doi:10.3772/j.issn.1002-0470.2023.01.008

考虑源荷不确定性下微电网能量调度的深度强化学习策略①

马冲冲②* 王一铮** 王 坤** 冯昌森③*

(*浙江工业大学信息工程学院 杭州 310023)

(**国网浙江省电力有限公司经济技术研究院 杭州 310008)

摘 要 针对微电网中源荷不确定性问题,本文提出一种基于连续型深度确定性策略梯度(DDPG)算法的微电网能量调度方法。首先,以日运行成本最低为目标构建优化调度模型,并将该调度模型转化成马尔可夫决策过程(MDP),定义了马尔可夫决策模型的状态空间、动作空间和奖励函数。其次,利用长短期记忆(LSTM)神经网络提取环境中时序数据的未来趋势作为状态,从而在连续调度动作空间下改善深度强化学习算法收敛效果。最后,通过训练深度强化学习模型,对比多种算法下最优能量调度策略,验证了本文所提方法的有效性。

关键词 微电网;能量管理;强化学习;深度确定性策略梯度(DDPG)

0 引 言

近年来,为实现“碳达峰”和“碳中和”的战略目标,我国提出构建以新能源为主体的新型电力系统[1-2]。微电网作为电网中重要组成部分,可极大提高分布式可再生能源渗透率。然而,由于可再生能源以及负荷需求的随机性使得微电网能量调度问题变得愈加复杂[3]。

目前,针对微电网的能量调度方法已有大量研究。经典线性优化算法[4]、启发式算法[5]、鲁棒优化算法[6-7]和随机规划方法[8]等都可用于求解微电网能量调度问题。例如,文献[9]利用混合粒子群算法对并网和孤岛运行方式的微电网进行能量优化。文献[10]考虑了可再生能源的随机性,提出了日前-实时两阶段随机规划方法。文献[11]建立微电网两阶段鲁棒优化模型,在最坏场景下求解最优调度计划。然而当实际场景存在高不确定性时,上述优化算法难以适用,会带来算法收敛慢、计算精度低、规划结果过于保守等问题。

随着深度强化学习技术的快速发展,因其适用于解决序贯决策问题而成为研究者关注的热点[12]。微电网能量调度问题作为一种时序控制问题,较为契合强化学习框架,因而深度强化学习在电力能量管理领域得到广泛应用[13-14]。文献[15]设计不同超参数的深度 Q 学习算法(deep Q network, DQN)对微能源网能量调度策略进行优化。文献[16]结合深度 Q 学习算法和混合整数线性规划方法,实现对园区多能源系统实时动态经济调度。文献[17]采用基于双深度 Q 学习算法(double deep Q network, DDQN),对微电网能量管理问题进行优化求解,避免了文献[15,16]中 Q 值过估计问题。以上研究通常将连续型决策变量离散化,从而会带来调度结果不精确、经济性差等问题。

为克服上述不足,基于策略梯度的连续型强化学习算法被提出。文献[18]采用双延迟深度确定性策略梯度算法(twin delayed deep deterministic policy gradient,TD3)实现对储能充放电策略的优化。文献[19]引入基于深度确定性策略梯度算法(deep

① 浙江省自然科学基金(LQ22E070007)资助项目。

② 男,1997 年生,硕士生;研究方向:微电网能量管理;E-mail: macc1325@163.com。

③ 通信作者,E-mail: fcs@zjut.edu.cn。

(收稿日期:2022-09-19)

deterministic policy gradient，DDPG），在自适应系统不确定性的基础上，实现任意场景下电力系统动态经济调度。基于上述分析，DDPG 算法较 DQN 算法和 DDQN 算法对环境有更强的探索能力，学习得到的能量调度策略更优。但上述研究并未对模型环境中不确定性因素进行建模，会导致模型收敛效果差，所得策略相对保守。

综上所述，本文提出了一种基于长短期记忆（long short-term memory，LSTM）神经网络和 DDPG 算法的微电网能量调度方法。首先，针对微电网优化运行问题建立相应的马尔可夫决策模型，该模型以调度周期内微电网运行的经济性为目标来寻求最优能量调度策略。其次，针对模型环境中光伏和负荷的随机特性，利用 LSTM 神经网络对其状态转移的不确定性进行建模。基于 LSTM 神经网络对历史光伏和负荷时序数据特征学习，进而得到有效的预测模型。然后，基于 LSTM-DDPG 方法构建微电网能量优化求解框架，并经过模型训练得到最优能量调度策略网络。最后，通过算例仿真验证了本文所提方法的有效性。

1　微电网优化调度模型

1.1　微电网基本结构

本文假设微电网以并网模式运行，其基本结构如图 1 所示。下面介绍并网型微电网的能量优化调度模型。

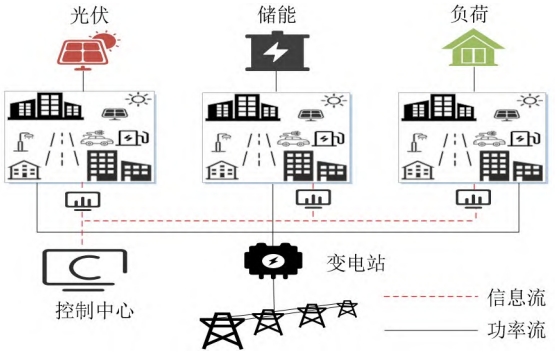

图 1　并网型微电网结构图

1.2　目标函数

在满足负荷需求以及微电网安全运行的前提下，通过对微电网能量优化管理，实现微电网运行成本最小化的目标。

目标函数包含 3 部分，分别是微电网的购电成本 C_{g}、储能设备的折旧成本[20] C_{pv} 和光伏发电设备的运维成本 C_{b}，可表示为

$$\min \quad C = C_{\mathrm{g}} + C_{\mathrm{pv}} + C_{\mathrm{b}} \tag{1}$$

（1）微电网购电成本为

$$C_{\mathrm{g}} = \sum_{t=1}^{T} \left(\lambda_{\mathrm{b},t} P_{\mathrm{b},t}^{\mathrm{g}} - \lambda_{\mathrm{s},t} P_{\mathrm{s},t}^{\mathrm{g}} \right) \tag{2}$$

式中，$\lambda_{\mathrm{b},t}$ 和 $\lambda_{\mathrm{s},t}$ 分别为 t 时段微电网向主电网购电和售电的价格，$P_{\mathrm{b},t}^{\mathrm{g}}$ 和 $P_{\mathrm{s},t}^{\mathrm{g}}$ 分别为 t 时段微电网向主电网的购、售电量。

（2）光伏设备运维成本为

$$C_{\mathrm{pv}} = \sum_{t=1}^{T} k_{\mathrm{pv}} P_t^{\mathrm{pv}} \tag{3}$$

式中，k_{pv} 为光伏的单位运维成本，P_t^{pv} 为 t 时段光伏出力。

（3）储能设备折旧成本

$$C_{\mathrm{b}} = \sum_{t=1}^{T} k_{\mathrm{b}} \left| P_t^{\mathrm{b}} \right| \tag{4}$$

式中，k_{b} 为储能设备的单位折旧成本；P_t^{b} 为 t 时段储能设备的工作功率，$P_t^{\mathrm{b}} < 0$ 表示储能设备充电，反之，则表示储能设备放电。

1.3　模型约束

（1）储能设备约束

储能设备的运行状态包括充电、不工作、放电 3 种状态。考虑到储能设备深度充电和放电会对自身造成损害，故需要在每个时段将储能设备功率和荷电状态约束在一定范围内。因此，储能设备所需满足的约束如下。

$$0 \leqslant \left| P_t^{\mathrm{b}} \right| \leqslant P_{\max}^{\mathrm{b}} \tag{5}$$

$$E_t^{\mathrm{b}} = \begin{cases} E_{t-1}^{\mathrm{b}} - P_t^{\mathrm{b}} u_{\mathrm{ch},t}^{\mathrm{b}} \eta_{\mathrm{ch}} \Delta t & u_{\mathrm{ch},t}^{\mathrm{b}} = 1 \\ E_{t-1}^{\mathrm{b}} & u_{\mathrm{ch},t}^{\mathrm{b}} = u_{\mathrm{dis},t}^{\mathrm{b}} \\ E_{t-1}^{\mathrm{b}} + P_t^{\mathrm{b}} u_{\mathrm{dis},t}^{\mathrm{b}} \Delta t / \eta_{\mathrm{dis}} & u_{\mathrm{dis},t}^{\mathrm{b}} = 1 \end{cases} \tag{6}$$

$$E_{\min}^{\mathrm{b}} \leqslant E_t^{\mathrm{b}} \leqslant E_{\max}^{\mathrm{b}} \tag{7}$$

式中，$P_{\max}^{\mathrm{b}}$ 为储能设备充放电功率的上限值；η_{ch} 和 η_{dis} 分别为储能设备的充电效率和放电效率；$u_{\mathrm{ch},t}^{\mathrm{b}}$ 和 $u_{\mathrm{dis},t}^{\mathrm{b}}$ 分别为 t 时段储能设备充电和放电的状态变量，值为 1 时表示储能设备处于工作状态，值为 0 时，则不工作；Δt 为时间间隔；$E_{\min}^{\mathrm{b}}$ 和 $E_{\max}^{\mathrm{b}}$ 分别为调

度周期内储能设备荷电状态的最小值和最大值;E_t^{b}为t时段储能设备的荷电状态。

由于储能设备的工作状态是单一进行的,即在一个工作时段内,储能设备的充放电状态无法同时存在,故采用式(8)对其约束。

$$u_{\mathrm{ch},t}^{\mathrm{b}} \cdot u_{\mathrm{dis},t}^{\mathrm{b}} = 0 \tag{8}$$

此外,为满足下一个调度周期起始时段对储能设备的蓄能和放能要求,需保证储能设备在调度周期初和调度周期末的荷电状态相等,故采用式(9)对其约束。

$$E_{\mathrm{end}}^{\mathrm{b}} = E_{\mathrm{pre}}^{\mathrm{b}} \tag{9}$$

式中,$E_{\mathrm{end}}^{\mathrm{b}}$为调度周期末储能设备的荷电状态,$E_{\mathrm{pre}}^{\mathrm{b}}$为储能设备在下一个调度周期的初始荷电状态。

(2) 功率平衡约束

$$P_{\mathrm{b},t}^{\mathrm{g}} - P_{\mathrm{s},t}^{\mathrm{g}} + P_t^{\mathrm{pv}} = P_t^{\mathrm{l}} - P_t^{\mathrm{b}} \tag{10}$$

式中,P_t^{l}为t时段的负荷需求。由于同一时段内不能同时存在购电和售电行为,故采用式(11)对其约束。

$$P_{\mathrm{b},t}^{\mathrm{g}} \cdot P_{\mathrm{s},t}^{\mathrm{g}} = 0 \tag{11}$$

(3) 与主电网功率交互约束

为保证在调度周期内变压器的安全运行,还需满足如下约束:

$$0 \leqslant P_{\mathrm{b},t}^{\mathrm{g}} \leqslant P_{\mathrm{b,max}}^{\mathrm{g}} \tag{12}$$

$$0 \leqslant P_{\mathrm{s},t}^{\mathrm{g}} \leqslant P_{\mathrm{s,max}}^{\mathrm{g}} \tag{13}$$

式中,$P_{\mathrm{b,max}}^{\mathrm{g}}$和$P_{\mathrm{s,max}}^{\mathrm{g}}$分别为微电网向主电网购电和售电的最大功率。

2 基于深度强化学习的能量管理方法

针对上述能量管理问题,还需建模成马尔可夫决策模型,从而利用强化学习算法进行求解。马尔可夫决策过程(Markov decision process, MDP)一般用一个5元组($\boldsymbol{S}$, $\boldsymbol{A}$, $\boldsymbol{P}^r$, R, γ)进行表示,其中,$\boldsymbol{S}$表示状态空间,$\boldsymbol{A}$表示动作空间,$\boldsymbol{P}^r$表示状态转移概率,R表示奖励函数,γ表示折扣因子。下面将进行具体建模。

2.1 基于马尔可夫决策过程的能量管理模型

(1) 状态空间

根据第1节对微电网调度模型的描述,本文将马尔可夫决策过程的状态空间用负荷、光伏出力、储能的荷电状态、购售电价以及微电网运行环境的调度时段表示。

$$\boldsymbol{S} = \{P_t^{\mathrm{l}},\ E_t^{\mathrm{b}},\ P_t^{\mathrm{pv}},\lambda_{\mathrm{b},t},\lambda_{\mathrm{s},t},\ t\} \tag{14}$$

(2) 动作空间

本文的连续型动作空间定义为储能设备的充放电功率,分别为储能设备的充电功率$P_{\mathrm{ch},t}^{\mathrm{b}}$、储能设备的放电功率$P_{\mathrm{dis},t}^{\mathrm{b}}$以及不工作。此外,还需根据约束式(5)对调度动作大小进行限制。

(3) 奖励函数

强化学习的目标是为了在与环境的交互探索中获得累计奖励最大。本文对奖励函数设置2个部分,第1部分是由微电网的运行成本函数转化而来,第2部分是由储能运行时的惩罚函数组成。

1) 微电网成本奖励函数

由于光伏出力和负荷的随机性使强化学习算法在每个时段获得的奖励呈现较大波动性,不合理的奖励设置机制会出现强化学习模型训练时间长、收敛性不佳等问题,从而难以学到有效的调度动作。基于此,本文为进一步提升模型训练效果,设置k_1和$k_2$2个正数,对奖励值大小进行缩放处理。因此,第1部分奖励函数设置如式(15)所示。

$$\tilde{C} = \frac{-C}{k_1} + k_2 \tag{15}$$

2) 调度周期内储能运行惩罚函数

储能设备在调度周期内动作时,可能会使储能设备的荷电状态在某个时段内出现越限情况,即违反约束式(7)。因此,需要给错误的调度动作给予惩罚。定义储能运行的惩罚函数,即:

$$D_t^{\mathrm{b}} = \begin{cases} \alpha_{\mathrm{dis}}(E_{\mathrm{min}}^{\mathrm{b}} - E_t^{\mathrm{b}}) & P_t^{\mathrm{b}} > 0,\ E_t^{\mathrm{b}} < E_{\mathrm{min}}^{\mathrm{b}} \\ \alpha_{\mathrm{ch}}(E_t^{\mathrm{b}} - E_{\mathrm{max}}^{\mathrm{b}}) & P_t^{\mathrm{b}} < 0,\ E_t^{\mathrm{b}} > E_{\mathrm{max}}^{\mathrm{b}} \end{cases} \tag{16}$$

式中,α_{dis}和α_{ch}分别表示在调度时段内储能违反最小和最大荷电状态约束的放电和充电惩罚系数。

3) 调度周期末储能惩罚函数

根据约束式(9)可知,还应设置储能调度周期末时储能的荷电状态惩罚函数,即:

$$D_{\mathrm{end},t} = \begin{cases} \lambda_{\mathrm{end}} \mid E_{\mathrm{end}}^{\mathrm{b}} - E_{\mathrm{pre}}^{\mathrm{b}} \mid & t = T_{\mathrm{end}} \\ 0 & t < T_{\mathrm{end}} \end{cases} \tag{17}$$

式中，λ_{end} 为储能调度周期末时段的惩罚系数，$E^{\text{b}}_{\text{end}}$ 为储能调度周期末时的荷电状态，$E^{\text{b}}_{\text{pre}}$ 为下一个调度周期储能的初始荷电状态，T_{end} 表示调度周期末。

因此，第 2 部分储能设备运行的惩罚函数表示为

$$D = \sum_{t=1}^{T} (D^{\text{b}}_t + D_{\text{end},t}) \tag{18}$$

综上所述，本文强化学习的奖励函数可表示为式(19)。

$$R = \tilde{C} + D \tag{19}$$

2.2　基于 LSTM 方法的源荷时序数据特征提取

考虑到光伏出力和负荷的随机性对能量调度策略有较大影响，本文引入 LSTM 对能量调度模型的随机环境进行建模。

LSTM 是一种改进的循环神经网络(recurrent neural network，RNN)，解决了 RNN 存在梯度消失、梯度爆炸以及时序数据长期依赖等问题[21]。由于具有长时间记忆方面特点以至于在时序预测中得到广泛应用[22-23]。针对本文的能量调度模型，在长时间尺度下，光伏出力和负荷具有较强的关联特性。因此，本文利用 LSTM 神经网络对源荷时序数据特征提取，并将提取到的未来时刻源荷特征与状态空间式(14)一起构成 DDPG 算法的策略网络输入，结构如图 2 所示。图 2 中 $P^{\text{pv\_n}}_t$ 和 $P^{\text{l\_n}}_t$ 分别表示未来调度时段的光伏出力和负荷的功率。

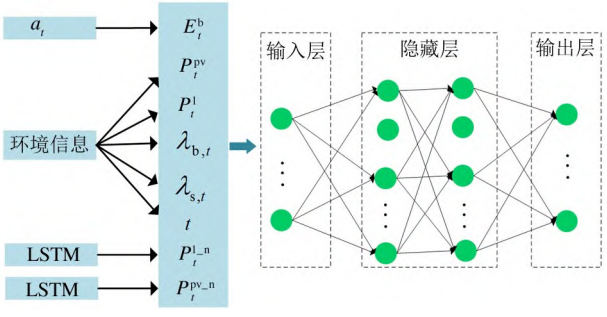

图 2　DDPG 策略网络输入结构图

2.3　微电网能量调度策略的 DDPG 算法求解

将具备感知和决策能力的深度强化学习算法[24]用于求解本文的微电网能量调度问题，实现微电网经济运行的目标。本文采用连续型强化学习的 DDPG 算法对第 1 节中微电网优化调度模型进行求解。

DDPG 算法是基于演员-评论家(actor-critic，AC)算法框架，并将演员网络作为策略网络，评论家网络作为价值网络。同时，DDPG 算法采用 DQN 算法的经验回放机制以及独立的目标网络结构。在 DDPG 算法探索学习过程中，对当前策略网络输入微电网运行环境状态 s_t，当前策略网络会基于确定性策略 μ 输出动作 a_t，即：

$$a_t = \mu(s_t \mid \theta^{\mu}) \tag{20}$$

式中，θ^{μ} 为当前策略网络参数。

为使 DDPG 算法在微电网的运行环境中有更强的探索能力，本文将策略网络输出的调度动作 a_t 作为均值，$\tilde{v}$ 为标准差构成一个正态分布 $N(a_t,\tilde{v})$。可通过衰减系数 ε 来控制 $\tilde{v}$ 衰减的速度，ε 越大，$\tilde{v}$ 衰减速度越慢；ε 越小，$\tilde{v}$ 衰减速度越快，以此控制 DDPG 算法在环境中的探索范围。然后从正态分布中随机输出新的动作作为实际调度动作，可表示为

$$a_t \sim N(a_t, \tilde{v}) \tag{21}$$

在 DDPG 算法训练前，由每个调度周期内微电网运行环境状态 s_t、能量调度动作 a_t、环境反馈的奖励 r_t 以及下一个状态 s_{t+1} 组成的状态转移序列样本 (s_t, a_t, r_t, s_{t+1}) 储存到经验池中作为训练样本。具体训练过程如下。

首先，从经验池中随机采样 N 个样本通过价值网络计算目标价值 y_i 和损失函数 L，如式(22)和式(23)所示。

$$L = \frac{1}{N}\sum_i [\, y_i - Q(s_i, a_i \mid \theta^{Q}) \,]^2 \tag{22}$$

$$y_i = r_i + \gamma Q'(s_{i+1}, \mu'(s_{i+1} \mid \theta^{\mu'}) \mid \theta^{Q'}) \tag{23}$$

式中，N 为采样的训练样本数；$\theta^{\mu'}$ 为目标策略网络的参数；θ^{Q} 为当前价值网络参数，$\theta^{Q'}$ 为目标价值网络的参数；i 表示采样的训练样本编号。

然后，再通过梯度下降策略和最小化损失函数对策略网络参数优化更新，如式(24)所示。

$$\nabla_{\theta^{\mu}} J \approx \frac{1}{N}\sum_i [\, \nabla_a Q(s,a \mid \theta^{Q}) \mid_{s=s_i, a=\mu(s_i)} \nabla_{\theta^{\mu}} \mu(s \mid \theta^{u}) \mid_{s_i}] \tag{24}$$

最后，在对策略网络和价值网络训练后，采用软更新的方法，对目标策略网络和目标价值网络参数

进行更新,即:

$$\theta^{\mu'} \leftarrow \tau\theta^{\mu} + (1-\tau)\theta^{\mu'} \tag{25}$$

$$\theta^{Q'} \leftarrow \tau\theta^{Q} + (1-\tau)\theta^{Q'} \tag{26}$$

式中,τ 为当前网络参数对目标网络参数进行软更新的系数,一般取值范围为 $0 < \tau \ll 1$。

基于 LSTM-DDPG 方法所建立的微电网能量优化调度框架如图 3 所示。由图 3 可知,利用 LSTM 预测未来调度时段的光伏和负荷,并构成图 3 的环境状态。通过 DDPG 算法输出调度动作作用于微电网环境并得到反馈奖励和下一时刻环境运行状态,实现了 DDPG 算法与微电网环境的交互。

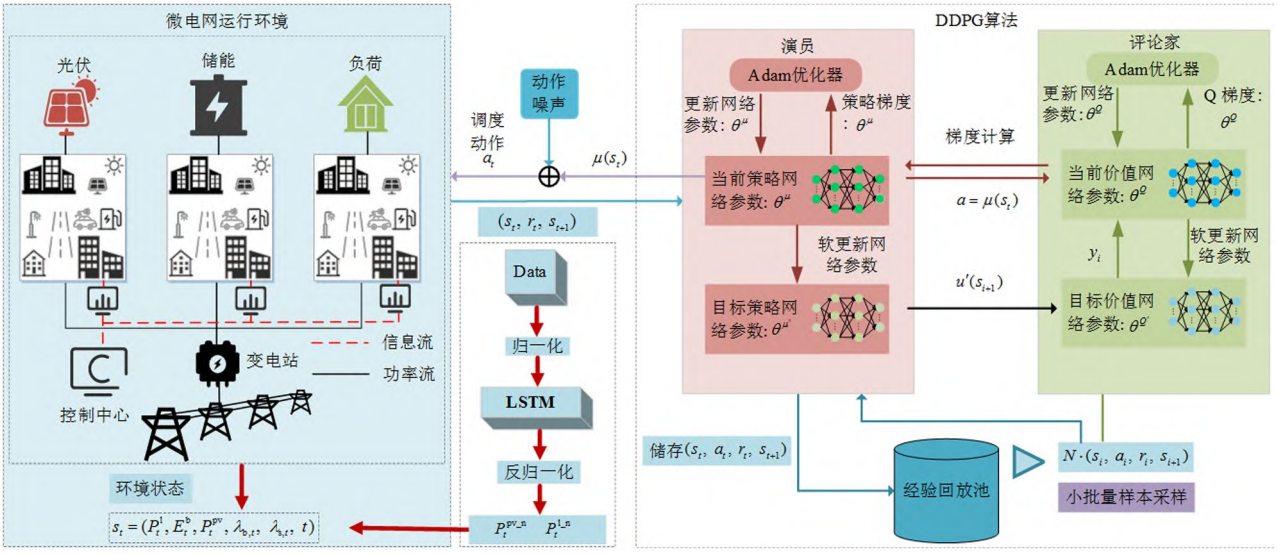

图 3　基于 LSTM-DDPG 方法的微电网能量优化调度框架

3　算例分析

3.1　算例参数介绍

本文在 Python 3.6 和 TensorFlow 框架下编写程序,进行仿真实验。计算机硬件配置为 i5-7300HQ CPU,2.50 GHz。算例中光伏数据集来源于 2019 年澳大利亚的 Yulara 光伏发电系统。储能设备容量设置为 1000 kW·h,充放电功率最大限定为 200 kW,储能设备的荷电状态最大、最小值分别设置为 0.8 和 0.2,初始荷电状态为 0.3。本文采用分时电价,分为峰时段、平时段以及谷时段,峰时段为 6:00—11:00、19:00—23:00,平时段为 11:00—19:00,谷时段为 23:00—6:00,电价参数如表 1 所示。

本文基于文献[19]关于超参数的设计思路,对本文 DDPG 算法的超参数进行设计,如表 2 所示。此外,实验过程中还对 DDPG 算法的神经网络隐藏层层数均设置为 2 层,每层神经元的个数设置为 100,隐藏层的激活函数设置为 ReLU 函数。

表 1　分时电价参数

电价 / 时段	价格/元/(kW·h)	
	购电电价	售电电价
峰时段	1.21	1.02
平时段	0.69	0.5
谷时段	0.43	0.27

表 2　DDPG 算法超参数

超参数	数值
演员网络学习率	0.001
评论家网络学习率	0.0001
折扣因子	0.95
软更新系数	0.001
经验池容量	20 000
小批量样本	128
网络优化器	Adam 优化器

针对光伏出力和负荷预测,本文基于一年的历史数据集并按照 6:2:2 比例划分为训练集、验证集和测试集。同时,为避免数据波动大以及数据异常

使神经网络收敛慢、预测精度较低等问题，需在 LSTM 模型训练前对神经网络输入数据进行预处理。本文利用 min-max 方法对神经网络的输入数据进行归一化处理，LSTM 神经网络的超参数如表 3 所示。

表 3　LSTM 神经网络超参数

超参数	数值
小批量样本	64
学习率	0.005
隐藏层神经元个数	128
隐藏层层数	3
网络优化器	Adam 优化器

3.2　LSTM 神经网络预测分析

本文在训练 LSTM 模型过程中，光伏训练集和验证集的网络损失均方误差（mean square error，MSE）分别为 0.001 35 和 0.001 27；负荷训练集和验证集的 MSE 分别为 0.003 25 和 0.003 54。LSTM 模型经过 110 次训练后，用该模型对光伏和负荷的测试集数据进行预测，光伏和负荷的平均绝对百分比误差（mean absolute percentage error，MAPE）分别为 34.0% 和 3.5%，预测效果可行。

然后，选取 2019 年 10 月 21 日至 2019 年 11 月 3 日光伏和负荷的预测结果绘制对比图，如图 4 和图 5 所示。

3.3　LSTM-DDPG 方法调度结果分析

在训练强化学习模型后，以 2019 年 12 月 26 日的能量调度情况为例，对所得的策略模型进行测试分析，结果如图 6 所示。

从图 6 可知，在谷电价和负荷需求较低时段，如 00:00—05:00时段，由于光伏不出力，需从主电网

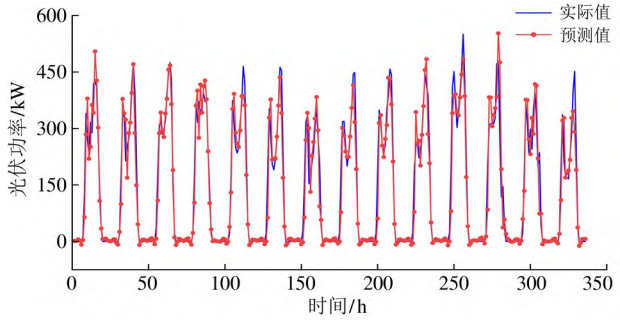

图 4　基于 LSTM 神经网络光伏预测曲线

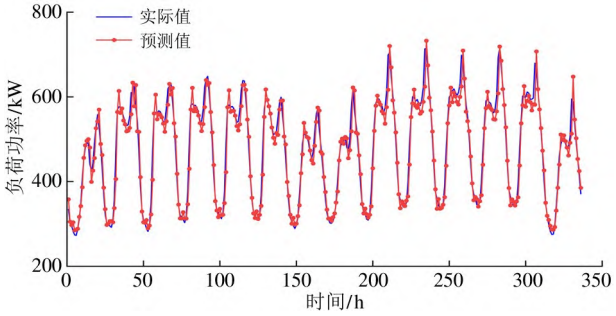

图 5　基于 LSTM 神经网络负荷预测曲线

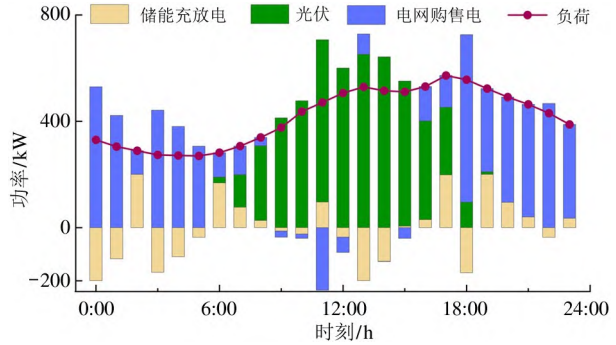

图 6　基于 LSTM-DDPG 算法的微电网日优化调度结果

大量购电以满足该时段的负荷需求。此外，该时段储能主要采取充电策略以满足电价较高或负荷高峰时的用电需求。在峰电价时段，如 6:00—8:00 时段，此时电价较高，会增加从主电网的购电成本；而光伏出力增大以及储能采取放电策略在很大程度上降低了微电网的购电成本。在平电价时段，如 11:00—15:00 时段，光伏出力较大，微电网向主电网的购电量大幅减少。同时，该时段储能处在充电状态，可在后续调度时段通过放电满足负荷需求。在 19:00—20:00 时段内，随着光伏出力减弱，储能采取放电策略，有效提升了微电网运行的经济性。由上述调度结果分析可知，储能在电价引导下，可在相应的时段内采取合理的充放电策略。

为分析光伏出力和负荷预测误差对能量调度结果的影响，设计如下 3 种算例模式进行对比分析。

模式 1：本文模型。

模式 2：假设在光伏和负荷完美预测情况下，利用对应的实际数据训练强化学习模型。

模式 3：不使用 LSTM 神经网络建模，直接采用 DDPG 算法求解。

由表 4 可知，模式 1 较模式 3 的运行成本降低

2.74%,说明本文所提 LSTM-DDPG 方法能够有效降低微电网运行成本。以模式 3 为基准,模式 2 下微电网运行成本较模式 3 降低 3.91%,相比模式 1 进一步降低了微电网运行成本。由此可知,可通过提高 LSTM 神经网络预测精度继而增强深度强化学习算法的寻优能力。

表 4　3 种模式下微电网运行成本

模式	成本/元
模式 1	4315.48
模式 2	4263.39
模式 3	4436.81

同时,为在不同划分比例的光伏和负荷数据集训练 LSTM 模型下分析本文方法对微电网的优化效果,设置如下 2 种算例模式进行对比分析。

模式 4:设数据集划分比例为 4∶3∶3,分别训练光伏和负荷的 LSTM 预测模型。

模式 5:设数据集划分比例为 8∶1∶1,分别训练光伏和负荷的 LSTM 预测模型。

由表 5 可知,模式 1 较模式 4 和模式 5 的微电网运行成本更低。因此,在本文采取的数据集划分比例下,对微电网运行成本优化更好。

表 5　3 种模式下微电网运行成本

模式	微电网运行成本
模式 1	4315.48
模式 4	4650.63
模式 5	4445.81

图 7 是 3 种模式下的储能荷电状态。由图 7 可知,在 16:00—19:00 时段内,可看出模式 1 和模式 2 储能的荷电状态呈现较大的上升趋势,而模式 3 表现较小的波动。并且,该时段内模式 3 的荷电状态水平整体偏低,难以在未来时段通过放电降低微电网的运行成本。通过上述对比分析,模式 1 和模式 2 中储能在调度周期内采取的能量调度策略较模式 3 要更优,使微电网运行成本降低更多。

图 8 为在模式 1 下对 DDPG 算法策略网络训练的奖励曲线。由图8可知,在5000回合前,奖励振

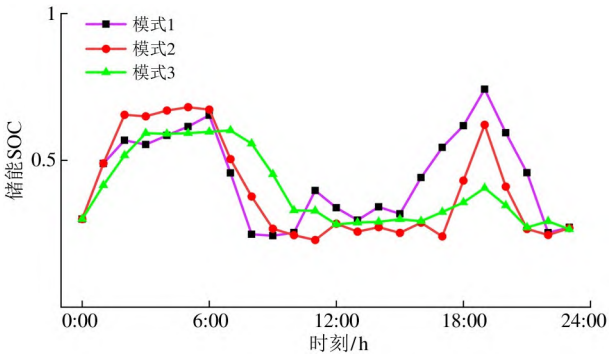

图 7　3 种模式下的储能荷电状态

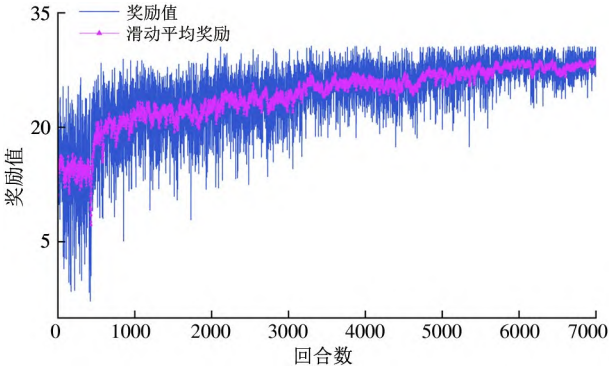

图 8　基于 LSTM-DDPG 算法的奖励曲线

荡较大且有明显的上升趋势。说明此阶段策略网络还不稳定,输出的调度动作可能会使储能的荷电状态发生越限而受到惩罚。在 6000 回合后,随着策略网络参数趋于稳定,滑动平均奖励曲线逐渐收敛,即奖励函数逐渐收敛到稳定值。需要指出最终的奖励曲线仍然处于较小的波动,这是因为在每个调度时段内微电网中光伏出力和负荷功率的不确定性而引发奖励小范围振荡,属于正常情况。图 9 为 DDPG 算法在训练阶段损失函数的变化过程,可以看出损失函数表现出较为明显的减小趋势。

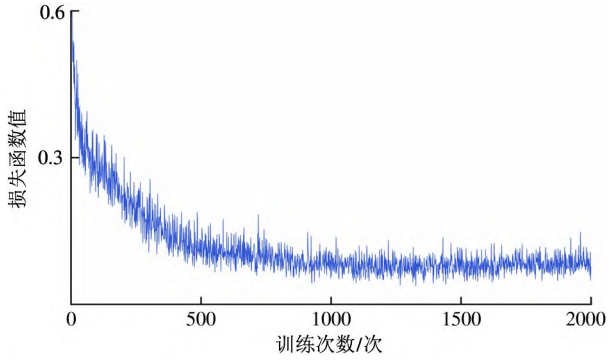

图 9　基于 LSTM-DDPG 算法的损失函数曲线

3.4　模型经济性分析

表 6 为对比了在多种深度强化学习算法下求解的微电网日运行成本，其中，离散度 d 表示对连续动作离散化的大小。

表 6　不同深度强化学习算法的优化结果

方法	离散度	微电网运行成本
LSTM-DDPG	—	4361.41
LSTM-DDQN	$d=5$	4753.36
	$d=50$	4860.16
DDQN	$d=5$	4846.48
LSTM-Dueling DQN	$d=5$	4871.63
	$d=50$	4930.50
Dueling DQN	$d=5$	4901.34

由表 6 可知，当 LSTM-DDQN 算法的动作离散度为 5 和 50 时，储能的调度动作离散度越大，动作空间越小，求解的微电网运行成本越高。这主要是因为动作离散度大使得 DDQN 算法中动作空间和状态空间中可描述的信息少，无法对环境充分探索，从而难以学到最优能量调度策略。因此，本文采取连续型深度强化学习算法有效提升微电网运行的经济性。

3.5　算法泛化性及稳定性分析

为分析本文所提深度网络的泛化性，选取其他区域的光伏和负荷数据，并设置 4 组对比实验，如表 7所示。表 7 中还对比了文献[18]中采用的 TD3 算法，且 4 组对比实验中本文算法均取得更低的运行成本。由此可知，本文所提算法在其他地区的数据集上也能表现较好的优化能力。即验证了本文所提 LSTM-DDPG 方法具有良好的泛化性。

表 7　4 组对比实验下的微电网运行成本

测试日	成本/元			
	LSTM-DDPG	DDPG	LSTM-TD3	TD3
测试日 1	4933.69	5050.28	4982.43	5116.17
测试日 2	5328.42	5487.20	5383.58	5498.71
测试日 3	6049.12	6125.33	6098.29	6250.61
测试日 4	5243.34	5360.51	5311.99	5429.88

为进一步分析本文所提方法的稳定性，采用 5 组随机种子进行仿真实验，并训练 7000 回合。图 10 是由 5 次实验得到的平均奖励值、最大奖励值和最小奖励值曲线组成，阴影部分表示奖励的最大最小值区间，中间实线表示 5 次实验的平均奖励值。由图 10 可知，在迭代次数为 6600 后，平均奖励和最大最小值之间差距逐步缩小直至收敛，说明本文所提方法具有良好的稳定性。

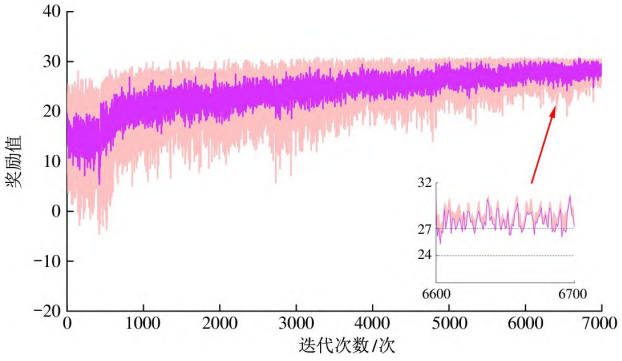

图 10　基于 DDPG 算法训练奖励曲线

4　结　论

本文提出了一种基于连续型深度确定性策略梯度算法的微电网能量调度方法，即利用深度强化学习对微电网能量优化调度问题进行建模和求解。通过建立基于 LSTM-DDPG 方法的能量调度框架，有效避免了连续调度动作离散化对调度策略的影响。仿真实验对比分析表明，所提方法可对微电网实时做出调度决策；通过对比多种深度强化学习算法可知，本文算法可有效应对随机变量的影响，提升微电网运行的经济性，在训练过程表现出较好的泛化性和稳定性。本文所采用的深度强化学习方法也可为其他优化调度问题提供思路，如综合能源系统、电动汽车充电站能量管理等。

参考文献

[1] 李晖，刘栋，姚丹阳. 面向碳达峰碳中和目标的我国电力系统发展研判[J]. 中国电机工程学报，2021，41(18)：6245-6259.
[2] 童光毅. 基于双碳目标的智慧能源体系构建[J]. 智慧电力，2021，49(5)：1-6.
[3] 鲁卓欣，徐潇源，严正，等. 不确定性环境下数据驱动的电力系统优化调度方法综述[J]. 电力系统自动化，2020，44(21)：172-183.
[4] SHI W B，LI N，CHU C C，et al. Real-time energy management in microgrids[J]. IEEE Transactions on Smart

Grid, 2017, 8(1): 228-238.
[5] 马天祥, 贾伯岩, 张智远, 等. 基于二层规划的能源互联微电网能量优化调度方法[J]. 电力系统自动化, 2019, 43(16): 34-43.
[6] WANG Z Y, CHEN B K, WANG J H, et al. Robust optimization based optimal DG placement in microgrids[J]. IEEE Transactions on Smart Grid, 2014, 5(5): 2173-2182.
[7] 贺帅佳, 阮贺彬, 高红均, 等. 分布鲁棒优化方法在电力系统中的理论分析与应用综述[J]. 电力系统自动化, 2020, 44(14): 179-191.
[8] ALIPOUR M, ZARE K, SEYEDI H. A multi-follower bilevel stochastic programming approach for energy management of combined heat and power microgrids[J]. Energy, 2018, 149:135-146.
[9] 吴定会, 高聪, 纪志成. 混合粒子群算法在微电网经济优化运行的应用[J]. 控制理论与应用, 2018, 35(4): 457-467.
[10] 黄张浩, 张亚超, 郑峰, 等. 基于不同利益主体协调优化的主动配电网日前-实时能量管理方法[J]. 电网技术, 2021, 45(6): 2299-2308.
[11] 刘一欣, 郭力, 王成山. 微电网两阶段鲁棒优化经济调度方法[J]. 中国电机工程学报, 2018, 38(14): 4013-4022.
[12] ZHANG Z D, ZHANG D X, QIU R C. Deep reinforcement learning for power system applications: an overview [J]. CSE-E Journal of Power and Energy Systems, 2020, 6(1): 213-225.
[13] MOCANU E, MOCANU D C, NGUYEN P H, et al. Online building energy optimization using deep reinforcement learning[J]. IEEE Transactions on Smart Grid, 2019, 10(4): 3698-3708.
[14] FORUZAN E, SOH L, ASGARPOOR S. Reinforcement learning approach for optimal distributed energy management in a microgrid[J]. IEEE Transactions on Power Systems, 2018, 33(5): 5749-5758.
[15] 刘俊峰, 陈剑龙, 王晓生, 等. 基于深度强化学习的微能源网能量管理与优化策略研究[J]. 电网技术, 2020, 44(10):3794-3803.
[16] 聂欢欢, 张家琦, 陈颖, 等. 基于双层强化学习方法的多能园区实时经济调度[J]. 电网技术, 2021, 45(4): 1330-1336.
[17] 梁宏, 李鸿鑫, 张华赢, 等. 基于深度强化学习的微网储能系统控制策略研究[J]. 电网技术, 2021, 45(10):3869-3877.
[18] 陈亭轩, 徐潇源, 严正, 等. 基于深度强化学习的光储充电站储能系统优化运行[J]. 电力自动化设备, 2021, 41(10): 90-98.
[19] 彭刘阳, 孙元章, 徐箭, 等. 基于深度强化学习的自适应不确定性经济调度[J]. 电力系统自动化, 2020, 44(9): 33-42.
[20] FENG C S, WEN F S, YOU S, et al. Coalitional game-based transitive energy management in local energy communities[J]. IEEE Transactions on Power Systems, 2020, 35(3): 1729-1740.
[21] 施海昕, 诸建超, 严骏驰, 等. 基于卷积神经网络和LSTM循环神经网络的客户复购预测方法[J]. 高技术通讯, 2021, 31(7): 713-722.
[22] 黄运有, 詹剑锋. 基于深度学习的短期负荷预测综述[J]. 高技术通讯, 2021, 31(3): 240-248.
[23] CHAI M K, XIA F, HAO S T, et al. PV power prediction based on LSTM with adaptive hyperparameter adjustment[J]. IEEE Access, 2019, (7): 115473-115486.
[24] 赵冬斌, 邵坤, 朱圆恒, 等. 深度强化学习综述:兼论计算机围棋的发展[J]. 控制理论与应用, 2016, 3306: 701-717.

Deep reinforcement learning based dispatch strategy for microgrid energy management considering uncertainty of source and load

MA Chongchong*, WANG Yizheng**, WANG Kun**, FENG Changsen*
(*College of Information Engineering, Zhejiang University of Technology, Hangzhou 310023)
(**Economic Research Institute of State Grid Zhejiang Electric Power Co. Ltd., Hangzhou 310008)

Abstract

Inorder to address the renewable energy and load uncertainty problem, a microgrid energy scheduling method based on a continuous deep deterministic policy gradient (DDPG) algorithm is proposed. Firstly, an optimization model is establised with the objective function to minimize the daily operating cost, and thereby the decision model is transformed into a Markov decision process (MDP) where the tuple of MDP, i.e., the state space, action space and reward function, is defined. Secondly, the long short-term memory (LSTM) neural network is used to extract the future trend of the time series data in the stochastic environment as the state, so as to improve the convergence effect of the deep reinforcement learning algorithm in the continuous action space. Finally, through training a deep reinforcement learning model and comparing the energy scheduling strategies obtained from the various algorithms, the effectiveness of the method proposed in this paper is verified.

Key words: microgrid, energy management, reinforcement learning, deep deterministic policy gradient (DDPG)