

基于深度强化学习的综合能源系统增量场景 动态优化调度方法

杨茂,王金鑫,朱一丹,王宇鑫,杜茜,苏欣

(东北电力大学 现代电力系统仿真控制与绿色电能新技术教育部重点实验室 吉林 吉林 132012)

摘要:在综合能源系统(IES)中,能量设备出力及负荷功率分布偏移会使深度强化学习优化调度模型的性能下降,由此提出基于深度强化学习的IES增量场景动态优化调度方法。针对以经济性为目标的IES优化调度模型,构建马尔科夫决策过程并对模型求解;提出新增场景筛选方法以选择出需要重新学习的增量场景,并采用基于知识蒸馏的增量学习方法对模型进行更新训练,以提升对增量场景的适配性;采用实地采集的IES运行场景数据集对所提方法进行验证。结果表明训练所得模型满足约束条件,在测试条件下,IES运行成本降低了24.4%,所提方法能够在提升调度模型响应功率偏移能力的同时提升系统运行的经济性。

关键词:深度强化学习;综合能源系统;近端策略优化;动态优化调度;新能源消纳率

中图分类号:TM73;TP18

文献标志码:A

DOI:10.16081/j.epae.202511029

0 引言

综合能源系统(integrated energy system, IES)集成多能源系统以实现多类能源间融合互补,为提升能源利用效率、降低碳排放提供了新的解决方案^[1-3]。IES优化调度是提升系统运行经济性、降低能源损耗的重要手段^[4-5]。

深度强化学习(deep reinforcement learning, DRL)作为解决时序决策问题的有效手段,因其实时性、对强随机的抗干扰性以及不依赖物理模型的强鲁棒性,近年来多应用于IES优化调度领域^[6-7]。例如,文献[8]考虑IES的经济和低碳目标,采用近端策略优化(proximal policy optimization, PPO)解决IES的优化调度问题;文献[9]采用双网络拟合联合奖励函数,利用多智能体DRL实现分布式非凸经济问题的求解。上述文献均采用深层次人工智能体构成强化学习的动作及奖励网络,利用深度学习非线性映射的强拟合能力以及深层特征的强挖掘能力,构建了更加经济、稳定的IES优化调度模型。

深度神经网络是DRL的重要构成单元,因而基于DRL的优化调度方法同样具有与深度学习相同的训练特征^[10-11]:DRL模型训练需要大量、多样的训练数据;实际应用场景需与训练样本场景独立同分布;DRL模型一次性学习全部训练样本后,学习过程即终止。然而,现实中IES应用场景多样,如大风、极寒、沙尘暴、阴雨等极端天气会造成新能源发电及用户电热负荷功率分布偏移^[12-14],同时,风、光发电

的强随机性导致风光发电功率序列易出现偏离训练集分布的样本^[15-16]。因极端天气、新能源发电强随机性等因素所产生的偏离训练集分布的新增场景发生概率低,在构建模型训练集时很难采集到大量的样本,而所构建的模型在对训练集学习后即终止训练,因此在实际遇到此类场景时,模型很难准确响应,进而导致发电功率动作越限或经济性下降。

对于增量场景的学习问题,传统方式是将新增场景与训练场景样本混合后重新训练,然而,大量的训练集样本与少量的增量场景样本间存在样本不平衡问题,增量场景样本的学习效果有限。同时,重复学习大量历史样本会造成极大的算力资源浪费。因此,通过模型修正更新逐步提升对新场景样本的泛化能力,是一种更为有效的解决手段,但目前在IES优化调度领域中未见相关研究。增量学习作为一种递进式学习新增样本的有效方法^[17],已在多个领域获得关注与应用。例如,文献[18]提出一种基于学习共享的增量学习方法,通过优化边缘设备上的深度神经网络增量学习过程,降低计算和内存的消耗;文献[19]提出一种基于Q学习和自适应核线性模型的增量学习框架,为动态环境中的任务适应和自主学习提供了解决方案。然而,上述方法在应对复杂场景中新增样本与历史样本之间的知识冲突问题上仍存在局限性,尤其在面临样本类别不平衡或多样化场景时,模型的泛化能力受限。而基于知识蒸馏(knowledge distillation, KD)的增量学习方法采用回放历史样本的方式记忆历史样本信息,训练时同时采用原模型的软标签和新增样本的硬标签对模型进行更新,可在记忆历史信息的同时持续学习新场景,例如文献[20]结合KD技术提出一种基于半监督小样本类增量学习方法,在训练过程中利用软标签保

收稿日期:2024-11-01;修回日期:2025-06-24

在线出版日期:2025-12-03

基金项目:吉林省产业技术研究与开发项目(2023C033-5)

Project supported by the Industrial Technology Research and Development Project of Jilin Province(2023C033-5)

留了历史知识的细粒度信息,在应对样本类别不均或多样化场景时能有效提升模型的泛化能力。基于此,可将基于KD的增量学习方法应用于IES优化调度中,通过识别发生概率较低的极端场景将其作为增量场景,并对基于KD的增量学习模型进行更新训练,在保留历史场景知识的同时高效适配新增调度场景,实现在低算力需求条件下对复杂动态环境的快速响应和优化调度。

综上所述,本文提出一种基于DRL的IES增量场景动态优化调度方法。首先,针对以经济性为目标的IES优化调度模型,构建马尔可夫决策过程,并基于PPO求解该决策问题;然后,提出基于局部离群因子(local outlier factor, LOF)离群检测的新增场景筛选方法以选择出需要重新学习的增量场景,并采用基于KD的增量学习方法,对模型进行更新训练,以提升对筛选出的增量场景的适配性;最后,利用在比利时某园区IES所采集的场景样本,对所提方法进行验证,结果表明所提方法能够高效地学习新增场景,训练所得模型的调度结果能可靠满足模型约束,提升系统经济效益。

1 基于PPO的IES动态优化调度

1.1 IES优化调度模型

本文所构建的IES主要包括电、热、气3种形式能源,如图1所示。IES与上级电网、气网交互,源侧设备包括风机(wind turbine, WT)、光伏(photo-voltaic, PV)、电锅炉(electric boiler, EB)、燃气锅炉(gas boiler, GB)、热电联产机组(combined heat and power, CHP)以及电储能(energy storage, ES)。荷侧包括电负荷、气负荷和热负荷。

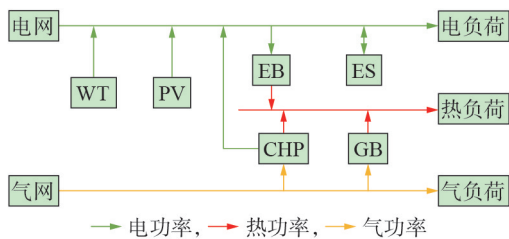


图1 IES结构示意图

Fig.1 Structure diagram of IES

以IES运行成本最小为目标,目标函数 F 如式(1)所示。

$$\begin{cases} \min F = C_E + C_G + C_m \\ C_E = \sum_{t=1}^T \varepsilon_E^t P_E^t \\ C_G = \sum_{t=1}^T \varepsilon_G^t M_G^t \\ C_m = \sum_{t=1}^T \sum_{n=1}^N \gamma_n P_n^t \end{cases} \quad (1)$$

式中: C_E 为IES与电网交互成本; ε_E^t 为 t 时段电网分时电价; P_E^t 为 t 时段IES与电网交互功率^[21]; C_G 为IES购气成本; ε_G^t 为 t 时段的天然气价格; M_G^t 为 t 时段输入IES的燃气量; C_m 为设备维护成本; γ_n 为第 n 种设备的单位维护成本; P_n^t 为第 n 种设备在 t 时段的输出功率; T 为调度周期; N 为IES的设备总数。

约束条件包括各设备的功率约束,电功率、热功率和天然气平衡约束,ES运行约束等,具体表达式见附录A式(A1)—(A5)。

1.2 IES优化调度马尔科夫决策模型

本文通过DRL求解IES的优化调度问题,将IES优化调度问题的数学表述转化为DRL的马尔科夫决策框架,包含环境空间 S ,动作空间 A ,状态转移概率 P ,奖励函数 R 和智能体策略 π 。IES的优化调度可建模为DRL智能体的交互过程,智能体通过调整机组出力优化调度决策。在 t 时段,环境向智能体提供观测到的系统状态 s_t ,智能体基于策略 π 和 s_t 生成动作 a_t ,执行动作后,智能体从环境中获得反馈奖励,并将 a_t 传递给环境,同时进行自我更新以优化未来的决策能力。

其中,状态是智能体决策的重要依据,应当覆盖所有对决策有影响的因素。本文IES的观测状态包括 t 时段IES与电网交互功率 P_E^t 、WT功率 P_{WT}^t 、PV功率 P_{PV}^t 、ES功率 P_{ES}^t 、EB功率 P_{EB}^t 、CHP机组的燃气量 M_{CHP}^t 、GB的燃气量 M_{GB}^t 、储能的荷电状态 S_{ES}^t 、电负荷 P_{load}^t 、气负荷 M_{load}^t 以及热负荷 $Q_{H,load}^t$ 。具体如式(2)所示。

$$s_t = \{t, P_E^t, P_{WT}^t, P_{PV}^t, P_{ES}^t, P_{EB}^t, M_{CHP}^t, M_{GB}^t, S_{ES}^t, P_{load}^t, M_{load}^t, Q_{H,load}^t\} \quad (2)$$

选择动作空间时需考虑完备性,以确保能够找到最优解。本文选取IES中各能源在下一时刻的计划出力及交互功率为决策变量。 a_t 如式(3)所示。

$$a_t = \{P_E^{t+1}, P_{WT}^{t+1}, P_{PV}^{t+1}, P_{ES}^{t+1}, P_{EB}^{t+1}, M_{CHP}^{t+1}, M_{GB}^{t+1}\} \quad (3)$$

式中: P_E^{t+1} 为 $t+1$ 时段IES与电网交互功率; P_{WT}^{t+1} 、 P_{PV}^{t+1} 分别为 $t+1$ 时段WT与PV的功率; P_{ES}^{t+1} 为 $t+1$ 时段储能功率; P_{EB}^{t+1} 为 $t+1$ 时段EB的热功率; M_{CHP}^{t+1} 为 $t+1$ 时段CHP机组的燃气量; M_{GB}^{t+1} 为 $t+1$ 时段GB的燃气量。

奖励函数是衡量智能体行为优劣的标准,本文将IES成本函数作为奖励项,同时通过引入功率不平衡和出力越限等惩罚机制,得到 t 时段的即时奖励函数 R_t 为:

$$R_t = \begin{cases} -e_t F_{all,t} & F_{all,t} > 0 \\ \beta - e_c F & F_{all,t} = 0 \end{cases} \quad (4)$$

$$F_{all,t} = h^2 + \max(g, 0) \quad (5)$$

式中: e_t 和 e_c 分别为约束惩罚系数和总运行成本系

数; $F_{\text{all},t}$ 为约束惩罚函数,具体表达式如附录A式(A6)~(A9)所示; h 表示 IES 优化调度模型的等式约束; g 表示 IES 优化调度模型的不等式约束; β 为正变换系数。

1.3 基于 PPO 的 IES 决策模型求解

PPO 是一种结合动态步长机制和重要性采样技术的 DRL 算法,可有效处理大规模、非线性和随机性场景下的问题^[22-23]。PPO 算法用于确定动作概率分布的动作网络(actor network)和用于估计评价动作价值的评价网络(critic network)以实现决策模型求解。基于 IES 优化调度的 DRL 模型训练框架如附录A图A1所示,具体流程如下。

1) 决策中心与 IES 环境不断交互,并将每个决策步的状态、动作和奖励存储到记忆库,直到完成当前调度周期。此时 critic network 没有更新。

2) 将得到的 s_t 输入 critic network, 获得状态函数 $V(s_t)$, 并计算折扣奖励函数 $Q(s_t, a_t)$; 将存储的所有状态观测值输入 critic network 得到所有状态的 $V(s_t)$, 计算优势函数 A_t 。

3) 利用 A_t 求得 critic network 的损失函数 c_{loss} , 然后反向传递以更新 critic network。 c_{loss} 如式(6)所示。

$$c_{\text{loss}} = E[Q(s_t, a_t) - V(s_t)]^2 \quad (6)$$

式中: $E[\cdot]$ 为期望运算。

4) 将存储的所有状态观测值输入到新动作网络(actor-new network)和旧动作网络(actor-old network), 分别得到新策略分布 $N_q(m, s^2)$ 与旧策略分布 $N_{\text{qold}}(m, s^2)$, 其中参数 m 与 s^2 分别为策略分布的均值与方差; 将储存的所有动作值分别输入新、旧策略分布, 得到每个动作在新、旧策略中被选择的概率值。新策略概率除以旧策略概率得到策略概率之比 r_t 。

5) 利用 A_t 与 r_t 计算求得目标函数值, 目标函数的负值作为 actor network 的损失函数 a_{loss} 反向传递更新 actor network。 a_{loss} 如式(7)所示。

$$a_{\text{loss}} = E\left\{\min[r_t(\theta)A_t, c_{\text{lip}}(r_t(\theta), 1-\varepsilon, 1+\varepsilon)A_t]\right\} \quad (7)$$

式中: c_{lip} 为裁剪函数; θ 为策略网络参数; ε 为裁剪范围。

2 基于场景筛选与 KD 的 IES 增量场景动态优化调度

2.1 基于 LOF 的增量场景识别

IES 优化调度的需求在于保证调度动作在满足约束的条件下, 尽可能地提升系统运行的经济性, 从而在调度结果满足约束条件且经济性较优时, 该场景不需要重新学习^[24]。在调度结果成本过高或调度动作越限时, 需要重新学习增量场景特征, 此时可以采用约束惩罚函数值及系统运行成本函数值作为识

别增量场景的特征参量, 构建增量场景识别特征参量 T_r , 如式(8)所示。

$$T_r = [F_{\text{all},t}, F] \quad (8)$$

当增量场景偏离原训练集场景分布, 其特征参量也会远离原训练集调度结果特征参量的分布, 此时通过检测特征参量的离群点可以实现增量场景识别。LOF 离群检测法是一种基于密度的离群检测方法^[25], 通过比较样本点与其邻域中其他点的局部密度差异, 来判断该点是否是离群点, 该方法不受数据分布影响, 适合处理复杂数据分布条件下的离群检测问题, 具体计算过程如附录B所示。

2.2 基于 KD 的增量学习方法

基于 KD 的增量学习方法原理如附录C图C1所示, 该方法以原模型作为教师模型, 新模型作为学生模型, 通过将少量回放的历史样本及新增样本输入教师模型以向学生模型提供软标签, 新模型通过学习历史样本的软标签与新增样本的硬标签, 实现学习新增样本的同时保持对历史模型的记忆。

训练损失函数 L_{LOSS} 如式(9)所示。

$$L_{\text{LOSS}} = \lambda L_{\text{soft}} + (1-\lambda) L_{\text{hard}} \quad (9)$$

式中: L_{soft} 为软标签产生的蒸馏损失; L_{hard} 为硬标签产生的回归损失; λ 为平衡参数, 本文取 0.5^[26]。

2.3 IES 增量场景动态调度方法

DRL 并不存在显性的标签, 其模型的损失是由环境反馈, 基于 KD 的增量学习方法的核心思想在于将历史模型的先验知识迁移至新模型^[27], 因而本文直接通过损失函数方式迁移先验知识, 构建 IES 增量场景动态调度方法, 该方法框架如附录C图C2所示, 具体过程如下。

1) 基础环境、模型构建。根据 IES 优化调度模型及式(4)搭建调度环境, 对采集的 IES 场景样本进行 DRL 模型训练, 训练结束后根据式(8)构建特征参量, 并根据式(B3)计算所有场景的 LOF, 根据式(B4)计算 LOF 阈值。

2) 增量场景识别。在每个调度周期结束后, 根据式(B3)计算该日场景的 LOF, 将 LOF 超出阈值的场景作为增量场景, 累积少数日增量场景后进行增量场景学习。

3) 模型更新训练。以增量场景作为训练场景训练新模型, 计算新模型动作与评价的回归损失; 回放少量原训练集场景, 将其与增量场景融合构建训练集, 以原模型的动作作为 actor-old network、新模型作为 actor-new network、原模型的评价作为 critic network, 计算蒸馏损失; 根据式(9)计算训练损失函数, 对新模型的动作与评价进行更新训练。

4) 模型参数更新。将训练后的动作更新为 actor network 应用于 IES 优化调度, 将原训练场景与增量

场景融合,以更新后的网络计算特征参量,重新计算全部场景的LOF,并更新LOF阈值。

3 算例分析

3.1 算例描述

为验证所提方法有效性,采集比利时某园区IES 2018—2019年全年风电、PV、电-热-气负荷数据集,采集晴天或少云的常规场景330 d,采集大风、多云、大雨、极寒4类极端场景各60 d。其中常规场景270 d用于原调度模型训练,常规场景及各极端场景均取30 d用于增量学习,其余场景用于模型性能测试,训练及测试场景划分如附录D表D1所示。取系统调度周期为24 h,时段间隔为1 h,电网分时电价如附录D表D2所示,天然气价格为3元/m³。设备运行参数及其他参数如附录D表D3、D4所示。

以pytorch作为深度学习搭建平台,actor network和critic network由全连接层神经网络构建。2个网络均包含4个隐含层,各层神经元个数分别为200、100、100、50。选取Adam优化器,学习率为 5×10^{-4} ,矩估计因子分别为0.5和0.9,训练总次数为500次。DRL奖励折扣因子取0.95,正则化参数为0.01。优先经验回放池的参数 ϕ 为0.6, β 为0.5。采用表D1中原始训练场景训练模型作为原始模型,在原始模型基础上,采用表D1中的增量场景进行增量学习,将得到的模型作为增量学习模型,采用原始模型和增量学习模型对表D1中测试场景的调度结果进行测试分析。

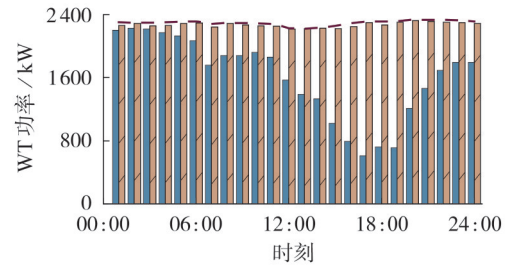
3.2 增量场景识别效果分析

设定LOF阈值系数 α_{LOF} 分别为0.9、0.95、1.00、1.05、1.10,对新增场景进行增量场景识别,不同 α_{LOF} 下的误检率与漏检率结果如附录D图D1所示。由图可知, α_{LOF} 高于1.00时存在漏检的情况,存在极端运行场景未被识别,而 α_{LOF} 达到1.00后,漏检率归0; α_{LOF} 低于1.00时误检率大幅增大。为了在尽可能保留极端运行场景样本的同时,尽可能突出其样本特征,需要在保证不漏检的条件下尽可能降低误检率,因而,本文LOF设定阈值系数 α_{LOF} 为1.00。

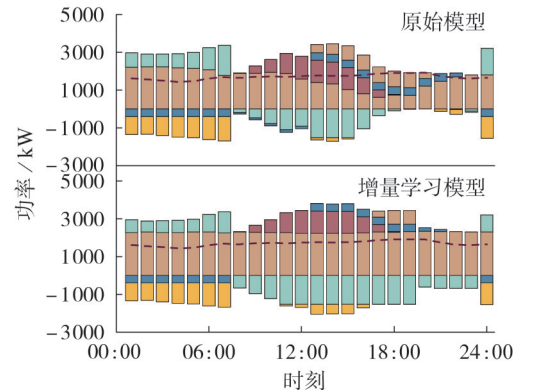
3.3 调度结果分析

为分析极端场景对调度结果的影响,随机抽取大风、多云、大雨、极寒的各1个场景,得到原始模型与增量学习模型在各测试场景中的调度结果,分析其功率曲线特征。在大风场景下,2个模型的电能设备调度结果如图2所示。

由图2(a)可知,在大风场景下,WT理论最大功率在各时段均较高,波动性不强。原始模型受原波动性较强的训练记忆影响,在06:00时刻后并未很好地响应较高的WT功率,调度结果远低于理论值,弃风情况较严重;而增量学习模型对大风场景进行了



(a) WT 功率调度结果对比



(b) 电能设备调度结果对比

图2 大风场景下各调度模型调度结果对比

Fig.2 Comparison of scheduling results from various scheduling models in high wind scenarios

学习更新,WT功率调度结果与理论值较为接近,很好地响应了与常规场景不同的风功率工况,弃风现象不明显。由图2(b)可知,由于原始模型调度结果弃风现象严重,尤其在10:00—22:00时段之间,WT功率较低,向电网售电量减少,相较于增量学习调度模型,经济收益降低。

在多云场景下,2个模型各电能设备调度结果如附录D图D2所示。在09:00、11:00、13:00时刻,由于云遮挡造成PV理论最大功率出现暂降,原始模型并未完全响应这种变化,所调度的PV功率超出了理论最大值;而增量学习模型的调度结果始终不高于理论最大值,且弃光量较少。由于需要补齐PV少发带来的缺额,原始模型售电量降低,售电收益减少,相较增量学习模型经济收益会有少量降低。

在大雨场景下,2个模型各电能设备调度结果如附录D图D3所示。大雨场景下PV理论最大功率大幅下降,原始模型未经过此类场景的训练,对此种工况难以响应,在存在光照的时间段,该模型的调度结果绝大部分出现了越限,由于响应困难,在14:00时刻也出现了较大的弃光量。而增量学习模型由于对此种场景进行过训练,调度所得PV功率均不超过理论最大功率,且弃光量较小。原始模型中IES通

过从电网增加购电量、降低售电量补偿PV功率缺额,且由于部分时段弃光量较大,进一步造成该时段售电量下降,因此其经济收益低于增量学习模型。

在极寒场景下,热负荷大幅上升,2个模型调度结果的热负荷功率差额及各热能设备调度结果如附录D图D4所示。由图D4(a)可知,增量学习模型很好地响应了这一变化,调度结果中热功率差额始终为0,而原始模型对极寒场景则不能很好响应,所发热功率较低,各时刻出现的功率缺额均较大。由图D4(b)可知,原始模型调度结果的热功率缺额均由EB从大电网交互中挪取电能发热补齐,在07:00—23:00时段,为补齐热功率缺额,IES需要在电价较高时从电网中购电或减少售电,以供EB发热,此时本该由更经济的其他机组提供的热功率被EB占用,因而极寒场景下,原始模型的运行成本较增量学习模型会有较大上升。

3.4 越限情况分析

调度结果可行的必要条件是符合调度模型的约束条件,调度结果越限会造成系统运行稳定性及经济性大幅降低,本文采用越限率 η_{lim} 作为评价调度模型调度结果越限情况的指标^[28],如式(10)所示。

$$\eta_{lim} = S_{num} / S_{sum} \times 100\% \quad (10)$$

式中: S_{num} 为测试场景中调度模型动作越限的场景个数; S_{sum} 为测试总场景个数。

对表D1中各类型测试场景进行测试,得到原始模型与增量学习模型在各测试场景中的越限率结果如图3所示。

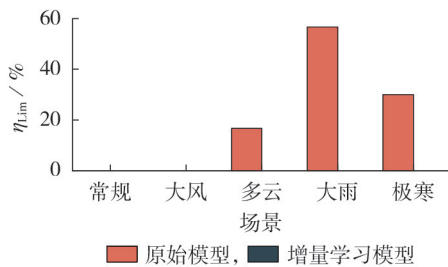


图3 各类测试场景调度结果越限情况对比

Fig.3 Comparison of over-limit scheduling results in various test scenarios

由图3可知,常规场景因与训练场景分布一致,原始模型可准确响应,该场景下并未出现违反约束的情况;在大风场景下,由于WT可发最大功率较常规场景更大,原始模型调度并未超出WT最大功率限制,也未出现越限情况;在多云场景下,由于此时部分时段PV最大可发功率较常规场景时大幅降低,原始模型并不能很好地响应这一变化,因而出现了少量越限情况;在大雨场景下,各时段PV最大可发功率均大幅降低,因而其出现的越限场景更多;在极寒场景下,用户热负荷大幅上升,原始模型未学习过

此场景调度方案,故而产生了热功率平衡失衡,出现了一部分越限的场景。而观察增量学习模型的越限情况可知,该模型在各场景下均未出现越限的情况,这是由于增量学习模型对各增量场景重新进行学习、更新,因而可以很好地响应各类型场景的功率变化,保证调度结果的有效性。

3.5 新能源消纳分析

新能源消纳率是节能减排的重要指标,为评价IES的新能源消纳能力^[29],定义新能源消纳率 R_{NEC} 为:

$$R_{NEC} = W_{actual} / W_{max} \times 100\% \quad (11)$$

式中: W_{actual} 为新能源实际消纳量; W_{max} 为新能源实际发电量。

对表D1中各类型场景进行测试,计算原始模型与增量学习模型在各测试场景中新能源消纳率的箱型图,结果如图4所示。

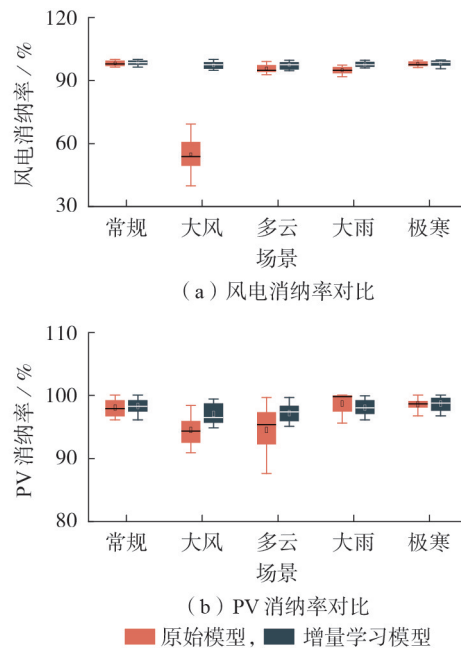


图4 各类测试场景新能源消纳率对比

Fig.4 Comparison of renewable energy integration rates in various test scenarios

对比常规场景下进行增量场景学习前、后的新能源消纳率可知,模型在常规场景下的新能源消纳率分布相差不大,在经过增量场景学习后的增量学习模型由于学习的常规场景样本稍多,其新能源消纳率分布较原始模型有轻微的上移趋势,可消纳的新能源电能也略多些。

对比大风场景下进行增量学习前、后的新能源消纳率可知,学习后的模型能有效响应高WT功率,其风电消纳率分布远高于原始模型,而原始模型受高WT功率扰动影响,PV消纳能力也略有下降,PV消纳率略低于增量学习模型。

对比多云场景下进行增量学习前、后的新能源

消纳率可知,原始模型的PV消纳率分布非常分散,由于多云场景部分时刻PV功率下降严重,原始模型对此种情况响应效果不佳,受到扰动后光照恢复时无法立即响应,提高发电功率,故而出现部分情况下PV消纳率很低,而多余PV功率下降不严重的场景,原始模型因其深度学习的泛化性能较好响应,此时PV消纳率较高,因而其PV消纳率分布非常分散、分布中心略低于增量学习模型,而WT发电也受PV功率变化扰动影响,电能消纳率略低于增量学习模型。

对比大雨场景下进行增量学习前、后的新能源消纳率可知,原始模型PV电能消纳率中位数处于100%,这是由于大雨场景下PV功率较常规场景下降过大,原始模型无法响应,调度结果中PV功率多处于越限情况,只能将PV发电保持在功率上限,此时本应由WT分担的发电功率在模型调度中被分配给了PV,造成风电消纳率下降。而对于少部分PV功率下降幅度不大的场景,原始模型虽可通过响应使得PV发电功率处于约束内,但由于分布较常规场景差异较大,其消纳率也有所下降,故其分布表现出大多数场景的PV消纳率集中于100%,而又存在很大占比场景的PV消纳率较低的情况。

对比极寒场景下进行增量学习前、后的新能源消纳率可知,此场景只对模型的热功率调度产生一定影响,对新能源发电影响不大,因此原始模型与增量学习模型的新能源消纳率分布近似。

3.6 经济性分析

以IES运行成本作为调度结果经济性的评价指标。对表D1中各类型场景进行测试,计算原始模型与增量学习模型在各测试场景中每个调度周期的系统运行成本。其中,调度结果因PV功率越限造成的功率缺额由上级电网交互购电补齐,热负荷缺额由EB从上级电网购电发热补齐,各类测试场景IES运行成本的箱型图如图5所示。

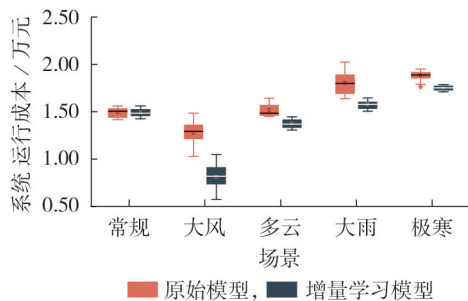


图5 各类测试场景系统运行成本对比

Fig.5 Comparison of operating cost in various test scenarios

除常规场景外,其他测试场景下采用增量学习模型IES的运行成本均低于原始模型。

在大风场景下,由于增量学习模型响应了大幅

提升的WT功率,向上级电网售出电能较多,使得此时IES运行成本处于较低状态,而原始模型并未很好地消纳风电功率,导致其运行成本增加,因此IES运行成本远高于增量学习模型。在多云场景下,原始模型一部分场景消纳新能源功率较增量学习模型低,另一部分场景原始模型调度的PV功率越限,需要由上级电网购电维持功率平衡,增加了系统运行成本,因而IES运行成本略高于增量学习模型。与多云场景类似,在大雨场景下,原始模型调度结果的PV功率越限更为严重,需从上级电网购买更多的电能,因而IES运行成本高于增量学习模型,且较多场景差额更大。在极寒场景下,由于原始模型未响应大幅增加的热负荷功率,热功率缺额由EB发热补齐,本应由系统内成本更低的其他能量设备协同发热的部分被成本更高的EB发热代替,因而原始模型IES运行成本较增量学习模型有小幅度上升。

3.7 方法泛用性评价

为验证本文所提方法的可推广性,依次训练基于深度确定性策略梯度(deep deterministic policy gradient, DDPG)^[30]和基于actor-critic^[31]的IES优化调度模型,得到原始模型及增量学习模型。将表D1中各类测试场景混合,测试各类方法训练所得模型的越限率、新能源消纳率与IES运行成本,其中,新能源消纳率与IES运行成本取平均值,得到结果如附录D表D5所示。

由表D5可知,PPO、DDPG、artor-critic这3种方法原始模型的调度结果中,artor-critic的越限率较低,DDPG的新能源消纳率较高,而PPO的系统运行成本最低。出现增量场景后,3种方法所训练得到的模型均出现了越限情况,经增量场景学习后,调度模型的动作均能满足约束条件,新能源消纳率均有所上升,系统运行成本均有所降低,其中,基于PPO的优化调度模型成本降低了24.4%。本文所提IES增量场景优化调度模型对常见的优化方法均表现出了优异的增量适配效果,因此具有可推广性。

4 结论

针对IES中能量设备及负荷功率分布偏移造成DRL优化调度模型性能下降的问题,本文提出了基于DRL的IES增量场景动态优化调度方法,所得结论如下。

1)由于能量设备及负荷功率分布发生偏移,调度模型对功率偏移响应困难,造成调度结果越限,对增量场景进行增量学习后可以有效响应功率变化,调度结果均能满足约束条件,提升模型的有效性。

2)新能源功率分布发生偏移后,调度模型对新能源消纳能力有所下降,通过对增量场景进行增量学习后,可以有效提升新能源消纳率。

3)由于调度模型对功率偏移响应困难,造成IES运行成本上升,通过对增量场景进行增量学习后,可以有效降低系统运行成本,提升系统运行的经济性。

4)本文所提方法对PPO、DDPG、actor-critic等常见DRL优化调度模型均存在增量适配效果,各方法的调度模型进行增量场景学习后,调度结果的有效性、经济性均有所提升,表现出了较强的泛用性。

附录见本刊网络版(<http://www.epae.cn>)。

参考文献:

- [1] YANG W T, LIU W J, CHUNG C Y, et al. Coordinated planning strategy for integrated energy systems in a district energy sector[J]. *IEEE Transactions on Sustainable Energy*, 2020, 11(3): 1807-1819.
- [2] 李鹏, 钟瀚明, 马红伟, 等. 基于深度强化学习的有源配电网多时间尺度源荷储协同优化调控[J]. *电工技术学报*, 2025, 40(5): 1487-1502.
LI Peng, ZHONG Hanming, MA Hongwei, et al. Multi-timescale optimal dispatch of source-load-storage coordination in active distribution network based on deep reinforcement learning[J]. *Transactions of China Electrotechnical Society*, 2025, 40(5): 1487-1502.
- [3] 陈静, 杨坤漓, 张岩, 等. 多时间尺度下计及综合需求响应和碳捕集-电转气联合运行的综合能源系统优化调度[J]. *电力自动化设备*, 2024, 44(6): 68-76.
CHEN Jing, YANG Kunli, ZHANG Yan, et al. Optimal scheduling of integrated energy system considering integrated demand response and joint operation of carbon capture and power-to-gas under multiple time scales[J]. *Electric Power Automation Equipment*, 2024, 44(6): 68-76.
- [4] 胡国峰, 李勇, 曹一家, 等. 考虑多能负荷波动实时平衡的综合能源系统多时间尺度优化调度[J]. *电力自动化设备*, 2024, 44(5): 120-126.
HU Guofeng, LI Yong, CAO Yijia, et al. Multi-time scale optimal scheduling of integrated energy system considering real-time balancing of multi-energy load fluctuation[J]. *Electric Power Automation Equipment*, 2024, 44(5): 120-126.
- [5] 刘妍, 胡志坚, 陈锦鹏, 等. 含碳捕集电厂与氢能多元利用的综合能源系统低碳经济调度[J]. *电力系统自动化*, 2024, 48(1): 31-40.
LIU Yan, HU Zhijian, CHEN Jinpeng, et al. Low-carbon economic dispatch of integrated energy system considering carbon capture power plant and multi-utilization of hydrogen energy[J]. *Automation of Electric Power Systems*, 2024, 48(1): 31-40.
- [6] 潘玺安, 艾欣, 胡俊杰, 等. 考虑网络安全约束的分布式智能电网边云协同优化调度方法[J]. *电工技术学报*, 2024, 39(19): 6104-6118.
PAN Xi'an, AI Xin, HU Junjie, et al. Network security constrained distributed smart grid edge-cloud collaborative optimization scheduling[J]. *Transactions of China Electrotechnical Society*, 2024, 39(19): 6104-6118.
- [7] LEE S, SEON J, SUN Y G, et al. Novel architecture of energy management systems based on deep reinforcement learning in microgrid[J]. *IEEE Transactions on Smart Grid*, 2024, 15(2): 1646-1658.
- [8] 杨挺, 刘豪, 王静, 等. 基于深度强化学习的园区综合能源系统低碳经济调度[J]. *电网技术*, 2024, 48(9): 3604-3613.
YANG Ting, LIU Hao, WANG Jing, et al. Deep reinforcement learning-based low-carbon economic dispatch of park integrated energy system[J]. *Power System Technology*, 2024, 48(9): 3604-3613.
- [9] DING L F, YAN G F, LIU J N. Multiagent reinforcement learning for strictly constrained tasks based on reward recorder[J]. *International Journal of Intelligent Systems*, 2022, 37(11): 8387-8411.
- [10] DOLATABADI A, ABDELTAWAB H, MOHAMED Y A I. Deep reinforcement learning-based self-scheduling strategy for a CAES-PV system using accurate sky images-based forecasting[J]. *IEEE Transactions on Power Systems*, 2023, 38(2): 1608-1618.
- [11] 杨楠, 叶迪, 林杰, 等. 基于数据驱动具有自我学习能力的机组组合智能决策方法研究[J]. *中国电机工程学报*, 2019, 39(10): 2934-2946.
YANG Nan, YE Di, LIN Jie, et al. Research on data-driven intelligent security-constrained unit commitment dispatching method with self-learning ability[J]. *Proceedings of the CSEE*, 2019, 39(10): 2934-2946.
- [12] 姜海洋, 杜尔顺, 马佳豪, 等. 考虑长周期供需不平衡风险的新型电力系统规划方法[J]. *中国电机工程学报*, 2024, 44(15): 5845-5858.
JIANG Haiyang, DU Ershun, MA Jiahao, et al. Power system optimal planning method considering long-term imbalance risk[J]. *Proceedings of the CSEE*, 2024, 44(15): 5845-5858.
- [13] YU G Z, LU L, TANG B, et al. Ultra-short-term wind power subsection forecasting method based on extreme weather[J]. *IEEE Transactions on Power Systems*, 2023, 38(6): 5045-5056.
- [14] CAMAL S, MICHIORRI A, KARINIOTAKIS G. Reliable provision of ancillary services from aggregated variable renewable energy sources through forecasting of extreme quantiles[J]. *IEEE Transactions on Power Systems*, 2023, 38(4): 3070-3084.
- [15] 杨茂, 王金鑫. 考虑可再生能源出力不确定的孤岛型微电网优化调度[J]. *中国电机工程学报*, 2021, 41(3): 973-985.
YANG Mao, WANG Jinxin. Optimal scheduling of islanded microgrid considering uncertain output of renewable energy[J]. *Proceedings of the CSEE*, 2021, 41(3): 973-985.
- [16] FANG F, YU S Y, XIN X L. Data-driven-based stochastic robust optimization for a virtual power plant with multiple uncertainties[J]. *IEEE Transactions on Power Systems*, 2022, 37(1): 456-466.
- [17] LI Z, HOIEM D. Learning without forgetting[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2018, 40(12): 2935-2947.
- [18] HUSSAIN M A, HUANG S H, TSAI T H. Learning with sharing: an edge-optimized incremental learning method for deep neural networks[J]. *IEEE Transactions on Emerging Topics in Computing*, 2023, 11(2): 461-473.
- [19] HU Y M, LI D C, HE Y Q, et al. Incremental learning framework for autonomous robots based on Q-learning and the adaptive kernel linear model[J]. *IEEE Transactions on Cognitive and Developmental Systems*, 2022, 14(1): 64-74.
- [20] CUI Y W, DENG W X, CHEN H Y, et al. Uncertainty-aware distillation for semi-supervised few-shot class-incremental learning[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2024, 35(10): 14259-14272.
- [21] 邸维斌. 分布式电源接入对电网配电系统的影响及优化措施[J]. *电力设备管理*, 2024(23): 42-44.
DI Weibin. The impact of distributed power generation access on the power grid distribution system and optimization measures[J]. *Electric Power Equipment Management*, 2024(23): 42-44.
- [22] 杨楠, 宋旭日, 董亮, 等. 基于生成式多对抗强化学习的高比例

- 新能源电网日内优化调度[J]. 电力自动化设备, 2025, 45(11):43-51.
YANG Nan, SONG Xuri, DONG Liang, et al. Intraday optimal scheduling for power system with high renewable energy based on generative multi-adversarial reinforcement learning [J]. Electric Power Automation Equipment, 2025, 45(11):43-51.
- [23] 李天明, 王小君, 窦嘉铭, 等. 基于约束强化学习的综合能源系统优化调度研究[J]. 电力系统保护与控制, 2025, 53(6):1-14.
LI Tianming, WANG Xiaojun, DOU Jiaming, et al. Research on optimal dispatch of integrated energy systems based on constrained reinforcement learning [J]. Power System Protection and Control, 2025, 53(6):1-14.
- [24] 尹博艺, 王灿, 翟昕. 考虑SOFC精细化建模的综合能源系统不确定性调度[J]. 电力系统自动化, 2025, 49(21):149-159.
YIN Boyi, WANG Can, ZHAI Xin. Uncertainty-aware scheduling of integrated energy system considering refined modeling for solid oxide fuel cell[J]. Automation of Electric Power Systems, 2025, 49(21):149-159.
- [25] 凌莉, 程张玉, 邹承明. 融合孤立森林和局部离群因子的离群点检测方法[J]. 计算机应用与软件, 2022, 39(12):278-283.
LING Li, CHENG Zhangyu, ZOU Chengming. Outlier detection method based on isolation forest and LOF[J]. Computer Applications and Software, 2022, 39(12):278-283.
- [26] 张翼, 朱永利. 结合知识蒸馏和图神经网络的局部放电增量识别方法[J]. 电工技术学报, 2023, 38(5):1390-1400.
ZHANG Yi, ZHU Yongli. Incremental partial discharge recognition method combining knowledge distillation with graph neural network[J]. Transactions of China Electrotechnical Society, 2023, 38(5):1390-1400.
- [27] 陈浩, 张文朝, 黄志光, 等. 考虑极端天气的先验知识引导风/光短期出力场景生成方法研究[J]. 智慧电力, 2025, 53(3):44-52.
CHEN Hao, ZHANG Wenchao, HUANG Zhiguang, et al. Research on prior knowledge-guided short-term wind/photovoltaic output scenario generation method considering extreme weather [J]. Smart Power, 2025, 53(3):44-52.
- [28] 符杨, 廖剑波, 李振坤, 等. 考虑越限风险的主动配电网日前优化调度及运行[J]. 中国电机工程学报, 2017, 37(21):6328-6338.
FU Yang, LIAO Jianbo, LI Zhenkun, et al. Day-ahead optimal scheduling and operating of active distribution network considering violation risk[J]. Proceedings of the CSEE, 2017, 37(21):6328-6338.
- [29] 潘志忠, 孔宁, 王燕涛. 改进新能源消纳的配电网资源优化配置研究[J]. 东北电力大学学报, 2023, 43(6):71-78.
PAN Zhongzhi, KONG Ning, WANG Yantao. Joint optimal allocation method of energy-load-storage in power supply area to improve new energy consumption capacity[J]. Journal of Northeast Electric Power University, 2023, 43(6):71-78.
- [30] 祁向龙, 陈健, 赵浩然, 等. 多时间尺度协同的配电网分层深度强化学习电压控制策略[J]. 电力系统保护与控制, 2024, 52(18):53-64.
QI Xianglong, CHEN Jian, ZHAO Haoran, et al. Multi-time scale cooperative voltage control strategy of a distribution network based on hierarchical deep reinforcement learning [J]. Power System Protection and Control, 2024, 52(18):53-64.
- [31] 李琛, 黄炎焱, 张永亮, 等. Actor-Critic框架下的多智能体决策方法及其在兵棋上的应用[J]. 系统工程与电子技术, 2021, 43(3):755-762.
LI Chen, HUANG Yanyan, ZHANG Yongliang, et al. Multi-agent decision-making method based on actor-critic framework and its application in wargame[J]. Systems Engineering and Electronics, 2021, 43(3):755-762.

作者简介:

杨 茂(1982—),男,教授,博士,主要研究方向为综合能源系统优化调度及新能源发电功率预测(E-mail: yangmao820@163.com);

王金鑫(1991—),男,博士研究生,主要研究方向为综合能源系统优化调度(E-mail:1436705235@qq.com)。

(编辑 丁欣欣)

Dynamic optimization and scheduling method for incremental scenarios of integrated energy system based on deep reinforcement learning

YANG Mao, WANG Jinxin, ZHU Yidan, WANG Yuxin, DU Qian, SU Xin

(Key Laboratory of Modern Power System Simulation and Control & Renewable Energy Technology, Ministry of Education, Northeast Electric Power University, Jilin 132012, China)

Abstract: In the integrated energy system (IES), the distribution deviation of energy equipment output and load power will cause the performance degradation of the deep reinforcement learning optimization scheduling models. A dynamic optimization and scheduling method of IES incremental scenarios based on deep reinforcement learning is proposed. Aiming at the IES optimization scheduling model with the goal of economy, the Markov decision-making process is constructed and the model is solved. A new scenario screening method is proposed to choose the incremental scenarios that need to be relearned, and an incremental learning method based on knowledge distillation is adopted to update and train the model, thereby improving the adaptability to incremental scenarios. The proposed method is validated using a dataset of operation scenarios collected from the field of IES. The results show that the trained model can meet the constraints. Under the test conditions, the operating cost of the IES is reduced by 24.4%. The proposed method can improve the ability of scheduling model to respond to power deviations while enhancing the economic efficiency of system operation.

Key words: deep reinforcement learning; integrated energy system; proximal policy optimization; dynamic optimization scheduling; consumption rate of new energy