控制与决策
Control and Decision

基于深度强化学习的微电网在线优化调度

季颖, 王建辉

引用本文:
季颖,王建辉. 基于深度强化学习的微电网在线优化调度[J]. 控制与决策, 2022, 37(7): 1675−1684.

在线阅读 View online: https://doi.org/10.13195/j.kzyjc.2021.0835

基于深度强化学习的微电网在线优化调度

季 颖, 王建辉†

(东北大学 信息科学与工程学院，沈阳 110004)

摘 要: 提出一种基于深度强化学习的微电网在线优化调度策略. 针对可再生能源的随机性及复杂的潮流约束对微电网经济安全运行带来的挑战,以成本最小为目标,考虑微电网运行状态及调度动作的约束,将微电网在线调度问题建模为一个约束马尔可夫决策过程. 为避免求解复杂的非线性潮流优化、降低对高精度预测信息及系统模型的依赖,设计一个卷积神经网络结构学习最优的调度策略. 所提出的神经网络结构可以从微电网原始观测数据中提取高质量的特征,并基于提取到的特征直接产生调度决策. 为了确保该神经网络产生的调度决策能够满足复杂的网络潮流约束,结合拉格朗日乘子法与soft actor-critic,提出一种新的深度强化学习算法来训练该神经网络. 最后,为验证所提出方法的有效性,利用真实的电力系统数据进行仿真. 仿真结果表明,所提出的在线优化调度方法可以有效地从数据中学习到满足潮流约束且具有成本效益的调度策略,降低随机性对微电网运行的影响.

关键词: 微电网；约束马尔可夫过程；深度强化学习；卷积神经网络

中图分类号: TP273 **文献标志码:** A

DOI: 10.13195/j.kzyjc.2021.0835

引用格式: 季颖,王建辉. 基于深度强化学习的微电网在线优化调度[J]. 控制与决策, 2022, 37(7): 1675-1684.

Online optimal scheduling of a microgrid based on deep reinforcement learning

JI Ying, WANG Jian-hui†

(College of Information Science and Engineering，Northeastern University，Shenyang 110004，China)

Abstract: This paper proposes an online scheduling strategy based on deep reinforcement learning (DRL). To overcome the challenges in economic and safe operation of microgrids posed by uncertain renewable energy resources and complex power flow constraints, in this paper, we formulate the microgrid online scheduling problem as a constrained Markov decision process (CMDP) with the objective of operating cost minimization while considering the constraints on the operating states and scheduling actions. To avoid solving complicated nonlinear optimal power flow and reduce the dependency on accurate forecasting information and system model, we design a convolutional neural network (CNN) architecture to learn the optimal scheduling policy. The neural network can extract high-quality features from the original observation data of the microgrid and directly make scheduling decisions based on the extracted features. To ensure the satisfaction of complex power flow constraints, we propose a novel DRL algorithm by combining the Lagrange multiplier method and the soft actor-critic algorithm to train the neural network. To verify the effectiveness of the proposed approach, we use real-world power system data to perform simulation studies. Simulation results demonstrate that the proposed online scheduling optimization approach can effectively learn a cost-effective scheduling strategy that satisfies power flow constraints, mitigating the effect of randomness on microgrids.

Keywords: microgrid; constrained Markov decision process; deep reinforcement learning; convolutional neural network

0 引 言

近年来,可再生能源的广泛接入极大地促进了电力系统的能源结构向清洁化、可持续化发展. 据统计[1],截至2019年,我国可再生能源发电量达2.04万亿千瓦时,占全部发电量的比重为27.9 %. 受能源供给侧结构改革及2030年碳达峰目标的驱动,电网中可再生能源的比例将进一步增加. 然而,风电、光伏等分布式可再生能源受自然条件影响,存在较大的波动性、间歇性,对大规模可再生能源并网消纳提出了巨大挑战.

微电网作为解决可再生能源消纳的有效方法,近年来受到广泛关注,其可通过先进的信息控制技术整

收稿日期: 2021-05-12；**录用日期:** 2021-08-09.
基金项目: 国家自然科学基金项目(61733003).
†通讯作者. E-mail: wangjianhui@ise.neu.edu.cn.

合分布式发电单元及可再生能源,为小规模区域内的电力用户提供可靠的能源供给,并实现经济效益与环境效益的最大化. 然而,不同于传统大电网,微电网的运行调度面临诸多挑战: 1) 间歇性可再生能源所具有的不确定性使微电网难以基于预测制定准确的日前调度计划,造成供需不平衡问题; 2) 分布式可再生能源接入导致的双向潮流易造成母线电压波动,引起微电网运行的稳定性问题.

为了解决随机环境下微电网的能量调度问题,许多研究提出了基于模型的在线优化调度方法. 文献[2]针对可再生能源和负荷的不确定性,提出了孤岛型的微电网鲁棒优化算法. 文献[3]建立了储能系统的混合整数线性规划模型,并对储能系统的充放电调度进行实时的滚动优化. 文献[4]提出了基于一种模型预测控制的微电网日前与日内滚动校正相结合的多时间尺度协调调度方法. 文献[5]提出了一种考虑可再生能源、系统负荷和电价等不确定性的两阶段随机规划模型,并用模型预测控制策略进行求解. 这些方法都依赖于准确的预测模型估计可再生能源的发电量,然而,由于可再生能源的不确定性,很多实际的微电网系统很难得到一个准确的预测模型. 此外,为了建立能量优化调度模型,这些方法也需要知道传输线参数及微电网潮流的物理模型. 因此,当预测模型不准确或者物理模型参数不确定时,这些方法的性能可能会受到影响.

为了减少对于模型的依赖,近年来许多学者提出基于数据驱动的强化学习方法以解决微电网的在线能量优化调度问题. 例如,文献[6]提出了一种基于Q学习的微电网经济调度算法,考虑可再生能源、电动汽车与储能的相互协调,以及系统的负荷波动和碳排放问题. 文献[7]提出了一种基于Q学习的微电网智能建筑群能量管理算法,以此减少运行成本. 文献[8]采用基于多智能体的强化学习解决微电网中的分布式能源管理问题. 文献[9]采用批量强化学习方法求解微电网的电池能量管理问题. 上述文献采用的大多数基于强化学习的方法都会受到维数灾的影响,并且难以处理具有高维状态变量和不确定性的微电网系统.

为了解决微电网中高维状态空间表征的问题,许多学者提出基于深度强化学习的能量调度方法. 文献[10]将基于深度Q网络(DQN)的方法应用于微电网的实时能量优化调度,该方法考虑了负荷需求和电价的不确定性. 文献[11]提出了考虑不确定性的深度期望Q学习算法来求解微电网的实时优化问题. 文献[12]考虑到未来电力消耗和光伏发电的不确定性,将深度强化学习应用于微电网储能设备的高效运行中. 文献[13]将深度Q学习应用到电池能源管理问题中,并利用卷积神经网络学习历史电价下的最优充电计划. 尽管上述基于学习的研究方法极大地减少了对于模型的依赖,但这些研究并没有考虑微电网的网架结构及对应的潮流约束. 因此,这些方法所产生的调度策略可能无法满足实际微电网系统中的运行要求,导致潮流过载、母线电压波动,甚至运行稳定性等问题.

针对上述问题,本文考虑微电网的潮流约束及可再生能源的不确定性,提出一种基于深度强化学习的微电网在线优化调度方法. 针对微电网在线优化多阶段决策的特点,首先构建约束马尔可夫决策过程(CMDP)模型,避免传统MDP模型难以处理约束的缺点; 然后设计一种卷积神经网络结构,通过从微电网的原始观测数据中提取高质量特征来近似最优的调度策略; 最后以soft-actor-critic (SAC)为基础提出拉格朗日soft-actor-critic (LSAC)强化学习算法训练该神经网络,从而获得近似最优的在线调度策略.

1 问题描述

1.1 考虑AC潮流约束的微电网优化调度问题

微电网优化调度通过调节分布式燃料机组输出的有功/无功功率及储能系统的充/放电功率,在满足分布式发电设备运行约束及AC潮流约束的前提下,最小化未来T个时段内微电网运行的期望成本. 其目标函数可表示为

$$\min E\Big[\sum_{t=0}^{T-1}\Big(\sum_{i\in\mathcal{N}}[C_d^i(P_{d,t}^i)+C_g(P_{g,t})]\Big)\Big], \tag{1a}$$

$$C_d^i(P_{d,t}^i)=(a_d^i\cdot(P_{d,t}^i)^2+b_d^i\cdot P_{d,t}^i+c_d^i)\Delta t, \tag{1b}$$

$$C_g(P_{g,t})=1_{\boldsymbol{R}^+}(P_{g,t})c_t^{\rm buy}|P_{g,t}|\Delta t- 1_{\boldsymbol{R}^-}(P_{g,t})c_t^{\rm sell}|P_{g,t}|\Delta t. \tag{1c}$$

式(1a)中: $C_d^i(\cdot)$为第i个母线上分布式燃料机组的发电成本,可以表示为该机组有功功率的二次函数(1b); $C_g(\cdot)$为微电网与大电网的能量交易成本,由式(1c)进行计算. 式(1b)中: $P_{d,t}^i$为第i个母线上的分布式燃料机组在t时段的输出功率, a_d^i、b_d^i、c_d^i为对应的发电成本曲线系数. 式(1c)中: $c_t^{\rm buy}$为从大电网购电的价格; $c_t^{\rm sell}$为向大电网售电的价格; $P_{g,t}$为微电网在t时段与大电网的交换功率,正值表示微电网从大电网购电,负值表示微电网向大电网售电; $1_{\boldsymbol{R}^+}$、$1_{\boldsymbol{R}^-}$为指示函数,表示为

$$1_{\boldsymbol{R}^+}(x)=\begin{cases}1,\ x>0;\\0,\text{otherwise.}\end{cases}$$

$$1_{\boldsymbol{R}^-}(x)=\begin{cases}1,\ x<0;\\0,\ \text{otherwise.}\end{cases}$$

同时,微电网的优化调度需满足如下约束:

1) AC潮流约束

$$\begin{aligned}&\underline{V}^i\leqslant V_t^i\leqslant\overline{V}^i,\ 0\leqslant I_t^{ij}\leqslant\overline{I}^{ij},\\&P_t^i=V_t^i\sum_{k\in\mathcal{N}}V_t^kY^{ik}\cos(\theta_t^{ik}+\delta_t^k-\delta_t^i),\\&Q_t^i=-V_t^i\sum_{k\in\mathcal{N}}V_t^kY^{ik}\sin(\theta_t^{ik}+\delta_t^k-\delta_t^i).\end{aligned}\tag{2}$$

其中:V_t^i 为 t 时段第 i 个母线上的电压,$\underline{V}^i$ 和 $\overline{V}^i$ 分别为其下限和上限约束;I_t^{ij} 为 t 时段传输线 ij 上的电流,$\overline{I}^{ij}$ 为其上限约束;δ_t^i 和 δ_t^k 分别为第 i 个和第 k 个母线上的相角;θ_t^{ik} 为母线 i 与 k 之间的导线阻抗角;Y^{ik} 为网络导纳矩阵的第 i 行第 k 列元素;P_t^i 和 Q_t^i 为第 i 个母线注入微电网的有功及无功功率,P_t^i 和 Q_t^i 同时满足

$$\begin{aligned}&P_t^i=P_{d,t}^i+P_{\mathrm{pv},t}^i+P_{\mathrm{wt},t}^i-P_{e,t}^i-P_{l,t}^i,\\&Q_t^i=Q_{d,t}^i-Q_{l,t}^i,\end{aligned}\tag{3}$$

$P_{\mathrm{pv},t}^i$、$P_{\mathrm{wt},t}^i$ 分别为第 i 个母线上光伏发电的有功功率、风力发电的有功功率,$P_{e,t}^i$ 为第 i 个母线上的储能设备在 t 时段的充/放电功率,$P_{l,t}^i$、$Q_{l,t}^i$ 为第 i 个母线上负荷的有功及无功功率,$Q_{d,t}^i$ 为第 i 个母线上分布式燃料机组在 t 时段输出的无功功率. 为了充分消纳可再生能源,假设光伏、风电以单位功率因数运行,不考虑其无功功率.

2) 分布式燃料机组运行约束

$$\begin{aligned}&\underline{P}_d^i\leqslant P_{d,t}^i\leqslant\overline{P}_d^i,\\&(P_{d,t}^i)^2+(Q_{d,t}^i)^2=(S_{d,t}^i)^2\leqslant(\overline{S}_d^i)^2.\end{aligned}\tag{4}$$

其中:$\underline{P}_d^i$ 和 $\overline{P}_d^i$ 分别为第 i 个母线上分布式燃料机组的有功输出下限及上限,$S_{d,t}^i$ 为该发电机在 t 时段输出的视在功率,$\overline{S}_d^i$ 为发电机的额定容量.

3) 储能系统运行约束

$$-\overline{P}_e^i\leqslant P_{e,t}^i\leqslant\overline{P}_e^i,\ \underline{\mathrm{SoC}}^i\leqslant\mathrm{SoC}_t^i\leqslant\overline{\mathrm{SoC}}^i.$$

$$\mathrm{SoC}_{t+1}^i=\begin{cases}\mathrm{SoC}_t^i+P_{e,t}^i\Delta t\cdot\eta_{\mathrm{ch}}^i,\ P_{e,t}^i>0;\\\mathrm{SoC}_t^i+P_{e,t}^i\Delta t/\eta_{\mathrm{dch}}^i,\ \text{otherwise.}\end{cases}\tag{5}$$

其中:$\overline{P}_e^i$ 为第 i 个母线上储能系统的最大充/放电功率;SoC_t^i 为 t 时段该储能的荷电状态,其上下限约束分别为 $\overline{\mathrm{SoC}}^i$ 和 $\underline{\mathrm{SoC}}^i$;$\eta_{\mathrm{ch}}^i$ 和 η_{dch}^i 分别为相应的充、放电效率.

4) 与大电网功率交换约束

$$(S_{g,t})^2\leqslant(\overline{S}_g)^2.\tag{6}$$

其中:$S_{g,t}$ 为微电网在 t 时段公共耦合点的视在功率,$\overline{S}_g$ 为其上限.

为了减少随机性对微电网优化运行的影响,传统基于模型的微电网在线调度方法需要根据未来 T 时段可再生能源出力及负荷的预测值,求解由式 (1)～(6) 组成的优化模型. 然而,由于AC潮流方程的存在,优化模型 (1)～(6) 具有很强的非线性,难以在有限的调度周期内求得最优解. 现有的基于强化学习的方法为了保证学习效果,也往往忽略系统的AC潮流,使得产生的调度策略难以应用到实际系统中. 为了解决上述问题,本文考虑将AC潮流约束的微电网优化调度问题建模为约束马尔可夫决策过程(CMDP),所提出的调度模型不需要可再生能源出力及负荷的预测信息,也不依赖于AC潮流的具体模型.

1.2　基于CMDP的在线调度模型

本节将微电网的在线优化调度问题建模成一个CMDP,目标是通过找到分布式燃料机组和储能的最优调度策略,使发电机组的日运行成本最小. 首先,介绍微电网在线优化调度问题的传统马尔可夫决策过程(MDP);然后,讨论传统MDP的局限性;最后,提出一个用辅助成本函数处理约束的CMDP建模方法.

1.2.1　传统的MDP建模方法

微电网在线优化调度问题传统上建模成一个MDP,用五元组 $(\mathcal{S},\mathcal{A},\mathcal{P},\mathcal{R},\gamma)$ 表示. 其中:$\mathcal{S}$ 为系统的状态集,$\mathcal{A}$ 为系统的动作集,$\mathcal{P}:\mathcal{S}\times\mathcal{A}\times\mathcal{S}\to[0,1]$ 为状态转移概率,$\mathcal{R}:\mathcal{S}\times\mathcal{A}\to\boldsymbol{R}$ 为奖励函数,$\gamma\in[0,1)$ 为折扣因子.

1) 状态变量.

t 时段的微电网系统状态 $s_t\in\mathcal{S}$ 可以描述为

$$\begin{aligned}&s_t=(P_T^1,\dots,P_T^N,Q_T^1,\dots,Q_T^N,c_T,\mathrm{SoC}_t);\\&P_T^i=(P_{t-\tau}^i,\dots,P_{t-1}^i),\ \forall i\in\mathcal{N};\\&Q_T^i=(Q_{t-\tau}^i,\dots,Q_{t-1}^i),\ \forall i\in\mathcal{N};\\&c_T=(c_{t-T}^{\mathrm{buy}},\dots,c_{t-1}^{\mathrm{buy}},c_{t-T}^{\mathrm{sell}},\dots,c_{t-1}^{\mathrm{sell}});\\&\mathrm{SoC}_t=(\mathrm{SoC}_t^1,\dots,\mathrm{SoC}_t^N).\end{aligned}\tag{7}$$

其中:P_T^i 为第 i 个母线上过去 T 个时段的有功功率,Q_T^i 为第 i 个母线上过去 T 个时段的无功功率,c_T 为过去 T 个时段大电网的购买/出售电价,SoC_t 为 t 时段所有母线上储能的荷电状态.

2) 动作变量.

根据 t 时段的微电网系统状态 s_t,微电网系统调度分布式燃料机组、储能单元和大电网. 微电网系统

的动作变量$a_t \in \mathcal{A}$描述为

$$a_t = (P^i_{d,t}, Q^i_{d,t}, P^i_{e,t}, P_{g,t}, Q_{g,t}),\ \forall i \in \mathcal{N}. \quad (8)$$

其中:$P^i_{d,t}$、$Q^i_{d,t}$为t时段第i个母线上分布式燃料机组的有功和无功功率,$P^i_{e,t}$为t时段第i个母线上储能系统的充/放电功率,$P_{g,t}$、$Q_{g,t}$为t时段微电网与大电网交换的有功和无功功率.

3) 转移概率.

状态转移概率$\mathcal{P}: \mathcal{S} \times \mathcal{A} \times \mathcal{S} \to [0,1]$表示为

$$\mathcal{P}^a_{ss'} = \Pr\{s_{t+1} = s' | s_t = s, s_t = a\}. \quad (9)$$

由于受到可再生能源及负荷不确定性的影响,状态转移概率(9)很难用精确的概率分布模型描述. 为了解决该问题,本文采用深度强化学习算法,从历史数据中隐式地学习这个分布.

4) 奖励函数.

为了最小化微电网系统的运行成本,定义时段t的奖励值r_t为总运行成本的负值,即

$$r_t(s_t, a_t) = -\Big[\sum_{i\in\mathcal{N}}(C^i_d(P^i_{d,t}) + C_g(P_{g,t}))\Big]. \quad (10)$$

5) 目标函数.

本文的目标是找到最优调度策略$\pi(a_t|s_t): s_t \to a_t$,最大化未来$T$个调度周期的期望奖励值(即总运行成本最小). 目标函数为

$$\max_{\pi\in\Pi} J(\pi) = E_{\tau\sim\pi}\Big[\sum_{t=0}^{T-1}\gamma^t \cdot r_t(s_t, a_t)\Big], \quad (11)$$

其中$\gamma \in [0,1)$为折扣因子.

1.2.2 处理约束条件的难点

在传统的MDP框架中,为了处理约束条件1)~4),必须引入一个惩罚项限制约束. 这种情况下目标函数将变成

$$\max_{\varpi\in\Pi} J(\pi) + \varrho \cdot \mathrm{Penalty}(\pi). \quad (12)$$

其中:$\mathrm{Penalty}(\pi)$为惩罚函数,ϱ为惩罚系数. 在实际应用中,惩罚系数ϱ很难确定. 如果惩罚系数设置得过小,则违反约束的行为不能受到足够的惩罚,导致调度决策不可行,从而危及微电网的稳定运行;相反,如果惩罚系数设置过大,则违反约束的行为可能会受到过度惩罚,从而导致调度决策缺少经济效益. 为了找到合适的惩罚系数,算法设计过程中往往需要通过试错不断地调整参数,增加了算法实施的难度.

1.2.3 CMDP

为了避免调整惩罚系数,本文采用CMDP对微电网在线调度问题进行建模. CMDP利用一个辅助成本函数扩充MDP模型,有

$$\begin{aligned} c_t(s_t, a_t) = {} & \sum_{i\in\mathcal{N}} \max(\max(0, V^i_t - \overline{V}^i), \underline{V}^i - V^i_t) + \\ & \sum_{ij\in\mathcal{M}} \max(0, I^{ij}_t/\overline{I}^{ij} - 1) + \max(0, S_{g,t}/\overline{S}_g - 1) + \\ & \sum_{i\in\mathcal{N}} \max(\max(0, \mathrm{SoC}^i_t - \overline{\mathrm{SoC}}^i), \underline{\mathrm{SoC}}^i - \mathrm{SoC}^i_t) + \\ & \sum_{i\in\mathcal{N}} \max(0, S^i_{d,t}/\overline{S}^i_d - 1). \end{aligned} \quad (13)$$

其中:第1项计算母线电压过载值,第2项计算配电线路过流值,第3项计算大电网与微电网间交换功率超过额定容量的部分,第4项评估储能系统荷电状态违反约束的情况,第5项计算分布式燃料机组发电量超出其额定容量的部分.

定义$J_C(\pi)$为与策略π有关的辅助成本函数的预期折扣收益,可表示为

$$J_C(\pi) = E_{\tau\sim\pi}\Big[\sum_{t=0}^{T-1}\gamma^t \cdot c_t(s_t, a_t)\Big]. \quad (14)$$

MDP可以扩充为如下CMDP:

$$\begin{aligned} & \max_{\pi\in\Pi}\ J(\pi) = E_{\tau\sim\pi}\Big[\sum_{t=0}^{T-1}\gamma^t \cdot r_t(s_t, a_t)\Big]; \\ & \text{s.t. } J_C(\pi) \leqslant d. \end{aligned} \quad (15)$$

其中d为容忍因子,可以将约束值限制在一个非常小的范围内. 在CMDP表达式(15)中,微电网的运行约束受到$J_C(\pi) \leqslant d$的严格限制,因此不需要调整惩罚系数.

2 求解算法

2.1 SAC (soft actor critic)算法原理

SAC是一种最大熵强化学习算法,与传统强化学习算法不同之处在于,其采用一种基于策略熵正则化的目标函数改进策略更新,从而提高训练过程的鲁棒性. 因此,其目标函数为

$$\tilde{J}(\pi) = \sum_{t=0}^{T-1} E_{(s_t,a_t)\sim\rho_\pi}\gamma^t[r(s_t, a_t) + \alpha\mathcal{H}(\pi(\cdot|s_t))]. \quad (16)$$

其中:$\mathcal{H}(\pi(\cdot|s_t)) = -\sum_{a_t}\pi(a_t|s_t)\log\pi(a_t|s_t)$为策略熵;$\alpha$为温度参数,控制策略的探索能力. 值得指出,当策略$\pi(a_t|s_t)$收敛到确定的最优策略时,策略熵$\mathcal{H}(\pi(\cdot|s_t))$的值为0,因此目标(16)与MDP的目标函数(11)等价. 由于策略熵的引入,状态-动作值函数的贝尔曼方程可以表示为如下正则化形式:

$$Q^\pi(s_t, a_t) = r(s_t, a_t) + \gamma E_{s_{t+1}\sim p}[V^\pi(s_{t+1})], \quad (17)$$

其中

$$V^\pi(s_t) = E_{a_t\sim\pi}[Q^\pi(s_t, a_t) - \log\pi(a_t|s_t)]. \quad (18)$$

为了求解最优策略$\pi^*(a_t|s_t)$, SAC采用策略迭代方法交替执行策略评估及策略改进.

策略评估为

$$Q^{\pi_{\rm new}}(s_t,a_t)=r(s_t,a_t)+\gamma E_{s_{t+1}\sim p}[V^{\pi_{\rm old}}(s_{t+1})]. \tag{19}$$

策略改进为

$$\pi_{\rm new}=\arg\min_{\pi}\overline{D}_{\rm KL}\Big(\pi(\cdot|s_t)\Big\|\frac{\exp(Q^{\pi_{\rm old}}(s_t,\cdot))}{Z^{\pi_{\rm old}}(s_t)}\Big). \tag{20}$$

其中: $\overline{D}_{\rm KL}(\cdot||s_t)=E_{s_t}D_{\rm KL}(\cdot||s_t)$为Kullback-Leibler散度的期望, $Z^{\pi_{\rm old}}(s_t)$为归一化项.

2.2 Lagrangian-SAC(LSAC)算法

尽管传统的SAC算法已经成功地应用于许多MDP问题,但不适合解决带约束的MDP. 因此,结合拉格朗日乘子法与SAC提出LSAC算法以解决考虑AC潮流约束的微电网在线优化调度问题.

LSAC算法将约束马尔可夫决策问题转化为如下极小-极大问题:

$$\min_{\pi}\max_{\lambda\geqslant 0}J^{\mathcal{L}}(\pi,\lambda)=-\tilde{J}(\pi)+\lambda\cdot g(\pi). \tag{21}$$

其中: $\tilde{J}(\pi)$为式(16)定义的目标函数, λ为拉格朗日乘子, $g(\pi)=J_C(\pi)-d$. 当存在策略π使得$g(\pi)$严格满足不等式约束,即$g(\pi)=J_C(\pi)-d<0$时,强对偶性成立. 根据对偶原理,当$\lambda\to\lambda^*$收敛到对偶最优时,策略$\pi\to\pi^*$收敛到问题(21)的最优解.

为了求解上述极小-极大问题,采用如下原始-对偶方法进行策略改进:

$$\begin{aligned}\pi_{\rm new}&=\arg\min_{\pi}J^{\mathcal{L}}_{\pi_{\rm old}}(\pi,\lambda_{\rm old}),\\ \lambda_{\rm new}&=\arg\max_{\lambda\geqslant 0}J^{\mathcal{L}}_{\pi_{\rm new}}(\pi_{\rm new},\lambda),\end{aligned} \tag{22}$$

其中$J^{\mathcal{L}}_{\pi'}(\pi,\lambda)$为拉格朗日函数$J^{\mathcal{L}}(\pi,\lambda)$在$\pi'$附近的近似,具体表示为

$$J^{\mathcal{L}}_{\pi'}(\pi,\lambda)=\overline{D}_{\rm KL}\Big(\pi(\cdot|s_t)\Big\|\frac{\exp(Q^{\pi'}(s_t,\cdot))}{\exp(\lambda Q^{\pi'}_C(s_t,\cdot)-\lambda d)}\Big), \tag{23}$$

$Q^{\pi'}_C(s_t,a_t)$为关于约束成本$c_t(s_t,a_t)$的状态-动作值函数, 表示为

$$\begin{aligned}Q^{\pi'}_C(s_t,a_t)&=c_t(s_t,a_t)+\gamma E_{s_{t+1}\sim p}[V^{\pi'}_C(s_t)],\\ V^{\pi'}_C(s_t)&=E_{a_t,s_{t+1},\cdots}\Big[\sum_{l=t}^{T-1}\gamma^{l-t}c_l(s_l,a_l)\Big].\end{aligned} \tag{24}$$

根据Kullback-Leibler散度的定义,目标(23)可以展开为如下形式:

$$\begin{aligned}&J^{\mathcal{L}}_{\pi'}(\pi,\lambda)=\\ &\overline{D}_{\rm KL}(\pi(\cdot|s_t)||\exp(Q^{\pi'}(s_t,\cdot)-\lambda Q^{\pi'}_C(s_t,\cdot)+\lambda d))=\\ &E_{s_t\sim\rho_\pi,a_t\sim\pi}[\log\pi(a_t|s_t)-Q^{\pi'}(s_t,a_t)+\\ &\lambda Q^{\pi'}_C(s_t,a_t)-\lambda d],\end{aligned} \tag{25}$$

其中$Q^{\pi'}(s_t,a_t)$、$Q^{\pi'}_C(s_t,a_t)$分别由式(17)和(24)计算. 当$\pi=\pi'$时,利用式(18)和(25)可以得到

$$J^{\mathcal{L}}_{\pi'}(\pi,\lambda)=-\tilde{J}(\pi)+\lambda J_C(\pi)-\lambda d=J^{\mathcal{L}}(\pi,\lambda). \tag{26}$$

2.3 基于卷积神经网络近似的LSAC算法

为了处理微电网在线优化问题中的高维连续状态空间和动作空间,采用参数化的函数结构近似策略函数$\pi(a_t|s_t)$、价值函数$V^\pi(s_t)$和$V^\pi_C(s_t)$、状态-动作值函数$Q^\pi(s_t,a_t)$和$Q^\pi_C(s_t,a_t)$.

有效的近似结构能够从微电网原始测量数据中提取高质量特征. 对于微电网在线优化调度问题,高质量的特征应包括系统的网络结构特征、可再生能源发电的时序特征以及微电网的运行状态特征. 为了高效地提取这些特征,设计一个基于卷积神经网络的参数化近似结构,如图1所示.

由图1可见,该卷积神经网络的输入为一个$T\times(1+N)\times 2$的张量$\mathcal{X}$,表示为

$$\mathcal{X}[:,:,0]=\begin{bmatrix}c^{\rm buy}_{t-T}&\cdots&P^2_{t-T}&P^1_{t-T}\\ \vdots&\vdots&\cdots&\vdots\\ c^{\rm buy}_{t-1}&\cdots&P^2_{t-1}&P^1_{t-1}\end{bmatrix}, \tag{27}$$

$$\mathcal{X}[:,:,1]=\begin{bmatrix}c^{\rm sell}_{t-T}&\cdots&Q^2_{t-T}&Q^1_{t-T}\\ \vdots&\vdots&\cdots&\vdots\\ c^{\rm sell}_{t-1}&\cdots&Q^2_{t-1}&Q^1_{t-1}\end{bmatrix}. \tag{28}$$

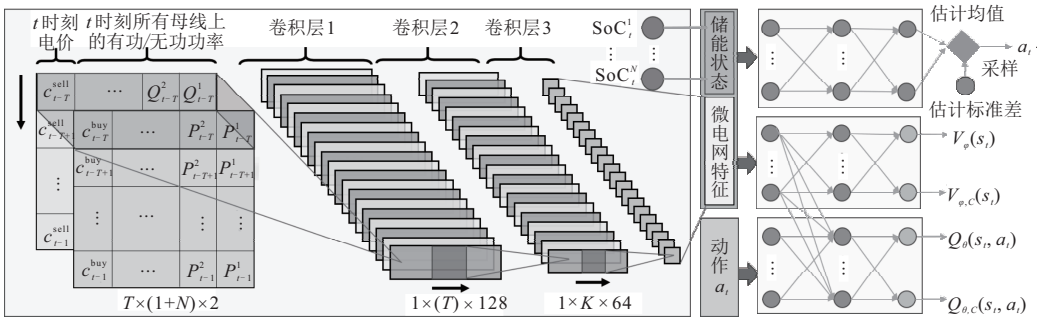


图 1　基于卷积神经网络的参数化近似结构

其中: $\mathcal{X}[:,:,0]$ 包括过去 T 个时段内从大电网购电的价格及所有母线上的有功功率, $\mathcal{X}[:,:,1]$ 包括过去 T 个时段内从大电网售电的价格及所有母线上的无功功率. 为了提取微电网的空间结构特征, 该卷积神经网络采用一个维度为 $1\times(1+N)$ 的核函数将每一个时段 t 内所有母线上的有功、无功信息映射到一个维度为 $1\times T\times 128$ 的卷积层(卷积层 1). 为了提取微电网中历史数据的时序特征, 在卷积层 1 之后, 该网络采用 $1\times T/K$ 的核函数将过去 T 个时段的特征映射到卷积层 $2(1\times K\times 64)$; 然后, 卷积层 2 中的特征通过一个维度为 $1\times K$ 的核函数映射到卷积层 $3(1\times 1\times 64)$; 最后, 该卷积神经网络将卷积层 3 中的特征输出作为微电网的时空特征.

基于该卷积网络提取的微电网时空特征, 将其与储能当前的荷电状态合并, 并将合并后的向量作为输入, 采用一个全连接神经网络近似最优策略和价值函数. 由于微电网调度中的控制变量都是连续的, 采用高斯策略函数, 即 $\pi_\phi(a_t|s_t)\sim\mathcal{N}(\mu_\phi|\Sigma_\phi)$. 通过神经网络近似其中的均值 μ_ϕ 和标准差 Σ_ϕ, 其表达式为

$$\begin{aligned}&[\mu_\phi,\Sigma_\phi]=\boldsymbol{W}_L^\phi\cdot f_L^\phi(s_t)+\boldsymbol{B}_L^\phi;\\&f_l^\phi(s_t)=\mathrm{ReLU}(\boldsymbol{W}_l^\phi\cdot f_{l-1}^\phi(s_t)+\boldsymbol{B}_l^\phi),\ l=1,2,\dots,L;\\&f_0^\phi(s_t)=s_t.\end{aligned}\tag{29}$$

其中: $\boldsymbol{W}_l^\phi$、$\boldsymbol{B}_l^\phi$ 为第 l 层神经网络的权重参数, ReLU 为 Rectified Linear Units 激活函数. 对于价值函数, 采用一个与式(29)结构相同的全连接神经网络进行近似, 近似的价值函数表示为 $[V_\psi(s_t),V_{\psi,C}(s_t)]$, 其中 ψ 为该神经网络的参数. 此外, 采用另一个与式(29)结构相同的全连接神经网络近似状态-动作值函数, 该神经网络的输入为上述两部分特征与动作 a_t 合并的向量, 表示为

$$\begin{aligned}&[Q_\theta(s_t,a_t),Q_{\theta,C}(s_t,a_t)]=\boldsymbol{W}_L^\theta\cdot f_L^\theta(s_t,a_t)+B_L^\theta;\\&f_l^\theta(s_t,a_t)=\mathrm{ReLU}\,(\boldsymbol{W}_l^\theta\cdot f_{l-1}^\theta(s_t,a_t)+\boldsymbol{B}_l^\theta),\\&l=1,2,\dots,L,\ f_0^\theta(s_t,a_t)=[s_t,a_t].\end{aligned}\tag{30}$$

其中 $\boldsymbol{W}_l^\theta$、$\boldsymbol{B}_l^\theta$ 为第 l 层神经网络的权重参数.

$V_\psi(s_t)$ 和 $V_{\psi,C}(s_t)$ 分别为奖励函数 $r_t(s_t,a_t)$ 和约束函数 $c_t(s_t,a_t)$ 的近似值函数. 该近似值函数神经网络可以通过最小化如下残差平方进行训练:

$$\begin{aligned}&J_V(\psi)=\\&\frac{1}{2}E_{s_t\sim\mathcal{D}}[(V_{\psi,C}(s_t)-E_{a_t\sim\pi_\phi}[Q_{\theta,C}(s_t,a_t)])^2+\\&(V_\psi(s_t)-E_{a_t\sim\pi_\phi}[Q_\theta(s_t,a_t)-\log\pi_\phi(a_t|s_t)])^2].\end{aligned}\tag{31}$$

其中: $\mathcal{D}$ 为经验放回池, $\pi_\phi(a_t|s_t)$ 为基于神经网络的策略函数, ϕ 为该策略网络的参数, $Q_\theta(s_t,a_t)$ 和 $Q_{\theta,C}(s_t,a_t)$ 分别为关于奖励函数 $r_t(s_t,a_t)$ 和约束函数 $c_t(s_t,a_t)$ 的近似状态-动作值函数.

状态-动作值函数网络通过最小化如下贝尔曼残差平方进行训练:

$$\begin{aligned}&J_Q(\theta)=\\&\frac{1}{2}E_{(s_t,a_t)\sim\mathcal{D}}[(Q_{\theta,C}(s_t,a_t)-\hat{Q}_C(s_t,a_t))^2+\\&(Q_\theta(s_t,a_t)-\hat{Q}(s_t,a_t))^2],\end{aligned}\tag{32}$$

$$\hat{Q}_C(s_t,a_t)=c_t(s_t,a_t)+\gamma E_{s_{t+1}\sim p}[V_{\bar{\psi},C}(s_t)],\tag{33}$$

$$\hat{Q}(s_t,a_t)=r_t(s_t,a_t)+\gamma E_{s_{t+1}\sim p}[V_{\bar{\psi}}(s_t)].\tag{34}$$

其中: $r_t(s_t,a_t)$ 为式(10)定义的奖励函数, $c_t(s_t,a_t)$ 为式(13)定义的约束成本.

在训练状态-动作值函数网络时, 采用另一个具有相同结构的目标值网络 $V_{\bar{\psi}}$、$V_{\bar{\psi},C}$ 产生样本数据, 以保证训练的稳定性, 其中目标值网络的参数 $\bar{\psi}$ 为值网络参数 ψ 历史值的滑动平均.

策略网络 π_ϕ 和拉格朗日系数 λ 分别通过最小、最大化如下目标函数进行训练:

$$J_\pi(\phi,\lambda)=\overline{D}_{\mathrm{KL}}\left(\pi_\phi(\cdot|s_t)\middle\|\frac{\exp(Q_\theta(s_t,\cdot))}{\exp(\lambda Q_{\theta,C}(s_t,\cdot)-\lambda d)}\right).\tag{35}$$

本文采用随机梯度下降的方法对值网络、状态-动作值函数网络、策略网络和拉格朗日系数进行优化. 为了避免样本的相关性, 每次迭代更新时从经验放回池 $\mathcal{D}$ 中随机采样以往的状态转移样本.

3 实验结果与分析

3.1 实验设置

为了对所提出方法进行有效性验证, 以欧洲 CIGRE 低压微电网系统[14] 为实验对象, 采用 pandapower 进行仿真, 该微电网结构如图 2 所示. 微

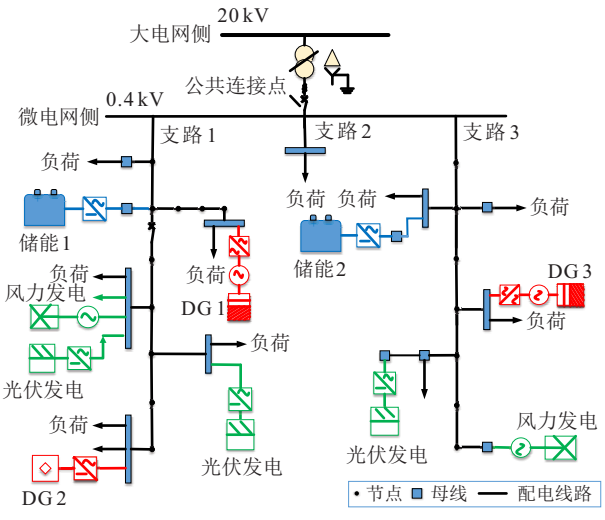


图 2 微电网结构

电网系统包括3个分布式燃料机组(DG)、2个电池储能、3个光伏发电单元、2个风力发电机和若干负荷. 对于每个光伏和风力发电机,其装机容量均为10 kW. 微电网与大电网之间的交换功率上限为500 kVA,其他发电机组的运行参数由表1给出,线路阻值参见文献[14].

表 1　微电网中分布式发电设备的运行参数

DG1	有功功率下限 $\underline{P}_d^1$	有功功率上限 $\overline{P}_d^1$	额定容量 $\overline{S}_d^1$	a_d^1	b_d^1	c_d^1
	0 kW	30 kW	37.5 kVA	0.000 1 \$/kW^{2}h	0.075 7 \$/kWh	0.471 2 \$/h
DG2	有功功率下限 $\underline{P}_d^2$	有功功率上限 $\overline{P}_d^2$	额定容量 $\overline{S}_d^2$	a_d^2	b_d^2	c_d^2
	0 kW	50 kW	62.5 kVA	0.000 1 \$/kW^{2}h	0.051 6 \$/kWh	0.501 1 \$/h
DG3	有功功率下限 $\underline{P}_d^3$	有功功率上限 $\overline{P}_d^3$	额定容量 $\overline{S}_d^3$	a_d^3	b_d^3	c_d^3
	0 kW	40 kW	50 kVA	0.000 1 \$/kW^{2}h	0.062 4 \$/kWh	0.461 5 \$/h
储能1	荷电状态下限 $\underline{\mathrm{SoC}}^1$	荷电状态上限 $\overline{\mathrm{SoC}}^1$	最大充/放电功率 $\overline{P}_e^1$	充电效率 η_{ch}^1	放电效率 η_{dch}^1	–
	0.1	1	60 kW	0.98	0.98	–
储能2	荷电状态下限 $\underline{\mathrm{SoC}}^2$	荷电状态上限 $\overline{\mathrm{SoC}}^2$	最大充/放电功率 $\overline{P}_e^2$	充电效率 η_{ch}^2	放电效率 η_{dch}^2	–
	0.1	1	60 kW	0.98	0.98	–

对于所提出的卷积神经网络,卷积层1的维度为$1\times24\times128$,卷积层2的维度为$1\times12\times64$,卷积层3的维度为$1\times1\times64$. 策略神经网络和值函数神经网络的维度为$(64,64)$,状态-动作值函数神经网络的维度为$(128,128)$. 对于所有隐藏层,采用ReLU神经元. 算法中使用的其他参数如表2所示. 本文所有实验在python环境中运行,采用TensorFlow 2.2.0执行神经网络训练,计算单元为8-core i7-6700K CPU.

表 2　基于LSAC算法的微电网在线调度方法参数

参数	值
调度时长 T	24
Batch size	256
λ 初始值	0
温度参数 α	0.02
折扣因子 γ	0.995
约束容忍因子 d	1e-3
目标值网络软更新参数 τ	5e-3
状态-动作值网络迭代步长 α_Q	1e-3
值网络迭代步长 α_V	5e-4
策略网络迭代步长 α_π	1e-3
λ 迭代步长 α_λ	5e-4
经验放回池	50 000

3.2　训练集和测试集

本文采用美国加州电力系统CASIO[15]真实的运行数据来模拟可再生能源发电、负荷和电价的不确定性,应用2018年~2019年CASIO的光伏、风电、负荷和电价数据作为训练集对所提出神经网络模型进行训练. 在模型训练结束后,应用2020年的数据作为测试集对训练好的模型进行测试. 为促进本地可再生能源的利用,假设出售给大电网的电价为购买价格的80%.

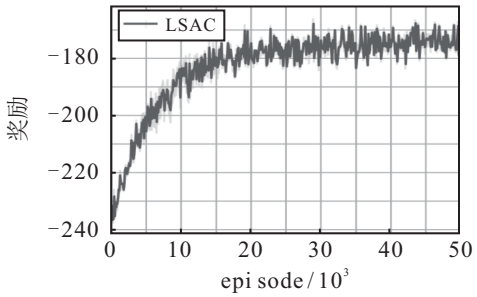
3.3　情景1:考虑DC潮流

为了与基于模型的方法进行对比,验证所提出学习算法的有效性,首先考虑DC潮流情景. 该情景不考虑复杂的非线性AC潮流约束,而是采用简化的DC潮流模型对微电网进行仿真. 该情景下,微电网优化调度可以建模成一个混合整数二次规划问题,从而便于用现有的优化工具箱进行求解.

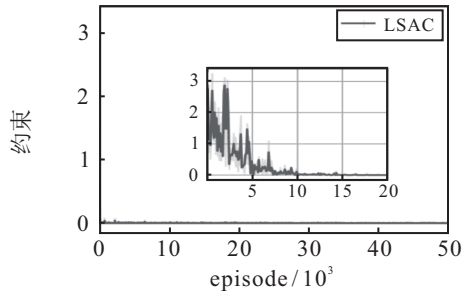
针对该情景设计3种对比策略: 1)基于完美预测信息的最优调度; 2)模型预测控制; 3)贪婪算法. 策略1): 假设微电网中的随机性可以准确地预测,并基于完美的预测信息求解最优的调度决策. 策略2): 采用在线预测和滚动优化制定实时调度决策. 本文假设其预测误差服从正态分布$\mathcal{N}(0,0.1^2)$,优化窗口为6 h. 策略3): 采用短视的贪婪策略,即不考虑当前决策对未来的影响,仅执行一步优化,因此该策略无法充分利用储能来帮助微电网降低成本.

图3展示了所提出LSAC算法在离线训练过程中的学习曲线. 由图3(a)可见,在经过约20 000个episodes训练后,约束违反值从最初的3降低到0附近. 这表明,在考虑DC潮流情景下,所提出算法可以迅速学到可行的微电网调度策略. 另外,由图3(b)可见,随着训练的进行, 奖励曲线快速升高,在经过约25 000个episodes训练后从-230逐渐收敛到-170左右. 这些训练结果表明,LSAC算法可以通过数据学习有效地改进调度策略. 离线训练结束后,训练好的神经网络模型可直接根据微电网的实时状态产生在线调度决策. 图3(c)比较了所提出在线调度策略与其他3种对比策略在测试集上的累积运行成本. 可以看出,与贪婪算法和模型预测控制相比,LSAC算法能获得更低的累积运行成本,且更接近完美信息下的最

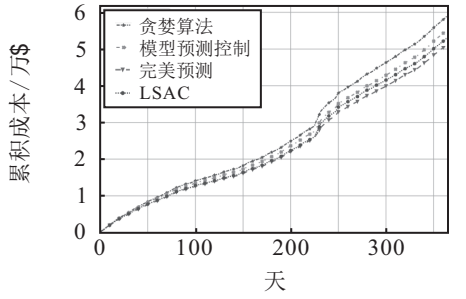
优策略. 这4种策略在测试集上的平均日运行成本分别为\$161.84(贪婪算法)、\$151.64(模型预测控制)、\$140.59(完美预测)、\$145.51(LSAC). 相对于贪婪算法和模型预测控制,LSAC算法可降低10.08 %和4.04 %的运行成本. 另外,LSAC算法仅比完美预测下最优策略的运行成本高3.38 %. 值得指出,由于不确定性的存在,完美预测下的最优策略在实际中不可能实现. 这些实验结果表明,在考虑DC潮流情景下,所提出算法可以学习到近似最优的微电网调度策略,有效地降低微电网的运行成本.



(a) 奖励曲线(训练集)



(b) 约束曲线(训练集)



(c) 测试集

图 3 考虑DC潮流时LSAC算法的学习曲线及不同调度策略在测试集上的累积运行成本

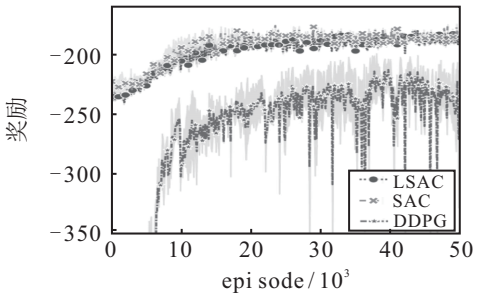
3.4 情景2:考虑AC潮流

考虑AC潮流时,基于模型的方法需要处理复杂的高维多周期非线性约束优化,导致其难以给出收敛的可行解. 为了验证所提出学习算法在考虑AC潮流情景下的有效性,与SAC和DDPG两种强化学习算法进行对比. 由于SAC和DDPG无法求解带约束的MDP,为了使得二者能够处理AC潮流约束,本文采用罚函数法,即在奖励函数中加入约束违反的惩罚项,有

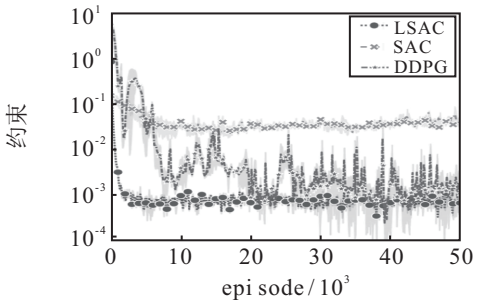
$$\bar{r}_t(s_t,a_t)=r_t(s_t,a_t)+\omega\times c_t(s_t,a_t). \tag{36}$$

罚函数法的一个缺点是惩罚系数ω难以确定:当ω值较大时,算法可能由于约束惩罚过大而损害性能;当ω值较小时,算法可能由于约束惩罚不足导致优化结果不可行. 而LSAC算法可根据拉格朗日法自动处理约束. 为了平衡奖励函数和约束函数,在对比实验中选择$\omega=1\,000$. 为了保证对比的公平性,这些对比方法都采用与所提出算法相同的网络结构.

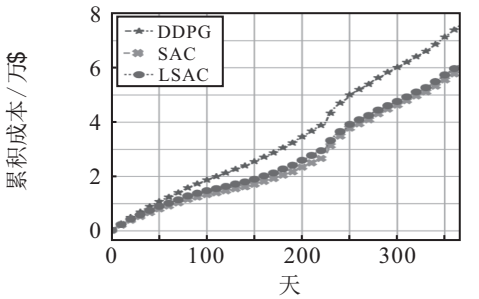
图4(a)和(b)对比了不同学习算法在离线训练过程中的学习曲线. 由图4(a)可见,相对于SAC和DDPG,LSAC算法可以有效地处理AC潮流约束,在



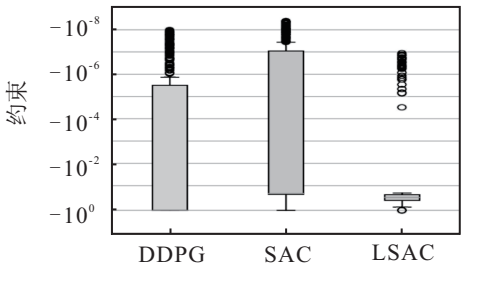
(a) 奖励曲线(训练集)



(b) 约束曲线(训练集)



(c) 累积成本(测试集)



(d) 约束(测试集)

图 4 考虑AC潮流时各算法的学习曲线、测试集上的累积运行成本及日约束违反值在测试集上的分布

10 000个训练episodes内即将约束违反值降低到1e-3以下. 而SAC和DDPG算法则无法有效控制约束, 导致其约束违反值过高且随训练样本变化较大. 这是由于在SAC和DDPG算法中, 惩罚系数ω需要人工设定, 其设定值往往不是最优的. 由图4(b)可见, 尽管SAC算法获得了更高的奖励, 但由于其约束违反值过高, 导致学习到的调度策略无法保证AC潮流约束, 给算法的实际应用带来困难. 对于DDPG算法, 其训练曲线波动较大, 尽管在某些训练样本上约束违反值更低, 但获得的奖励总体上低于LSAC算法.

图4(c)比较了不同学习算法获得的在线调度策略在测试集上的累积运行成本. 由图4(c)可见, 与DDPG算法相比, SAC和LSAC算法能获得更低的累积运行成本. 这3种学习算法在测试集上的平均日运行成本分别为\$205.95(DDPG)、\$161.06(SAC)、\$166.48(LSAC). 尽管SAC算法获得的运行成本比LSAC算法更低, 但其无法保证满足微电网的AC潮流约束. 图4(d)给出了3种算法日累积约束违反值在测试集上的分布. 由图4(d)可见, 相对于SAC算法, LSAC算法具有更小的均值和方差, 且LSAC算法有更少的离群点, 其值也相对较低. 这些对比结果表明, LSAC算法可以通过学习获得满足AC潮流约束的调度策略, 且对于测试集上不同日期的随机性, 表现出更好的稳定性和鲁棒性. 此外, 在保证AC潮流约束的前提下, 所提出学习算法也可以获得具有成本效益的在线调度策略.

为进一步验证所提出方法的有效性, 选取测试集(共366天)中负荷较高的2天, 对所提出方法获得的调度结果进行分析. 图5展示了该时间段内微电网出力情况及调度决策. 由图5(a)和(c)可见, 当大电网电价较高时(如15～20小时、40～44小时), 分布式发电机输出功率增加, 微电网从大电网购电量减少; 当大电网电价较低时(如0～5小时、22～33小时), 分布式发电机输出功率减少, 微电网从大电网购电量相对较高. 这表明所提出调度策略可以根据电价的波动调整分布式发电机的输出功率, 从而相应地增加或减少从大电网的购电量, 以降低运行成本. 图5(d)和(e)展示了储能1和储能2的充/放电调度结果. 可以看出, 在16～21小时期间, 储能1和储能2同时放电, 荷电状态(SoC)迅速降低, 并在21小时左右到达SoC下限附近. 这是由于在此期间, 可再生能源发电量迅速减少(图5(b)), 而微电网负荷和大电网电价仍然较高(图5(a)), 所提出调度策略控制储能设备放电, 保证微电网供电可靠性, 并避免从大电网高价购电. 相

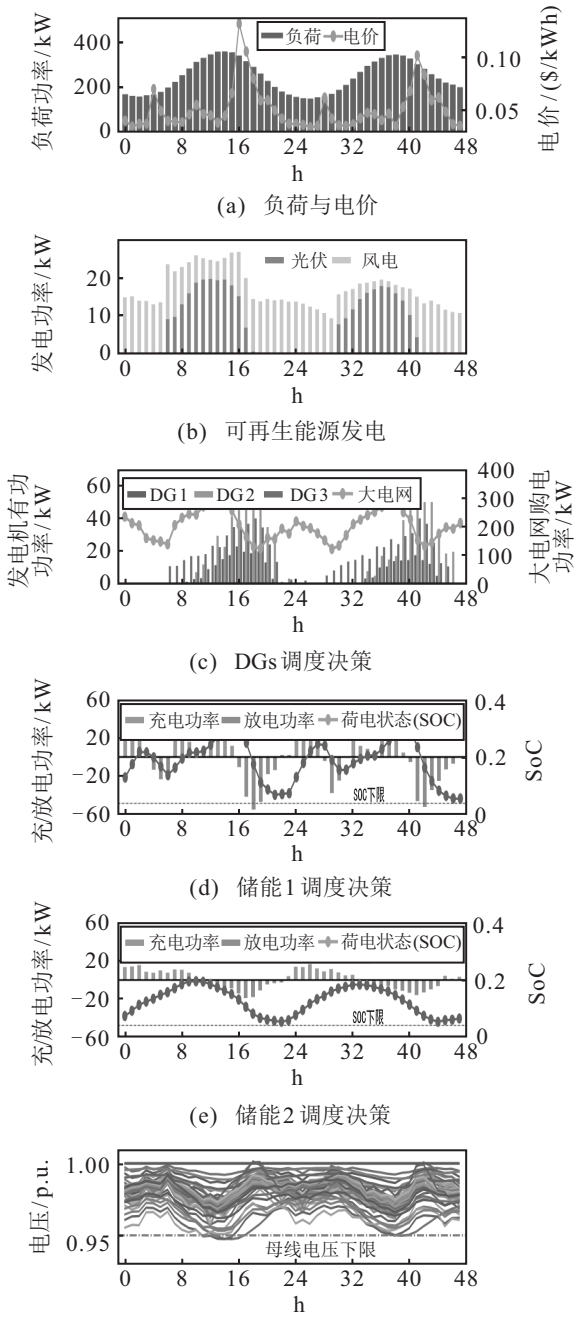


图5　测试集中负荷较高的2天内微电网出力情况及调度结果

似的调度模式也见于40～44小时. 此外, 为保证充足的电量, 储能设备大多在负荷较少、电价较低的时段进行充电, 如0～3小时、24～27小时, 从而降低充电成本. 图5(f)展示了微电网内各母线/节点的电压曲线. 可以看出, 各母线/节点的电压总体保持在安全区间内, 即0.95p.u～1.05p.u., 仅有个别母线在负荷高峰期(13～16小时)时的电压值稍低于0.95. 值得指出, 所提出调度方法仅根据微电网当前的运行状态制定实时的在线调度决策, 完全不需要可再生能源及负荷的预测信息或微电网的潮流模型. 这些结果表明了所提出调度方法对可再生能源的随机波动具有一定

的鲁棒性,可以有效降低微电网的运行成本,提高微电网运行的经济性和安全性.

4 结 论

本文提出了一种基于深度强化学习的微电网在线优化调度方法,考虑AC潮流约束和可再生能源的不确定性,将微电网的在线优化调度问题建模成一个CMDP,该建模方法不需要精确的系统模型和预测信息,避免了传统MDP模型难以处理约束的缺点. 为了求解已建立的CMDP模型,设计了一种卷积神经网络学习最优调度策略. 该神经网络可从微电网的原始观测数据中提取高质量的特征,并基于提取到的特征直接生成调度决策. 为了保证生成的调度决策能够满足复杂的网络潮流约束,基于拉格朗日乘子法与SAC算法,提出了一种新的深度强化学习算法训练该神经网络. 为了验证所提出方法的有效性,利用真实的电力系统数据进行仿真. 仿真结果表明,所提出方法在考虑AC潮流约束的情境下,优于现有的深度强化学习算法,包括DDPG和SAC.

参考文献(References)

[1] 国家能源局. 国家能源局2020年一季度网上新闻发布会文字实录[DB/OL].(2020-03-06)[2021-05-02]. www.nea.gov.cn/2020-03/06/c138850234.htm.

[2] 王栋, 郑鹏远, 任祎丹, 等. 不确定性环境下的孤岛型微电网鲁棒优化算法[J]. 现代电力, 2021, 38(2): 147-155.
(Wang D, Zheng P Y, Ren Y D, et al. Robust optimization algorithm for islanded microgrid in uncertain environment[J]. Modern Electric Power, 2021, 38(2): 147-155.)

[3] Valencia F, Collado J, Sáez D, et al. Robust energy management system for a microgrid based on a fuzzy prediction interval model[J]. IEEE Transactions on Smart Grid, 2016, 7(3): 1486-1494.

[4] 肖浩, 裴玮, 孔力. 基于模型预测控制的微电网多时间尺度协调优化调度[J]. 电力系统自动化, 2016, 40(18): 7-14.
(Xiao H, Pei W, Kong L. Multi-time scale coordinated optimal dispatch of microgrid based on model predictive control[J]. Automation of Electric Power Systems, 2016, 40(18): 7-14.)

[5] Li Z W, Zang C Z, Zeng P, et al. Combined two-stage stochastic programming and receding horizon control strategy for microgrid energy management considering uncertainty[J]. Energies, 2016, 9(7): 499.

[6] 刘金华, 柯钟鸣, 周文辉. 基于强化学习的微电网能源调度策略及优化[J]. 北京邮电大学学报, 2020, 43(1): 28-34.
(Liu J H, Ke Z M, Zhou W H. Reinforcement learning based energy dispatch strategy and control optimization of microgrid[J]. Journal of Beijing University of Posts and Telecommunications, 2020, 43(1): 28-34.)

[7] Kim S, Lim H. Reinforcement learning based energy management algorithm for smart energy buildings[J]. Energies, 2018, 11(8): 2010.

[8] Foruzan E, Soh L K, Asgarpoor S. Reinforcement learning approach for optimal distributed energy management in a microgrid[J]. IEEE Transactions on Power Systems, 2018, 33(5): 5749-5758.

[9] Mbuwir B, Ruelens F, Spiessens F, et al. Battery energy management in a microgrid using batch reinforcement learning[J]. Energies, 2017, 10(11): 1846.

[10] Ji Y, Wang J H, Xu J C, et al. Real-time energy management of a microgrid using deep reinforcement learning[J]. Energies, 2019, 12(12): 2291.

[11] 冯昌森, 张瑜, 文福拴, 等. 基于深度期望Q网络算法的微电网能量管理策略[J]. 电力系统自动化, 2022, 46(3): 14-22.
(Feng C S, Zhang Y, Wen F S, et al. Energy management strategy for microgrid based on deep expected Q network algorithm[J]. Automation of Electric Power Systems, 2022, 46(3): 14-22.)

[12] Francois-Lavet V, Taralla D, Ernst D, et al. Deep reinforcement learning solutions for energy microgrids management[C]. The 13th European Workshop on Reinforcement Learning. Barcelona, 2016: 39019554.

[13] Cao J, Harrold D, Fan Z, et al. Deep reinforcement learning-based energy storage arbitrage with accurate lithium-ion battery degradation model[J]. IEEE Transactions on Smart Grid, 2020, 11(5): 4513-4521.

[14] Papathanassiou S, Hatziargyriou N, Strunz K. A benchmark low voltage microgrid network[C]. Proceedings of the CIGRE Symposium: Power Systems with Dispersed Generation. Athens, 2005: 8.

[15] California ISO open access same-time information system(OASIS)[DB/OL]. (2020-05-20)[2021-02-26]. http://oasis.caiso.com/mrioasis/logon.do.

作者简介

季颖(1987−), 女, 博士生, 从事智能电网、微电网优化调度等研究, E-mail: jiying@stumail.neu.edu,cn;

王建辉(1957−), 女, 教授, 博士生导师, 从事智能控制理论及其应用等研究, E-mail: wangjianhui@ise.neu.edu.cn.

(责任编辑: 郑晓蕾)