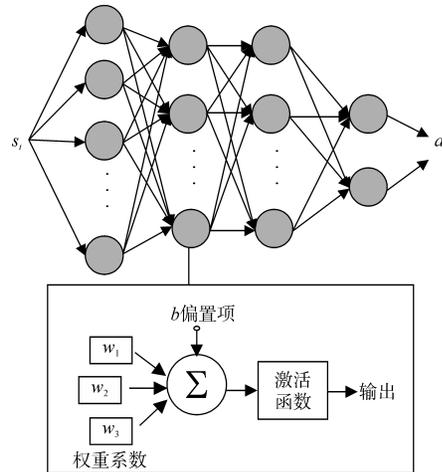


图3 神经网络及神经元数学模型

Fig.3 Mathematical model of neuron network and neuron

指导策略网络更新。价值网络输入 s_t ,输出价值函数 $V(s_t)$, $V(s_t)$ 表示当前状态下可获取的累积奖励的期望,即

$$V(s_t) = r_t + \gamma r_{t+1} + \dots + \gamma^{T-t} r_T \quad (14)$$

式中: T 为决策序列的长度。

为更准确地对 $V(s_t)$ 进行估计,价值网络的更新目标为最小化时序差分误差:

$$\min [\gamma V(s_{t+1}) + r_t - V(s_t)] \quad (15)$$

第 n 次迭代时策略网络以梯度下降法进行更新,更新的目标函数为

$$L(\theta) = E_t \{ \min \{ r_t(\theta) A_\pi(s_t, a_t), \text{clip}[r_t(\theta), 1 - \varepsilon, 1 + \varepsilon] A_\pi(s_t, a_t) \} \} \quad (16)$$

其中

$$A_\pi(s_t, a_t) = r_t + \gamma r_{t+1} + \dots + \gamma^{T-t+1} r_{T-1} + \gamma^{T-t} V(s_T) - V(s_t) \quad (17)$$

式中: $E_t(\cdot)$ 为期望函数; $r_t(\theta)$ 为新策略 π_θ 和旧策略 π_{θ^-} 的比值; ε 为超参数; $A_\pi(s_t, a_t)$ 为优势函数, 表示状态 s_t 下选取动作 a_t 相对平均动作的优势。clip 函数将新旧策略比值限定在 $[1 - \varepsilon, 1 + \varepsilon]$, 防止了训练过程中智能体获得的奖励波动过大, 其具体数学表达为

$$\text{clip}(x, a, b) = \begin{cases} x & a \leq x \leq b \\ a & x < a \\ b & x > b \end{cases} \quad (18)$$

5) 判断是否达到最大训练轮数 N , 如果达到则结束训练, 否则返回步骤 2)。

4 算例分析

4.1 算例设置

本文以成都市某地区微网为例进行仿真研究, 该地区某一典型单日的光伏、负荷功率如图 4 所示。一天共设置 24 个调度时刻, 各个时刻的分时电价参考文献[17]。微网中设备参数如表 1 所示, 本文假设光伏发电成本为 0。

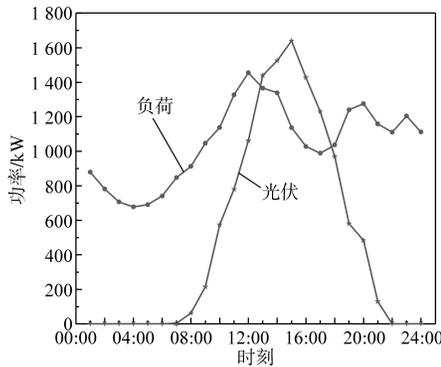


图4 典型日光伏和负荷数据

Fig.4 PV and load data of a typical day

对于智能体策略网络设置 4 层隐藏层, 每层神经元的个数往往按经验选取, 本文参考文献[18]的设计, 从输入到输出层每层神经元数量分别为 128, 128, 32 和 2; 对智能体价值网络也设置 4 层隐藏层, 由于价值网络仅输出 $V(s_t)$, 因此其与策略网络不同的是输出层神经元数量为 1, 从输入到输出层每层神经元数量分别为 128, 128, 32 和 1。神经元激活函数均为 Tanh, 该函数为双曲正切函数, 其输出均值为 0, 相较选择其他函数具

表 1 设备参数

Tab.1 The equipment paraments

设备	参数	数值
燃气轮机	功率上/下限	1 200/100 kW
	爬坡功率上/下限	150/-150 kW
	成本系数	0.33
电储能	功率上/下限	600/-600 kW
	荷电状态上下限	0.9/0.1
	充放电效率	0.97
	度电成本系数	0.15

有更快的收敛速度; 超参数 ε 设置为 0.01; 最大训练轮数 N 为 1 700; 奖励偏置常数设为 1 200。

本文仿真程序基于 Pytorch 框架编写, 硬件条件为 i7 11 800H CPU, 2.3 GHz, 24 G 内存。

4.2 结果分析

智能体更新过程中每个调度周期所获得的奖励之和变化规律如图 5 所示, 奖励值最终收敛于 8 700。训练开始时, 奖励变化幅值变化较大, 这是因为智能体尚未寻得最优策略, 尚处于探索阶段; 随着智能体神经网络更新, 智能体逐渐探索出最优策略, 因此奖励趋于收敛。

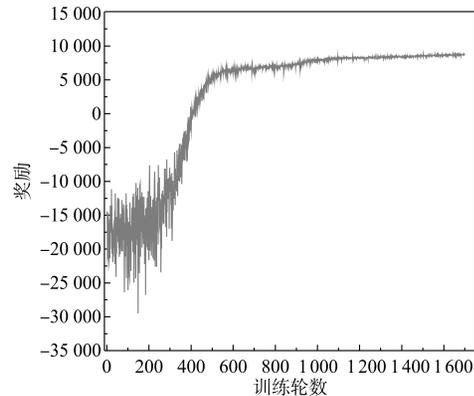


图5 奖励变化曲线

Fig.5 Reward change curve

智能体对该日的优化调度结果如图 6 所示。可见在 01:00—04:00 电价低谷时刻, 系统通过外网购电给电储能充电, 以用于其他电价较高时刻负荷供电, 降低购电成本和燃气轮机发电成本。在 13:00—17:00, 光伏发电功率高于负荷功率, 此时燃气轮机发电成本始终维持在下限, 多余的发电功率全部存储于储能中, 有效提升了可再生能源利用率。

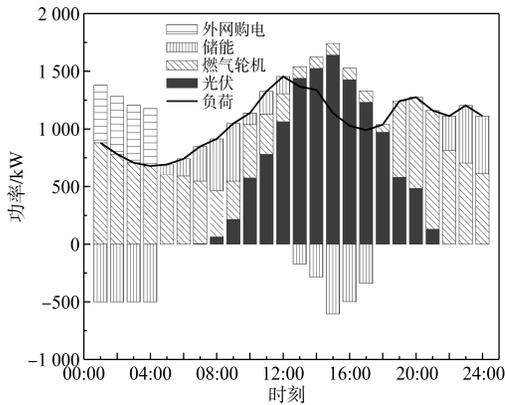


图6 微网优化调度结果

Fig.6 Optimization dispatch results of microgrid

4.3 对比分析

为了验证本文所提出的基于PPO的微网优化调度方法的有效性,将本文方法与文献[19]提出的改进粒子群算法(improved particle swarm optimization, IPSO)、文献[20]提出的基于深度双向长短期记忆网络(bidirectional long-short term memory, Bi-LSTM)的优化调度方法和CPLEX优化器进行对比。多次对该日进行优化调度,3种方法的最优运行成本如图7所示。

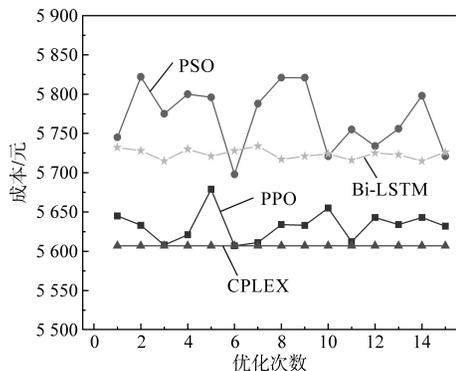


图7 不同方法的优化成本对比

Fig.7 Comparison of optimal cost of different methods

由图7可见,CPLEX优化结果始终保持不变,本文方法由于每次需重新更新神经网络参数,其结果具有波动性,但经过大量的训练,其调度结果始终与最优的CPLEX基本一致。本文方法下第3次优化结果最优,仅比最优的运行成本多0.0018%。

由于基于Bi-LSTM的优化调度方法需大量样本进行神经网络的训练,本文在图4所示光伏、负荷功率基础上叠加服从正态分布的随机性^[21],生成100个样本,利用求解器求解各样本对应的最优调度结果。通过100个样本对深度神经网络进行训练,可见经过训练后其优化成本始终高于本文所提方法,原因是基于Bi-LSTM的优化调度

方法的决策精度取决于样本容量大小,当实际调度情况与训练样本差异较大时,其优化结果会与最优解产生偏差。

IPSO方法容易陷入局部最优解,本文方法优化成本始终低于IPSO方法,验证了本文所提方法的有效性。

5 结论

针对微网优化调度问题,本文提出了一种基于PPO算法的优化调度方法。首先构建了以微网运行经济性为目标的优化模型,并结合深度强化学习框架,对微网优化智能体的状态、动作和奖励进行设计,在此基础上设计了优化调度算法流程,通过实验仿真结果可知本文优化方法优于粒子群算法。本文为人工智能方法在电力系统中的应用提供了参考。

参考文献

- [1] 吕天光. 含多微网配电系统多元主体互动运行策略研究[D]. 上海:上海交通大学,2018.
LÜ Tianguang. Study on interactive operation strategies for multiple agents in power distribution systems with multi-microgrids[D]. Shanghai:Shanghai Jiao Tong University, 2018.
- [2] 沈儒茹. 多微网系统的优化调度策略研究[D]. 哈尔滨:哈尔滨工业大学,2020.
SHEN Ruru. Research on optimal scheduling strategy of multi-microgrid system[D]. Harbin: Harbin Industrial University, 2020.
- [3] 帅轩越,王秀丽,吴雄,等. 计及条件风险价值下基于合作博弈的多微网协同优化调度[J]. 电网技术,2022,46(1):130-138.
SHUAI Xuanyue, WANG Xiuli, WU Xiong, et al. Cooperative optimal scheduling of multi-microgrids based on cooperative game considering conditional value at risk[J]. Power System Technology, 2022, 46(1): 130-138.
- [4] 李鹏,吴迪凡,李雨薇,等. 基于综合需求响应和主从博弈的多微网综合能源系统优化调度策略[J]. 中国电机工程学报, 2021, 41(4): 1307-1321, 1538.
LI Peng, WU Difan, LI Yuwei, et al. Optimal dispatch of multi-microgrids integrated energy system based on integrated demand response and stackelberg game[J]. Proceedings of the CSEE, 2021, 41(4): 1307-1321, 1538.
- [5] 许琼果,张炜焘,刘光伟,等. 计及储能调节特性的直流微电网优化调度研究[J]. 电气传动, 2022, 52(15): 53-60.
XU Qiongguo, ZHANG Weitao, LIU Guangwei, et al. Research on optimal scheduling of DC microgrid considering energy storage regulation characteristics[J]. Electric Drive, 2022, 52(15): 53-60.

- [6] 宋晓英,王艳松. 基于协同进化遗传算法的微网经济环保调度[J]. 电力系统保护与控制, 2014, 42(5): 85-89.
SONG Xiaoying, WANG Yansong. Economic and environmental dispatch of microgrid using co-evolutionary genetic algorithm[J]. Power System Protection and Control, 2014, 42(5): 85-89.
- [7] WANG L, LAN X G, ZHANG H B. The theory application of deep reinforcement learning[J]. Pattern Recognition and Artificial Intelligence, 2019, 32(1): 67-81.
- [8] 熊珞琳,毛帅,唐漾,等. 基于强化学习的综合能源系统管理综述[J]. 自动化学报, 2021, 47(10): 2321-2340.
XIONG Luolin, MAO Shuai, TANG Yang, et al. Reinforcement learning based integrated energy system management: a survey [J]. Acta Automatica Sinica, 2021, 47(10): 2321-2340.
- [9] 吴利刚,张梁,周倩,等. 基于强化学习的微电网能量调度优化策略研究[J]. 控制工程, 2022, 29(7): 1162-1172.
WU Ligang, ZHANG Liang, ZHOU Qian, et al. Research on energy scheduling optimization strategy of micro-grid based on reinforcement learning[J]. Control Engineering of China, 2022, 29(7): 1162-1172.
- [10] 邢放,李云. 面向主动配电网的调峰需求系统调度学习优化[J]. 自动化与仪器仪表, 2021(3): 150-153, 157.
XING Ao, LI Yun. System scheduling learning optimization for peak load regulation in active distribution network[J]. Automation & Instrumentation, 2021(3): 150-153, 157.
- [11] 谢启跃,应雨龙. 基于深度确定性策略梯度的电-气综合能源微网优化调度研究[J]. 热力发电, 2022, 51(3): 141-147.
XIE Qiyue, YING Yulong. Research on optimal dispatch of electricity-gas integrated energy microgrid based on deep deterministic policy gradient[J]. Thermal Power Generation, 2022, 51(3): 141-147.
- [12] MOCANU E, MOCANU D C, NGUYEN P H, et al. On-line building energy optimization using deep reinforcement learning [J]. IEEE Transactions on Smart Grid, 2019, 10(4): 3698-3708.
- [13] ZHANG J W, ZHANG Z H, HAN S, et al. Proximal policy optimization via enhanced exploration efficiency[J]. Information Sciences, 2022, 609: 750-765.
- [14] 叶宇剑,王卉宇,汤奕,等. 基于深度强化学习的居民实时自治最优能量管理策略[J]. 电力系统自动化, 2022, 46(1): 110-119.
YE Yujian, WANG Huiyu, TANG Yi, et al. Real-time autonomous optimal energy management strategy for residents based on deep reinforcement learning[J]. Automation of Electric Power System, 2022, 46(1): 110-119.
- [15] 闫庆友,史超凡,秦光宇,等. 基于近端策略优化算法的电化学/氢混合储能系统双层配置及运行优化[J]. 电力建设, 2022, 43(8): 22-32.
YAN Qingyou, SHI Chaofan, QIN Guangyu, et al. Research on two-layer configuration and operation optimization based on proximal policy optimization for electrochemical/hydrogen hybrid energy storage system[J]. Electric Power Construction, 2022, 43(8): 22-32.
- [16] 李咸善,陈敏睿,程杉,等. 基于双重激励协同博弈的含电动汽车微电网优化调度策略[J]. 高电压技术, 2020, 46(7): 2286-2296.
LI Xianshan, CHEN Minrui, CHENG Shan, et al. Research on optimal scheduling strategy of microgrid with electric vehicles based on dual incentive cooperative game[J]. High Voltage Engineering, 2020, 46(7): 2286-2296.
- [17] 董雷,刘雨,乔骥,等. 基于多智能体深度强化学习的电热联合系统优化运行[J]. 电网技术, 2021, 45(12): 4729-4738.
DONG Lei, LIU Yu, QIAO Ji, et al. Optimal dispatch of combined heat and power system based on multi-agent deep reinforcement learning[J]. Power System Technology, 2021, 45(12): 4729-4738.
- [18] 于新海,王鑫,苏日古格,等. 基于GA和IPSO的微电网调度优化与对比分析[J]. 电气传动, 2022, 52(22): 38-43.
YU Xinhai, WANG Xin, SU Riguge, et al. Microgrid dispatching optimization and comparative analysis based on GA and IPSO[J]. Electric Drive, 2022, 52(22): 38-43.
- [19] 陈卫东,吴宁,黄彦璐,等. 基于深度学习的微电网优化调度辅助决策方法[J]. 南方电网技术, 2022, 16(1): 117-126.
CHEN Weidong, WU Ning, HUANG Yanlu, et al. Auxiliary decision-making method of optimal dispatching for microgrid based on deep learning[J]. Southern Power Technology, 2022, 16(1): 117-126.
- [20] 袁桂丽,贾新潮,房方,等. 虚拟电厂源荷双侧热电联合随机优化调度[J]. 电网技术, 2020, 44(8): 2932-2940.
YUAN Guili, JIA Xinchao, FANG Fang, et al. Joint stochastic optimal scheduling of heat and power considering source and load sides of virtual power plant[J]. Power System Technology, 2020, 44(8): 2932-2940.
- [21] 陈明昊,孙毅,谢志远. 基于双层深度强化学习的园区综合能源系统多时间尺度优化管理[J]. 电工技术学报, 2023, 38(7): 1864-1881.
CHEN Minghao, SUN Yi, XIE Zhiyuan. The multi-time-scale management optimization method for park integrated energy system based on the bi-layer deep reinforcement learning[J]. Transactions of China Electrotechnical Society, 2023, 38(7): 1864-1881.

收稿日期:2022-08-29
修改稿日期:2022-10-27