

DOI: 10.19783/j.cnki.pspc.211685

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

蔺伟山¹, 王小君¹, 孙庆凯¹, 刘 翌¹, 和敬涵¹, 蒲天骄²

(1. 北京交通大学电气工程学院, 北京 100044; 2. 中国电力科学研究院有限公司, 北京 100192)

摘要: 随着综合能源系统中间歇性能源和负荷不确定性的逐步增强, 传统的调度方法局限于固定物理模型及参数设定, 难以较好地动态响应源荷的随机波动。针对这一问题, 提出了一种基于深度强化学习的综合能源系统动态调度方法。首先, 以数据驱动方式构建面向综合能源系统的深度强化学习模型, 通过智能体与综合能源系统的持续交互, 自适应学习调度策略, 降低对物理模型的依赖程度。其次, 通过添加随机扰动的方式表征源荷不确定性变化特征, 针对不确定性变化特征改进深度强化学习模型的状态空间、动作空间、奖励机制以及训练流程等关键环节, 并经由近端策略优化算法优化求解, 实现了综合能源系统的动态调度决策。最后, 通过算例仿真验证了所提方法在不同时间尺度以及不确定性环境下的可行性和有效性。

关键词: 综合能源系统; 动态调度; 不确定性; 深度强化学习; 近端策略优化

Dynamic dispatch of an integrated energy system based on deep reinforcement learning in an uncertain environment

LIN Weishan¹, WANG Xiaojun¹, SUN Qingkai¹, LIU Zhao¹, HE Jinghan¹, PU Tianjiao²

(1. School of Electrical Engineering, Beijing Jiaotong University, Beijing 100044, China;

2. China Electric Power Research Institute Co., Ltd., Beijing 100192, China)

Abstract: As the uncertainties of intermittent energy and load in the integrated energy system gradually increase, traditional dispatch methods are limited to fixed physical models and parameter settings that can hardly respond to the random fluctuations in the dynamic system with source-load. In this paper, a deep reinforcement learning-based dynamic dispatch method for the integrated energy system is proposed to address this problem. First, a data-driven deep reinforcement learning model is constructed for the integrated energy system. Through the continuous interaction between agent and integrated energy system, the dispatch strategies are learned adaptively to reduce dependence on the physical models. Secondly, the variations of source-load uncertainties are characterized by adding random disturbances. Pivotal aspects such as state spaces, action spaces, reward mechanisms and the training process of the deep reinforcement learning model are improved according to the characteristics of uncertainties. Then a proximal policy optimization algorithm is used to solve the problem, and the dynamic dispatch decisions of the integrated energy system are realized. Finally, simulation results verify the feasibility and effectiveness of the proposed method over different time scales and in uncertain environments.

This work is supported by the National Natural Science Foundation of China (No. 51977005).

Key words: integrated energy system; dynamic dispatch; uncertainties; deep reinforcement learning; proximal policy optimization

0 引言

随着全球传统化石能源的日渐枯竭和环境问题的日益突出, 世界各国正积极调整能源结构, 通过

开发和利用可再生能源, 以减少对传统化石能源的依赖^[1-2]。近年来, 通过构建综合能源系统(Integrated Energy System, IES), 将多种异质能源互联融合、互补利用, 为提高能源的利用效率和缓解全球的环境压力提供了重要的解决方案^[3]。

IES 能实现对多种能源的综合管理和经济调

基金项目: 国家自然科学基金项目资助(51977005)

度,但随着多种能源之间的耦合程度不断提升以及可再生能源和负荷的不确定性逐渐增强,如何实现 IES 的实时、准确、动态调度成为亟待解决的问题^[4-5]。对于 IES 调度问题,文献[6]研究了 IES “源-网-荷-储”协调优化调度问题,并采用商业求解器 Cplex 进行求解。文献[7]采用改进的粒子群算法对电-热综合能源系统的经济调度模型进行求解。文献[8]提出一种含电转气两阶段运行的热电耦合 IES 优化调度模型。以上研究在 IES 优化调度方面取得一定的成果,但多依赖于详细的物理模型,局限于固定调度计划,未对源荷的不确定性问题做深入研究。

为应对源荷不确定性问题,文献[9]基于源荷不确定性客观规律和偏差预控提出一种多时间尺度滚动优化调度模型。文献[10]采用场景分析法对风电、光伏出力随机性进行建模求解。文献[11]考虑不确定性负荷和风电出力的极端场景,构建了两阶段鲁棒优化调度模型。文献[12]采用随机场景和鲁棒优化两种方法对风光出力和需求响应的不确定性因素进行了精细化建模。以上研究方法在系统实际运行时无法对源荷进行动态调整,而模型预测控制基于反馈校正的思想可以对调度结果进行滚动优化。文献[13]基于模型预测控制构建了 IES 多时间尺度优化调度模型。上述研究对解决考虑源荷不确定性的 IES 优化调度问题提供了良好的思路,但调度决策仍对源荷的精准预测过分依赖。

近年来,强化学习理论的不断发展为解决含不确定性因素的优化决策问题提供了新的思路。强化学习是一种数据驱动的方法,可经由智能体与环境的持续交互来不断改进策略。将深度学习的非线性拟合能力与强化学习的决策能力相结合的深度强化学习是解决高维复杂问题的有效手段^[4],可以避免对精细化物理模型的过分依赖。训练完成的深度强化学习模型可以进行短时间尺度的实时决策,满足算法实时性的需求,同时模型可以根据实际运行结果数据进行自我更新并对决策结果进行持续修正。目前已有研究将深度强化学习用于 IES 的经济调度和能量管理中。文献[15]提出了一种基于深度 Q 学习的 IES 能量管理模型,该模型将状态量和动作量进行了离散化,但离散程度精细化会造成维数灾难问题。文献[16-17]采用深度确定性策略梯度(Deep Deterministic Policy Gradient, DDPG)算法求解 IES 经济调度问题,该算法过分依赖超参数,收敛不稳定。

本文在现有研究的基础上采用近端策略优化(Proximal Policy Optimization, PPO)算法,提出不确

定性环境下基于深度强化学习的 IES 动态调度方法,具体做法为:(1)本文以数据驱动的方式构建 IES 动态调度模型,通过与 IES 环境的交互训练实现调度计划的自适应寻优,降低在处理 IES 调度问题时对固定物理模型的依赖程度;(2)针对源荷的不确定性问题,通过将训练数据随机抽样、为状态与动作空间添加扰动以及为奖励函数添加惩罚的方式改进深度强化学习模型,充分挖掘源荷的不确定性变化特征,实现 IES 在不确定性条件下的动态调度决策。

1 IES 动态调度问题

1.1 IES 介绍

IES 一般包括能源转换设备、能源存储设备、能源传输网络和负荷。本文对如图 1 所示的 IES 进行研究,系统中的能源转换设备包括热电联产机组(Combined Heat and Power, CHP)、电锅炉(Electric Boiler, EB)、燃气锅炉(Gas Boiler, GB)和光伏阵列(Photovoltaic, PV),能源存储设备为电储能(Battery Energy Storage, BES),负荷包括电负荷和热负荷。IES 能够与上级主电网和天然气供应商进行能量交互,同时利用协调优化控制方法对系统内多种能源进行综合使用,以提高能源的使用效率。

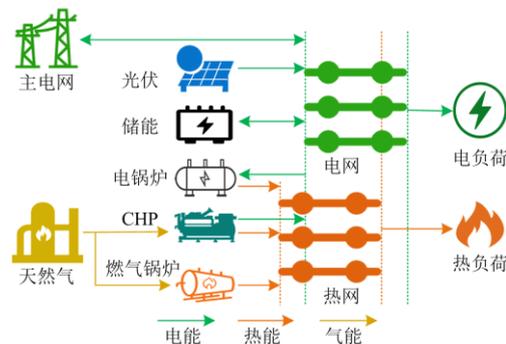


图 1 IES 结构示意图

Fig. 1 Schematic diagram of structure of IES

系统中各设备的运行特性介绍如下。

1.1.1 热电联产机组

热电联产机组消耗天然气产生电能和热能。 t 时刻热电联产机组输出的电功率为

$$P_{\text{CHP}}(t) = M_{\text{CHP}}(t)H_{\text{NG}}\eta_{\text{CHP}} \quad (1)$$

式中: $P_{\text{CHP}}(t)$ 为 t 时刻热电联产机组输出的电功率; $M_{\text{CHP}}(t)$ 为 t 时刻热电联产机组天然气的消耗量; H_{NG} 为天然气的热值; η_{CHP} 为热电联产机组的电转化效率。

热电联产机组输出的热功率与电功率之间的耦合关系称为“电热特性”,通常分为定热电比和变热电比两种类型,一般采用定热电比^[18],表示为

$$b = \frac{H_{\text{CHP}}(t)}{P_{\text{CHP}}(t)} \quad (2)$$

式中, $H_{\text{CHP}}(t)$ 为 t 时刻热电联产机组输出的热功率。

1.1.2 电锅炉

电锅炉消耗电能产生热能, t 时刻电锅炉输出的热功率为

$$H_{\text{EB}}(t) = P_{\text{EB}}(t)\eta_{\text{EB}} \quad (3)$$

式中: $H_{\text{EB}}(t)$ 为 t 时刻电锅炉输出的热功率; $P_{\text{EB}}(t)$ 为 t 时刻电锅炉消耗的电功率; η_{EB} 为电锅炉的转化效率。

1.1.3 燃气锅炉

燃气锅炉消耗天然气产生热能, t 时刻燃气锅炉输出的热功率为

$$H_{\text{GB}}(t) = M_{\text{GB}}(t)H_{\text{NG}}\eta_{\text{GB}} \quad (4)$$

式中: $H_{\text{GB}}(t)$ 为 t 时刻燃气锅炉输出的热功率; $M_{\text{GB}}(t)$ 为 t 时刻燃气锅炉天然气的消耗量; η_{GB} 为燃气锅炉的转化效率。

1.1.4 电储能

t 时刻电储能的荷电状态(State of Charge, SOC)为

$$C_{\text{SOC}}(t) = C_{\text{SOC}}(t-1) - \frac{\eta_{\text{BES}}P_{\text{BES}}(t)\Delta t}{Q_{\text{BES}}} \quad (5)$$

$$C_{\text{SOC}}(0) = C_{\text{SOC}}^{\text{ini}} \quad (6)$$

式中: $P_{\text{BES}}(t)$ 为 t 时刻电储能的充/放电功率, $P_{\text{BES}}(t)$ 为正时表示放电, $P_{\text{BES}}(t)$ 为负时表示充电; Q_{BES} 为电储能的容量; $C_{\text{SOC}}^{\text{ini}}$ 为电储能初始时刻的荷电状态; Δt 为调度时间尺度大小; η_{BES} 为电储能的充/放电系数, 具体如式(7)所示。

$$\eta_{\text{BES}} = \begin{cases} \eta_{\text{ch}}, & P_{\text{BES}}(t) < 0 \\ \frac{1}{\eta_{\text{dis}}}, & P_{\text{BES}}(t) \geq 0 \end{cases} \quad (7)$$

式中, η_{ch} 和 η_{dis} 分别为电储能的充、放电效率。

1.2 动态调度问题的数学描述

IES 的动态调度问题是在满足负荷需求的前提下, 以经济运行最优为目标, 有效协调系统中各设备在每个时段的出力, 以此来提升系统的经济效益。

1.2.1 目标函数

IES 的运行总成本包括常规成本和随机性成本两部分。

1) 常规成本

常规成本包括与主电网的电能交互成本、购买天然气的成本、电储能的充放电折旧成本和设备维护成本。由于设备维护成本相对其他成本较小, 故未考虑在常规成本中。因此常规成本表示为^[19]

$$C_{\text{E}} = \sum_{t=1}^T \varepsilon_{\text{E}}(t)P_{\text{grid}}(t)\Delta t + \sum_{t=1}^T \varepsilon_{\text{NG}}(t)M_{\text{NG}}(t)\Delta t + \quad (8)$$

$$\sum_{t=1}^T \rho_{\text{BES}} |P_{\text{BES}}(t)|$$

$$M_{\text{NG}}(t) = M_{\text{CHP}}(t) + M_{\text{GB}}(t) \quad (9)$$

式中: C_{E} 为系统运行的常规成本; $P_{\text{grid}}(t)$ 为 t 时刻系统与主电网交互的电功率, 为正时表示从主电网购电, 为负时表示从主电网售电; $\varepsilon_{\text{E}}(t)$ 为 t 时刻的电价, $P_{\text{grid}}(t)$ 为正时电价取购电电价, $P_{\text{grid}}(t)$ 为负时电价取售电电价; $\varepsilon_{\text{NG}}(t)$ 为 t 时刻的天然气单价; $M_{\text{NG}}(t)$ 为 t 时刻系统购买天然气的总量; ρ_{BES} 为电储能的折旧成本系数。

2) 随机性成本

由于 IES 中电负荷、热负荷以及光伏出力具有不确定性, 系统在运行过程中会有小概率出现源荷供需不平衡的情况, 对供需不平衡设置随机性成本, 如式(10)所示。

$$C_{\text{R}} = \sum_{t=1}^T \varepsilon_{\text{P}}^{\text{VAC}}(t)P_{\text{VAC}}(t)\Delta t + \sum_{t=1}^T \varepsilon_{\text{H}}^{\text{VAC}}(t)H_{\text{VAC}}(t)\Delta t \quad (10)$$

式中: C_{R} 为系统运行的随机性成本; $P_{\text{VAC}}(t)$ 为 t 时刻电功率供需不平衡量; $H_{\text{VAC}}(t)$ 为 t 时刻热功率供需不平衡量; $\varepsilon_{\text{P}}^{\text{VAC}}(t)$ 和 $\varepsilon_{\text{H}}^{\text{VAC}}(t)$ 分别为 t 时刻的电功率和热功率供需不平衡单价。

因此, 系统运行的总成本 F 表示为

$$F = \min(C_{\text{E}} + C_{\text{R}}) \quad (11)$$

1.2.2 约束条件

IES 动态调度问题的约束包括电功率平衡约束、热功率平衡约束、外部能源供应约束和各设备运行约束。

1) 电功率平衡约束

$$P_{\text{grid}}(t) + P_{\text{CHP}}(t) + P_{\text{PV}}(t) + P_{\text{BES}}(t) - P_{\text{EB}}(t) = P_{\text{load}}(t) \quad (12)$$

式中: $P_{\text{grid}}(t)$ 为与主电网交互的电功率; $P_{\text{CHP}}(t)$ 为热电联产机组输出的电功率; $P_{\text{PV}}(t)$ 为光伏阵列输出的电功率; $P_{\text{BES}}(t)$ 为电储能的充/放电功率; $P_{\text{EB}}(t)$ 为电锅炉消耗的电功率; $P_{\text{load}}(t)$ 为 t 时刻的电负荷。

2) 热功率平衡约束

$$H_{\text{CHP}}(t) + H_{\text{GB}}(t) + H_{\text{EB}}(t) = H_{\text{load}}(t) \quad (13)$$

式中: $H_{\text{CHP}}(t)$ 为热电联产机组输出的热功率; $H_{\text{GB}}(t)$ 为燃气锅炉输出的热功率; $H_{\text{EB}}(t)$ 为电锅炉输出的热功率; $H_{\text{load}}(t)$ 为 t 时刻的热负荷。

3) 能源交互约束

为保证系统能安全稳定地运行, 需对系统与主电网交互的电功率和系统天然气的购买总量进行约

束, 具体约束为

$$P_{\text{grid}}^{\min} \leq P_{\text{grid}}(t) \leq P_{\text{grid}}^{\max} \quad (14)$$

$$M_{\text{NG}}^{\min} \leq M_{\text{NG}}(t) \leq M_{\text{NG}}^{\max} \quad (15)$$

式中: P_{grid}^{\min} 和 P_{grid}^{\max} 分别为系统与主电网交互电功率的最小值和最大值; M_{NG}^{\min} 和 M_{NG}^{\max} 分别为系统天然气购买总量的最小值和最大值。

4) 设备运行约束

系统中各设备均有设备运行的上、下限范围, 对于热电联产机组输出电功率、电锅炉和燃气锅炉输出热功率、电储能设备充放电功率, 运行约束分别为

$$P_{\text{CHP}}^{\min} \leq P_{\text{CHP}}(t) \leq P_{\text{CHP}}^{\max} \quad (16)$$

$$H_{\text{EB}}^{\min} \leq H_{\text{EB}}(t) \leq H_{\text{EB}}^{\max} \quad (17)$$

$$H_{\text{GB}}^{\min} \leq H_{\text{GB}}(t) \leq H_{\text{GB}}^{\max} \quad (18)$$

$$P_{\text{BES}}^{\min} \leq P_{\text{BES}}(t) \leq P_{\text{BES}}^{\max} \quad (19)$$

式中: P_{CHP}^{\min} 和 P_{CHP}^{\max} 分别为热电联产机组输出电功率的最小值和最大值; H_{EB}^{\min} 和 H_{EB}^{\max} 分别为电锅炉输出热功率的最小值和最大值; H_{GB}^{\min} 和 H_{GB}^{\max} 分别为燃气锅炉输出热功率的最小值和最大值; P_{BES}^{\min} 和 P_{BES}^{\max} 分别为电储能设备充放电功率的最小值和最大值。

对于电储能设备, 为避免过度充放电对电储能设备造成损害, 需要对电储能设备的荷电状态进行约束, 如式(20)所示。

$$S_{\text{SOC}}^{\min} \leq S_{\text{SOC}}(t) \leq S_{\text{SOC}}^{\max} \quad (20)$$

式中, S_{SOC}^{\min} 和 S_{SOC}^{\max} 分别为电储能设备荷电状态的最小值和最大值。

2 基于深度强化学习的 IES 动态调度模型

强化学习的数学基础是马尔可夫决策过程, 一个马尔可夫决策模型通常被描述为一个五元组 $\langle \mathbf{S}, \mathbf{A}, \mathbf{P}, \mathbf{R}, \gamma \rangle$, 其中 \mathbf{S} 是状态空间; \mathbf{A} 是动作空间; $\mathbf{S} \times \mathbf{A} \times \mathbf{S}' \rightarrow \mathbf{P}$ 是状态转移矩阵; $\mathbf{S} \times \mathbf{A} \rightarrow \mathbf{R}$ 是奖励函数; γ 是折扣因子。

IES 中可再生能源出力易受不确定性环境的影响, 同时负荷需求也具有随机性, 这给 IES 动态调度问题的求解带来一定难度。本质上, IES 动态调度问题是一个随机序贯决策问题, 与强化学习决策过程相似。利用强化学习方法求解考虑源荷不确定性的 IES 动态调度问题具有良好的优势, 强化学习智能体以“试错”的方式进行自适应学习, 通过与环境的持续交互逐步调整自己的行动策略以适应环境^[20]。强化学习的目的是找到策略 π 使得系统获得最大的累积奖励值, 即获得最大价值。随着交互的进行, 策略 π 会被逐步更新, 通常采用动作-价值

函数 $Q_{\pi}(\mathbf{s}, \mathbf{a})$ 和状态-价值函数 $V_{\pi}(\mathbf{s})$ 衡量某策略 π 优劣程度, 定义分别为

$$Q_{\pi}(\mathbf{s}, \mathbf{a}) = E_{\pi} \left[\sum_{t=0}^{\infty} \gamma^t R_{t+1} \mid \mathbf{S} = \mathbf{s}, \mathbf{A} = \mathbf{a} \right] \quad (21)$$

$$V_{\pi}(\mathbf{s}) = E_{\pi} \left[\sum_{t=0}^{\infty} \gamma^t R_{t+1} \mid \mathbf{S} = \mathbf{s} \right] \quad (22)$$

式中: $E_{\pi}[\cdot]$ 为策略 π 下累计奖励的期望; R_{t+1} 为 $t+1$ 时刻的即时奖励; $\gamma \in [0, 1]$ 为折扣因子, 表示未来时刻的奖励对当前时刻的重要程度, γ 越大表示未来奖励越重要。

本节采用深度强化学习方法构建 IES 动态调度模型, 包括定义状态空间、动作空间和奖励函数。同时, 为灵活应对源荷的不确定性以及考虑模型求解的便捷性和训练过程的稳定性, 对状态空间、动作空间以及奖励函数进行了改进。

2.1 改进状态空间

本文中的 IES 是智能体的环境, 智能体通过观测 IES 的状态对系统中各设备出力进行最优调度决策, 观测状态包括电负荷、热负荷、光伏出力和电储能的荷电状态, 为充分考虑源荷的不确定性且避免对不确定因素分布知识的复杂建模, 本文对训练数据中的 $P_{\text{load}}(t)$ 、 $H_{\text{load}}(t)$ 和 $P_{\text{PV}}(t)$ 添加随机扰动 N_t 作为新的训练样本, 增强智能体在与环境交互时对环境的探索能力, 从而更好地应对源荷的不确定性。为便于对深度强化学习模型进行求解分析, 其中 N_t 采用高斯分布, 其概率密度函数如式(23)所示。上述随机扰动不仅限于高斯分布这一种形式, 但目的均为训练模型使其学习到不确定性环境的分布特征。

$$f(N_t) = \frac{1}{\sqrt{2\pi}\sigma_N} \exp \left[-\frac{(N_t - \mu_N)^2}{2\sigma_N^2} \right] \quad (23)$$

式中: N_t 为电负荷/热负荷/光伏出力的波动偏差; μ_N 、 σ_N 分别为电负荷/热负荷/光伏出力波动偏差的期望和标准差。

因此状态空间可表示为

$$\mathbf{s}_t = [P_{\text{load}}(t), H_{\text{load}}(t), P_{\text{PV}}(t), C_{\text{SOC}}(t-1)] \quad (24)$$

2.2 改进动作空间

动作空间中的各分量可以由 IES 中各设备的出力表示。由式(2)可知, $P_{\text{CHP}}(t)$ 确定后 $H_{\text{CHP}}(t)$ 可以通过计算确定; 由式(13)可知, $H_{\text{GB}}(t)$ 确定后 $H_{\text{EB}}(t)$ 可以通过计算确定, 进而可以根据式(3)计算确定 $P_{\text{EB}}(t)$; 进一步由式(12)可知, 确定 $P_{\text{BES}}(t)$ 后, $P_{\text{grid}}(t)$ 也可以通过计算确定。因此动作空间可表示为

$$\mathbf{a}_t = [P_{\text{CHP}}(t), H_{\text{GB}}(t), P_{\text{BES}}(t)] \quad (25)$$

为进一步提升对环境的探索能力, 训练过程中

同样对动作空间添加随机扰动,改进后的动作空间为

$$\tilde{\mathbf{a}}_t = \chi \mathbf{a}_t + (1 - \chi) \mathbf{n}_t \quad (26)$$

式中: $\tilde{\mathbf{a}}_t$ 为 t 时刻的实际动作空间; \mathbf{n}_t 为服从 $[-1, 1]$ 范围内截断均匀分布的随机扰动; χ 为初始动作空间各分量所占的比例。随着训练的进行,逐渐增大 χ 值,使动作空间的探索程度逐渐降低, χ 的上限设置为 0.9, 保证训练后期动作空间仍具有探索性。

2.3 改进奖励函数

IES 动态调度的目标是最小化运行总成本,将该目标转化为强化学习奖励最大化的形式;另外,对于智能体的决策动作,即各设备出力应满足设备运行约束,对违反约束的动作添加惩罚项,同时要对源荷功率不平衡造成的功率误差添加惩罚项。因此,奖励函数可以表示为

$$R_t(\mathbf{s}_t, \mathbf{a}_t) = \lambda [F_t(\mathbf{s}_t, \mathbf{a}_t) + \delta_t(\mathbf{s}_t, \mathbf{a}_t) F_B] \quad (27)$$

$$\delta_t(\mathbf{s}_t, \mathbf{a}_t) = |P_{\text{CHP}}(t) - P_{\text{CHP}}| + |H_{\text{GB}}(t) - H_{\text{GB}}| + |P_{\text{BES}}(t) - P_{\text{BES}}| + |H_{\text{EB}}(t) - H_{\text{EB}}| + |P_{\text{VAC}}(t)| + |H_{\text{VAC}}(t)| \quad (28)$$

式中: λ 为奖励函数的缩放系数; $\delta_t(\mathbf{s}_t, \mathbf{a}_t)$ 为 t 时刻的惩罚系数; F_B 为惩罚基值,动作违反约束时 F_B 取一个较大的常数,否则取 0; P_{CHP} 取值如式(29)所示,同理 H_{GB} 、 P_{BES} 和 H_{EB} 分别取各自的功率约束上下限。

$$P_{\text{CHP}} = \begin{cases} P_{\text{CHP}}^{\min}, & P_{\text{CHP}}(t) \leq 0 \\ P_{\text{CHP}}^{\max}, & P_{\text{CHP}}(t) > 0 \end{cases} \quad (29)$$

2.4 最优策略

智能体观测到 IES 某一状态 \mathbf{s}_t 后做出动作 \mathbf{a}_t 的优劣程度可以用状态-价值函数 $V_\pi(\mathbf{s})$ 衡量。

$$V_\pi(\mathbf{s}) = E_\pi \left[\sum_{k=t}^T \gamma^{k-t} R_k(\mathbf{s}_k, \mathbf{a}_k) | \mathbf{s} = \mathbf{s}_t \right] \quad (30)$$

IES 优化调度的目标寻找最优策略 π^* 使得状态-价值函数 $V_\pi(\mathbf{s})$ 最大化,即

$$\pi^* = \arg \max V_\pi(\mathbf{s}) \quad (31)$$

3 基于深度强化学习的 IES 动态调度问题求解方法

传统强化学习方法在处理低维状态空间和动作空间的问题时,具有超参数少、收敛性稳定等优势。随着状态空间和动作空间的维度提升,传统强化学习方法的适用范围受到限制。本文状态空间中的 $P_{\text{load}}(t)$ 、 $H_{\text{load}}(t)$ 、 $P_{\text{PV}}(t)$ 和 $C_{\text{SOC}}(t)$ 以及动作空间中的 $P_{\text{CHP}}(t)$ 、 $H_{\text{GB}}(t)$ 和 $P_{\text{BES}}(t)$ 均为连续量,如果对状

态空间和动作空间进行离散化,会存在维数灾难问题^[21];同时动作空间的离散化会造成关键信息的丢失,传统强化学习无法有效求解。

为此本文采用具有连续决策能力的深度强化学习求解 IES 动态调度问题,具体算法采用 PPO 算法,该算法基于 Actor-Critic 网络架构。由于使用深度神经网络对强化学习进行函数拟合,维数灾难和信息丢失问题被很好地解决。PPO 算法为 On-Policy 方法,网络更新之后的原始数据会被废弃,导致训练过程需要多次采样,降低了数据使用率。针对这一问题,本文在训练过程中使用经验回放机制,利用重要性采样技术限制新旧策略的概率分布差异,将 PPO 算法由 On-Policy 方法转换为 Off-Policy 方法,提高了数据利用率,加快了训练速度。

3.1 PPO 算法原理

PPO 算法的 Actor 网络和 Critic 网络相互独立,拥有不同的损失函数和权重更新方式,其训练过程如图 2 所示。算法训练过程主要包括 3 部分内容:一是 Actor 网络与环境进行交互并将获得的关键信息存储在经验池内;二是利用 Critic 网络计算状态价值和折扣奖励,进而计算优势函数并对 Critic 网络进行权重更新;三是利用采样网络计算重要性采样比,从而对 Actor 网络进行权重更新。其中采样网络权重与 Actor 网络权重完全一致,但更新滞后于 Actor 网络。

3.1.1 Critic 网络训练

对于 Critic 网络,通过优化损失函数 $L(\boldsymbol{\omega})$ 来更新网络权重。

$$L(\boldsymbol{\omega}) = E[y_t - V(\mathbf{s}_t; \boldsymbol{\omega})]^2 \quad (32)$$

$$y_t = R_t + \gamma V(\mathbf{s}_{t+1}; \boldsymbol{\omega}) \quad (33)$$

式中: $L(\boldsymbol{\omega})$ 为 Critic 网络的损失函数; $V(\mathbf{s}_t; \boldsymbol{\omega})$ 为 t 时刻 Critic 网络的输出值; y_t 为 $t+1$ 时刻对 $V(\mathbf{s}_t; \boldsymbol{\omega})$ 做出的估计值; R_t 为 t 时刻的即时奖励; $\boldsymbol{\omega}$ 为 Critic 网络权重。

根据梯度下降算法更新 Critic 网络权重,更新方式为

$$\boldsymbol{\omega} \leftarrow \boldsymbol{\omega} - \alpha \nabla_{\boldsymbol{\omega}} L(\boldsymbol{\omega}) \quad (34)$$

式中: ∇ 为梯度计算函数; α 为 Critic 网络的学习率。

3.1.2 Actor 网络训练

对于 Actor 网络,通过优化损失函数 $J(\boldsymbol{\theta})$ 来更新网络权重。

$$J(\boldsymbol{\theta}) = E_{(\mathbf{s}_t, \mathbf{a}_t): \pi(\cdot; \boldsymbol{\theta})} \left[\min(r_t(\boldsymbol{\theta}) A(\mathbf{s}_t, \mathbf{a}_t), \text{clip}(r_t(\boldsymbol{\theta}), 1 - \varepsilon, 1 + \varepsilon) A(\mathbf{s}_t, \mathbf{a}_t)) \right] \quad (35)$$

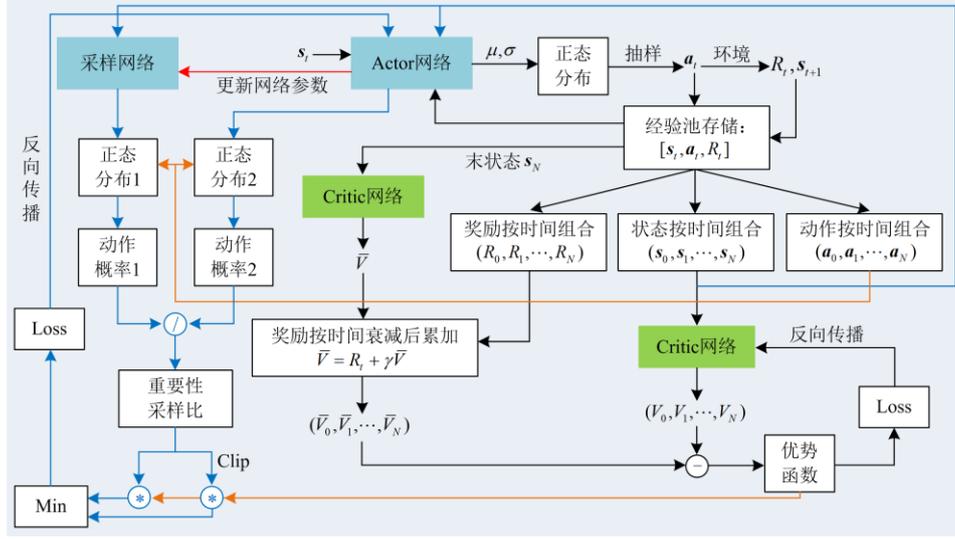


图 2 PPO 算法训练过程示意图

Fig. 2 Schematic diagram of the training process of PPO algorithm

$$r_t(\theta) = \frac{\pi(a_t | s_t; \theta)}{\pi(a_t | s_t; \theta')} \quad (36)$$

$$\text{clip}(x, x^{\min}, x^{\max}) = \begin{cases} x, & x^{\min} \leq x \leq x^{\max} \\ x^{\min}, & x < x^{\min} \\ x^{\max}, & x > x^{\max} \end{cases} \quad (37)$$

式中: $J(\theta)$ 为 Actor 网络的损失函数; $\pi(a_t | s_t; \theta)$ 为 t 时刻 Actor 网络的输出值; $A(s_t, a_t)$ 为优势函数; $r_t(\theta)$ 为 t 时刻重要性采样比; θ 为 Actor 网络权重; ε 是一个超参数, 用来限制 $r_t(\theta)$ 的变化范围。 $J(\theta)$ 随 $r_t(\theta)$ 的变化如图 3 所示。

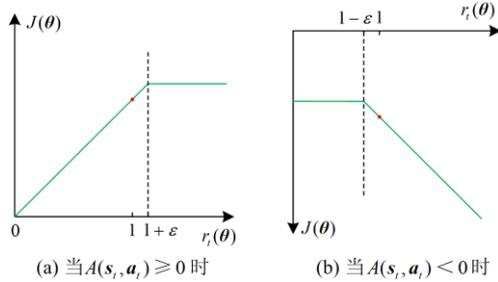


图 3 损失函数与重要性采样比的函数示意图
Fig. 3 Schematic diagram of loss as a function of importance sampling ratio

式(35)中优势函数 $A(s_t, a_t)$ 的定义为

$$A(s_t, a_t) = Q(s_t, a_t) - V(s_t) \quad (38)$$

为提升训练速度, 对 $Q(s_t, a_t)$ 做蒙特卡洛近似, 优势函数转化为

$$A(s_t, a_t) = y_t - V(s_t; \omega) \quad (39)$$

根据梯度上升算法更新 Actor 网络权重, 更新方式为

$$\theta \leftarrow \theta + \beta \nabla_{\theta} J(\theta) \quad (40)$$

式中, β 为 Actor 网络的学习率。

3.2 应用流程

利用深度强化学习方法求解 IES 动态调度问题的框架如图 4 所示。

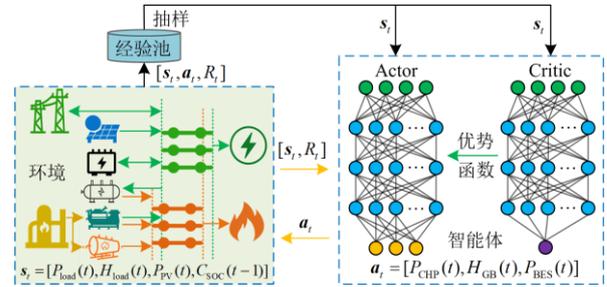


图 4 动态调度问题求解流程

Fig. 4 Solving process of dynamic dispatch problem

Actor 网络和 Critic 网络的输入均为状态 s_t , Actor 网络的输出为动作 a_t , Critic 网络的输出为 V_t 值。训练过程中智能体与环境按时段顺序交互, 智能体获取环境状态 s_t 并做出动作 a_t , 环境返还奖励值 R_t 给智能体。将每个时段的状态 s_t 、动作 a_t 和奖励值 R_t 在经验池中进行存储。

深度强化学习模型的训练结果与环境的初始状态关系密切, 为降低利用训练完成的模型获得调度结果的偶然性, 同时为进一步提升模型在应对源荷

不确定性时的适用性，将模型训练过程中每轮训练的初始状态以随机抽样的方式获得，并从经验池随机抽样小批量样本对智能体的网络权重进行训练更新。利用训练数据对基于 PPO 算法的深度强化学习模型完成离线训练后将模型保存并应用于 IES 的动态调度决策。

4 算例仿真与分析

本节以图 1 所示的 IES 作为算例进行仿真研究，通过对比分析不同时间尺度和不同调度方法下的求解结果，验证本文所提方法的有效性。

4.1 算例仿真设置

IES 中的电负荷、热负荷和光伏出力数据来源于国内某重点项目园区，该数据包含随机性，能充分体现 IES 中源与荷的不确定性。系统内设备的运行参数见表 1，设备的其他参数见表 2^[22-23]。

IES 与主电网交互功率范围为 $[-2.5, 2.5]$ MW，天然气购买量的范围为 $[0, 400]$ m³，电储能的容量为 2 MWh。系统电价采用如表 3 所示的分时电价，其中谷时段为 23:00—07:00，平时段为 07:00—12:00、19:00—23:00，峰时段为 12:00—19:00。天然气单价为固定价格 3.95 元/m³，天然气热值为 9.88 kW/m³，电功率和热功率的供需不平衡单价均为固定价格 0.3 元/kWh。

表 1 IES 设备的运行参数

Table 1 Operational parameters of the IES equipment

设备类型	功率下限/MW	功率上限/MW
热电联产机组	0	1.6
电锅炉	0	1.5
燃气锅炉	0	1.6
电储能	-0.5	0.5

表 2 IES 设备的其他参数

Table 2 Other parameters of the IES equipment

参数	取值	参数	取值
η_{CHP}	0.35	C_{SOC}^{\min}	0.2
b	1.2	C_{SOC}^{\max}	0.8
η_{EB}	0.95	η_{ch}	0.95
η_{GB}	0.8	η_{dis}	0.95
C_{SOC}^{\min}	0.3	ρ_{BES}	0.05

表 3 分时电价

Table 3 Time-of-use electricity price

时段	购电电价/(元/kWh)	售电电价/(元/kWh)
谷时段	0.15	0
平时段	0.8	0.2
峰时段	1.3	0.5

本文提出的深度强化学习方法通过 TensorFlow 平台实现。智能体接收电负荷、热负荷和光伏出力的训练数据后对深度强化学习模型进行训练，训练过程中的超参数依据经验值选取，并根据参考文献和训练过程做动态调整，直至算法收敛至最大奖励值。深度强化学习模型 Actor 网络和 Critic 网络的隐藏层层数均为 3，每层含有 64 个神经元，隐藏层的激活函数均为 ReLU，网络权重更新采用 Adam 优化器。经验池的大小为 1000，小批量样本数为 96，模型训练轮数为 10 000，剩余的其他超参数见表 4。

表 4 其他超参数

Table 4 Other hyperparameters

参数	取值	参数	取值
α	0.002	γ	0.9
β	0.001	λ	0.001
ε	0.2	F_B	6000

4.2 动态调度结果分析

首先选取系统调度总时段为 24 h、时间尺度大小为 1 h 的测试数据进行求解分析，求解结果如图 5 和图 6 所示。

图 5 是模型训练过程中智能体奖励值随训练轮数的变化曲线。从图 5 中可以看出该模型经过约 7000

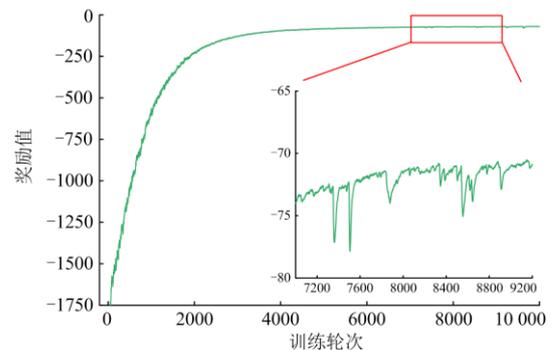
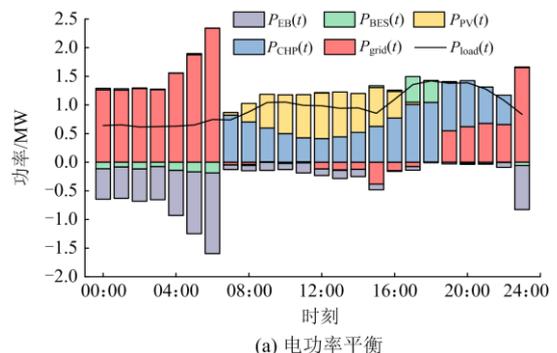


图 5 智能体训练过程的奖励值变化

Fig. 5 Change of reward of the agent in the training process



(a) 电功率平衡

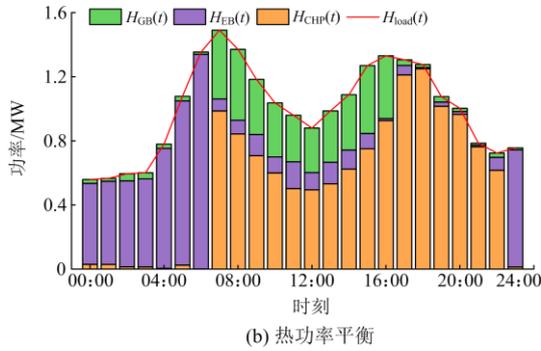


图 6 电功率和热功率的 1 h 调度结果

Fig. 6 Dispatch results of electric power and heat power under dispatch period of 1 hour

轮训练后收敛, 训练初期智能体对环境不熟悉, 智能体做出调度决策时获得的奖励值较小。随着训练的进行, 智能体与环境不断交互并积累经验从而更新网络权重, 因此智能体的奖励值会逐渐增大直至收敛, 训练过程中奖励值的波动性来自于源荷的不确定性。

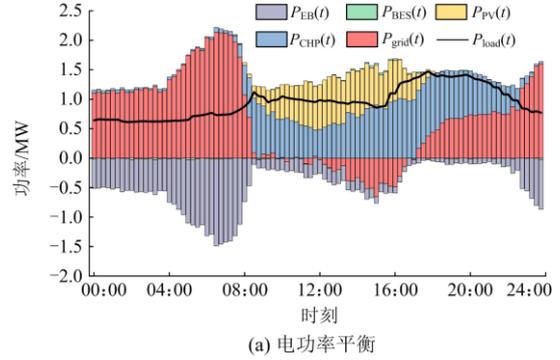
图 6 是时间尺度大小为 1 h 时系统电功率和热功率的动态调度结果。从图 6(a)可以看出, 系统与主电网交互的电功率是跟随电价变化的。在电价的谷时段, 系统从主电网购电满足电负荷需求; 在电价的平时段和峰时段, 电负荷主要由热电联产机组和光伏出力满足, 而且在峰时段系统将热电联产机组多产生的电能售给主电网, 如 12:00—15:00 时段等。电储能设备通过在电价谷时段充电并在电价峰时段放电来降低系统运行成本, 如 04:00—07:00 时段和 17:00—18:00 时段。从图 6(b)可以看出, 电锅炉输出的热功率同样跟随电价变化, 在电价谷时段, 系统热负荷主要通过电锅炉购电制热来满足; 在电价平时段和峰时段, 系统热负荷主要由热电联产机组和燃气锅炉满足。

本文所构建的深度强化学习模型具有良好的扩展性和适应性, 可以方便地应用至不同时间尺度的动态调度问题, 只需在模型训练时调整与环境的交互步长。图 7 展示的是时间尺度为 15 min 时相应的动态调度结果。

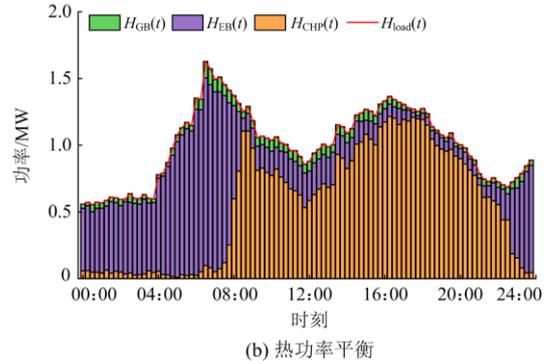
从图 7 可以看出, 系统与主电网交互的电功率和电储能设备的充放电情况基本跟随电价变化。在电价的谷时段, 系统从主电网购电满足电负荷需求, 同时利用电锅炉购电制热满足热负荷需求。在电价的峰时段和光伏出力较大的时段, 系统通过向主电网售电来降低系统运行成本。

为验证本文所提方法能够有效应对源荷的不确

定性, 提出 3 个场景, 对不同场景训练过程中智能体动作 $P_{CHP}(t)$ 的越限情况进行对比分析, 求解结果如图 8 所示。



(a) 电功率平衡



(b) 热功率平衡

图 7 电功率和热功率的 15 min 调度结果

Fig. 7 Dispatch results of electric power and heat power under dispatch period of 15 minutes

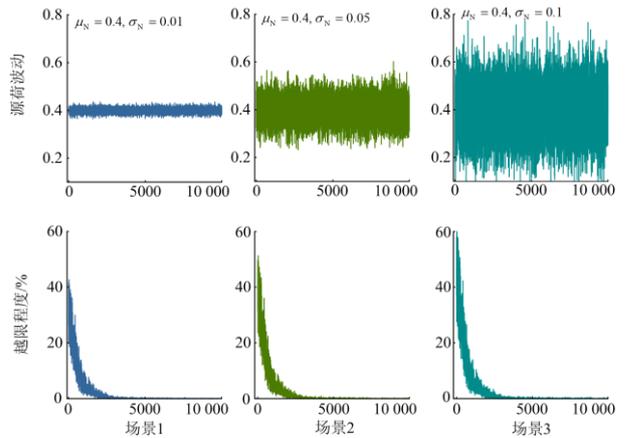


图 8 不同场景下训练过程智能体的动作越限变化

Fig. 8 Change of action of the agent over the limit in the training process under different scenarios

场景 1: 将电负荷、热负荷和光伏出力功率波动偏差的期望均设置为 0.4, 标准差均设置为 0.01。

场景 2: 将电负荷、热负荷和光伏出力功率波动偏差的期望均设置为 0.4, 标准差均设置为 0.05。

场景 3: 将电负荷、热负荷和光伏出力功率波动偏差的期望均设置为 0.4, 标准差均设置为 0.1。

从图 8 可以看出, 随着源荷不确定性增加, 训练过程中智能体动作均被约束在限定范围内, 从而说明深度强化学习模型能够自适应学习系统中源荷的不确定性, 而无需在模型中进行人为假定。

为进一步验证本文所提方法在考虑热力损耗后同样可以自适应调度决策, 将测试数据中的 $H_{load}(t)$ 添加一个增量 Δh 作为等效热力损耗后对模型进行求解分析。限于篇幅, 图 9 展示了考虑热力损耗前后部分设备输出热功率的变化情况。

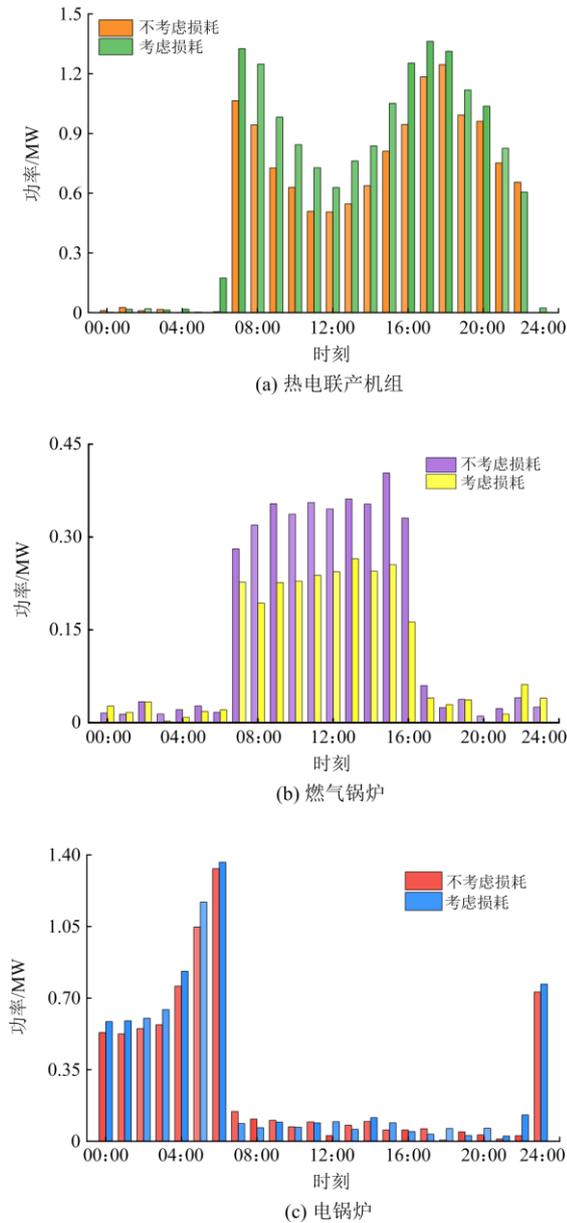


图 9 部分设备的热功率变化
Fig. 9 Change of heat power of parts of equipment

由图 9 可知, 考虑热力损耗之后各类设备能够自适应调整输出功率以满足园区的负荷需求, 在电价的平时段和峰时段, 热电联产机组和燃气锅炉输出热功率的变化较为明显; 在电价的谷时段, 电锅炉输出热功率的变化较为明显, 因此本文提出的方法对实际运行系统具有较好的跟踪学习能力。

4.3 敏感性分析

为分析奖励函数的关键参数对本文所提方法的影响, 选取 6 组典型参数对模型进行训练并求解, 图 10 展示了不同参数下智能体在训练过程中奖励值的变化情况, 图 11 展示了不同参数下热电联产机组电功率的调度结果。

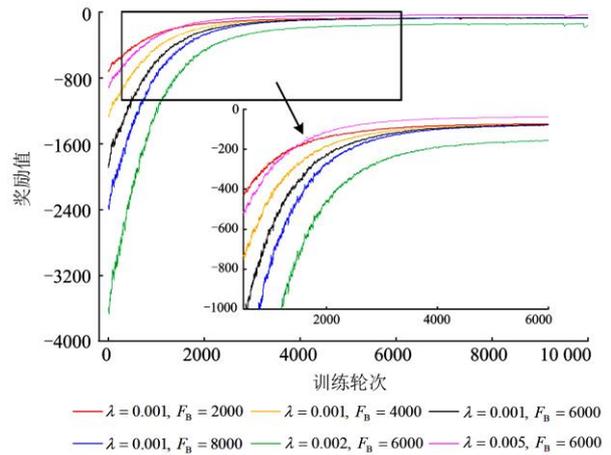


图 10 不同参数下训练过程智能体的奖励值变化
Fig. 10 Change of reward of the agent in the training process under different parameters

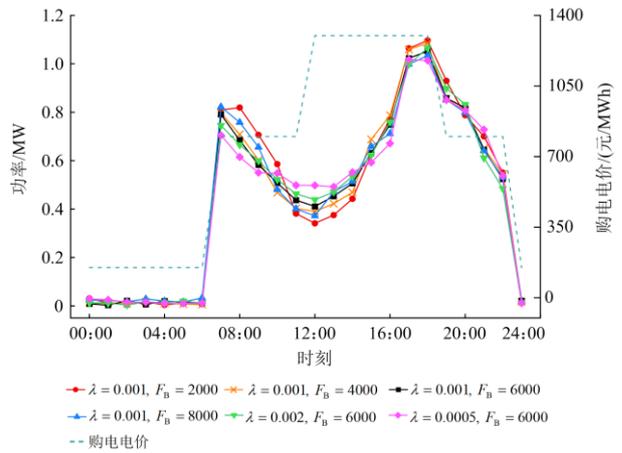


图 11 不同参数下热电联产机组电功率的调度结果
Fig. 11 Dispatch results of electric power of the CHP under different parameters

由图 10 可知, 智能体的奖励值在不同参数下均可以实现较好的收敛。奖励函数的缩放系数 λ 会影

响训练过程收敛值的大小, 但不会影响收敛速度; 而奖励函数的惩罚基值 F_B 会影响训练过程的收敛速度, 但不会影响收敛值的大小。

由图 11 可知, 热电联产机组输出的电功率跟随购电电价的变化而变化。在购电电价的谷时段, 热电联产机组输出的电功率较小, 缩放系数 λ 和惩罚基值 F_B 对输出功率的影响较小; 在购电电价的平时段和峰时段由于电价提高和光伏对系统电功率的补充, 热电联产机组输出功率的变化情况如图中所示, 缩放系数 λ 和惩罚基值 F_B 会影响热电联产机组输出功率在各时段的协调, 但对系统运行总成本维持在较低水平的目标影响较小。

4.4 对比结果分析

为进一步比较本文所提方法与其他调度方法的性能, 将本文所提方法与基于 DDPG 算法的动态调度方法以及传统场景分析方法进行比较。从测试数据集中随机选取 7 个测试日并采用上述 3 种方法对得到的系统运行成本进行对比分析。其中调度总时段为 24 h, 时间尺度大小为 1 h, 传统场景分析方法采用求解软件进行求解, 运行成本的统计结果如表 5 所示。

表 5 不同调度方法的运行成本

运行成本/元	PPO	DDPG	传统场景分析
最大值	32 633	33 170	34 892
最小值	28 982	30 844	31 352
平均值	31 222	31 778	32 327

从表 5 可以看出, 基于 PPO 算法的平均日运行成本为 31 222 元; 基于 DDPG 算法的平均日运行成本为 31 778 元, 较 PPO 算法增加了 1.78%; 基于传统场景分析方法的平均日运行成本为 32 327 元, 较 PPO 算法增加了 3.54%。从日运行成本的最大值、最小值以及平均值来看, 基于 PPO 算法的动态调度方法较其他两种调度方法获得了更好的经济性能, 能有效地降低系统运行成本。

5 结论

本文提出了一种基于深度强化学习的 IES 动态调度方法, 该方法将 IES 动态调度问题转化为具有连续状态和连续动作的强化学习问题来处理, 有效解决了维数灾难和信息缺失而造成的模型求解困难问题, 同时避免了对源荷的不确定性进行详细建模分析。

算例结果表明, 本文所提方法能够较好地应对不确定性环境造成光伏出力变化和负荷变化的随机性; 同时所构建的深度强化学习模型在不同时间尺度下

具有较好的扩展性和适应性。与其他优化调度方法对比, 本文所提方法在收敛性和经济性方面均具有较好的表现。

本文的研究重点在于满足能源供需平衡的前提下合理安排机组出力以保证 IES 的经济性, 缺乏对实现调度决策机理的解释, 未来将以数据驱动与物理模型相融合的方式对决策机理做进一步研究。

参考文献

- [1] 余晓丹, 徐宪东, 陈硕翼, 等. 综合能源系统与能源互联网综述[J]. 电工技术学报, 2016, 31(1): 1-13.
YU Xiaodan, XU Xiandong, CHEN Shuoyi, et al. A brief review to integrated energy system and energy internet[J]. Transactions of China Electrotechnical Society, 2016, 31(1): 1-13.
- [2] WANG Y L, WANG Y D, HUANG Y J, et al. Optimal scheduling of the regional integrated energy system considering economy and environment[J]. IEEE Transactions on Sustainable Energy, 2019, 10(4): 1939-1949.
- [3] 贾宏杰, 穆云飞, 余晓丹. 对我国综合能源系统发展的思考[J]. 电力建设, 2015, 36(1): 16-25.
JIA Hongjie, MU Yunfei, YU Xiaodan. Thought about the integrated energy system in China[J]. Electric Power Construction, 2015, 36(1): 16-25.
- [4] KESHAVARZZADEH A H, AHMADI P. Multi-objective techno-economic optimization of a solar based integrated energy system using various optimization methods[J]. Energy Conversion and Management, 2019, 196: 196-210.
- [5] 郑国太, 李昊, 赵宝国, 等. 基于供需能量平衡的用户侧综合能源系统电/热储能设备综合优化配置[J]. 电力系统保护与控制, 2018, 46(16): 8-18.
ZHNEG Guotai, LI Hao, ZHAO Baoguo, et al. Comprehensive optimization of electrical/thermal energy storage equipments for integrated energy system near user side based on energy supply and demand balance[J]. Power System Protection and Control, 2018, 46(16): 8-18.
- [6] 帅挽澜, 朱自伟, 李雪萌, 等. 考虑风电消纳的综合能源系统“源-网-荷-储”协同优化运行[J]. 电力系统保护与控制, 2021, 49(19): 18-26.
SHUAI Wanlan, ZHU Ziwei, LI Xuemeng, et al. “Source-network-load-storage” coordinated optimization operation for an integrated energy system considering wind power consumption[J]. Power System Protection and Control, 2021, 49(19): 18-26.
- [7] 刘洪, 陈星屹, 李吉峰, 等. 基于改进 CPSO 算法的区域电热综合能源系统经济调度[J]. 电力自动化设备, 2017, 37(6): 193-200.
LIU Hong, CHEN Xingyi, LI Jifeng, et al. Economic dispatch based on improved CPSO algorithm for regional power-heat integrated energy system[J]. Electric Power Automation Equipment, 2017, 37(6): 193-200.

- [8] 崔杨, 闫石, 仲伟之, 等. 含电转气的区域综合能源系统热电优化调度[J]. 电网技术, 2020, 44(11): 4254-4264.
CUI Yang, YAN Shi, ZHONG Wuzhi, et al. Optimal thermoelectric dispatching of regional integrated energy system with power-to-gas[J]. Power System Technology, 2020, 44(11): 4254-4264.
- [9] 袁泉, 吴云亮, 李豹, 等. 计及源荷不确定性的多时间尺度滚动调度计划模型与方法[J]. 电力系统保护与控制, 2019, 47(16): 8-16.
YUAN Quan, WU Yunliang, LI Bao, et al. Multi-timescale coordinated dispatch model and approach considering generation and load uncertainty[J]. Power System Protection and Control, 2019, 47(16): 8-16.
- [10] 张大海, 负韞韵, 王小君, 等. 考虑广义储能及光热电站的电热气互联综合能源系统经济调度[J]. 电力系统自动化, 2021, 45(19): 33-42.
ZHANG Dahai, YUN Yunyun, WANG Xiaojun, et al. Economic dispatch of integrated electricity-heat-gas energy system considering generalized energy storage and concentrating solar power plant[J]. Automation of Electric Power Systems, 2021, 45(19): 33-42.
- [11] 郭尊, 李庚银, 周明, 等. 考虑网络约束和源荷不确定性的区域综合能源系统两阶段鲁棒优化调度[J]. 电网技术, 2019, 43(9): 3090-3100.
GUO Zun, LI Gengyin, ZHOU Ming, et al. Two-stage robust optimal scheduling of regional integrated energy system considering network constraints and uncertainties in source and load[J]. Power System Technology, 2019, 43(9): 3090-3100.
- [12] 马国真, 林毓军, 张泽亚, 等. 计及源荷多重不确定性的综合能源系统鲁棒经济调度方法[J]. 电力系统保护与控制, 2021, 49(20): 43-52.
MA Guozhen, LIN Yujun, ZHANG Zeya, et al. A robust economic dispatch method for an integrated energy system considering multiple uncertainties of source and load[J]. Power System Protection and Control, 2021, 49(20): 43-52.
- [13] 王成山, 吕超贤, 李鹏, 等. 园区型综合能源系统多时间尺度模型预测优化调度[J]. 中国电机工程学报, 2019, 39(23): 6791-6803, 7093.
WANG Chengshan, LÜ Chaoxian, LI Peng, et al. Multiple time-scale optimal scheduling of community integrated energy system based on model predictive control[J]. Proceedings of the CSEE, 2019, 39(23): 6791-6803, 7093.
- [14] 乔骥, 王新迎, 张擎, 等. 基于柔性行动器-评判器深度强化学习的电-气综合能源系统优化调度[J]. 中国电机工程学报, 2021, 41(3): 819-833.
QIAO Ji, WANG Xinying, ZHANG Qing, et al. Optimal dispatch of integrated electricity-gas system with soft actor-critic deep reinforcement learning[J]. Proceedings of the CSEE, 2021, 41(3): 819-833.
- [15] 王新迎, 赵琦, 赵黎媛, 等. 基于深度 Q 学习的电热综合能源系统能量管理[J]. 电力建设, 2021, 42(3): 10-18.
WANG Xinying, ZHAO Qi, ZHAO Liyuan, et al. Energy management approach for integrated electricity-heat energy system based on deep Q-learning network[J]. Electric Power Construction, 2021, 42(3): 10-18.
- [16] NAUG A, AHMED I, BISWAS G. Online energy management in commercial buildings using deep reinforcement learning[C] // 2019 IEEE International Conference on Smart Computing (SMARTCOMP), June 12-15, 2019, Washington DC, USA: 249-257.
- [17] SCHREIBER T, ESCHWEILER S, BARANSKI M, et al. Application of two promising reinforcement learning algorithms for load shifting in a cooling supply system[J]. Energy and Buildings, 2020, 229.
- [18] 王磊, 姜涛, 宋丹, 等. 基于灵活热电比的区域综合能源系统多目标优化调度[J]. 电力系统保护与控制, 2021, 49(8): 151-159.
WANG Lei, JIANG Tao, SONG Dan, et al. Multi-objective optimal dispatch of a regional integrated energy system based on a flexible heat-to-electric ratio[J]. Power System Protection and Control, 2021, 49(8): 151-159.
- [19] YU L, XIE W W, XIE D, et al. Deep reinforcement learning for smart home energy management[J]. IEEE Internet of Things Journal, 2020, 7(4): 2751-2762.
- [20] PERERA A T D, WICKRAMASINGHE P U, NIK V M, et al. Introducing reinforcement learning to the energy system design process[J]. Applied Energy, 2020, 262.
- [21] YANG T, ZHAO L Y, LI W, et al. Reinforcement learning in sustainable energy and electric systems: a survey[J]. Annual Reviews in Control, 2020, 49: 145-163.
- [22] 杨挺, 赵黎媛, 刘亚闯, 等. 基于深度强化学习的综合能源系统动态经济调度[J]. 电力系统自动化, 2021, 45(5): 39-47.
YANG Ting, ZHAO Liyuan, LIU Yachuang, et al. Dynamic economic dispatch for integrated energy system based on deep reinforcement learning[J]. Automation of Electric Power Systems, 2021, 45(5): 39-47.
- [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.

收稿日期: 2021-12-10; 修回日期: 2022-01-24

作者简介:

蔺伟山(1997—), 男, 硕士研究生, 研究方向为人工智能在综合能源系统中的应用; E-mail: 20121455@bjtu.edu.cn

王小君(1978—), 男, 通信作者, 博士, 教授, 研究方向为电力系统分析与控制、综合能源系统优化运行; E-mail: xjwang1@bjtu.edu.cn

孙庆凯(1992—), 男, 博士研究生, 研究方向为人工智能在综合能源系统中的应用。E-mail: sunqingkai_123@163.com

(编辑 周金梅)