·自动化与信息化·

文章编号:1005-6548(2023)01-0054-10 中图分类号:TM711 文献标识码:A 学科分类号:47040
DOI:10.13357/j.dlxb.2023.006 开放科学(资源服务)标识码(OSID):

基于深度强化学习的微电网优化调度研究

罗建勋[1],张 玮[1],王 辉[2],邓立军[3]
(1.齐鲁工业大学(山东省科学院)信息与自动化学院,济南 250353;2.山东大学 电气工程学院,济南 250000;
3.国网济南供电公司,济南 250001)

摘 要:针对微电网的调度优化问题,以一个包括风力发电机、储能系统、恒温控制负荷和价格响应负荷的新型微电网模型为研究对象,以实现新型微电网经济运行成本最小为目标,计及风力发电的波动性和随机性对微电网安全经济运行带来的影响,在基于AC算法的框架上,提出了一种改进的A3C算法。通过采用多线程的方法实现异步训练,并将DQN算法的经验回放机制由均匀性采样改进为重要性采样加入到A3C算法的训练中。试验分别对DQN、AC和改进的A3C算法进行了训练仿真并进行了对比。结果表明,改进的A3C算法提高了样本的利用率,训练时间为1.46 min,缩短了训练时间。当风力发电出现波动时,依据主电网的实时电价,使用改进的A3C算法模型通过控制储能装置的充放电,减少了该微电网在高电价时从主电网的购电量,使其在低电价时再大量购电,从而降低了购电成本。该模型给出的调度策略提高了经济效益,有效地降低了风电波动对微电网的影响。所提方案可为微电网智能调度提供参考。

关键词:微电网;调度优化;调度策略;深度强化学习;重要性采样

Research on Optimal Scheduling of Micro-Grid Based on Deep Reinforcement Learning

LUO Jianxun[1], ZHANG Wei[1], WANG Hui[2], DENG Lijun[3]
(1.*School of Information and Automation Engineering, Qilu University of Technology(Shandong Academy of Sciences), Jinan* 250353, *China*; 2.*School of Electrial Engineering, Shandong University, Jinan* 250000, *China*;
3.*State Grid Jinan Power Supply Company, Jinan* 250001, *China*)

Abstract: In the view of the scheduling optimization problem of microgrid, a novel microgrid model that consists of a wind turbine generator, an energy storage system, a set of thermostatically controlled loads and a set of price-responsive loads is used as the research object. With the goal of achieving the minimum economic operating cost, and considering the impact of the volatility and randomness of wind power generation on the safe and economic operation of the microgrid, based on the framework of AC algorithm, an improved A3C algorithm is proposed. Asynchronous training is realized by adopting multi-threading method. The experience replay mecha-

* 收稿日期:2022-05-26
基金项目:国家自然科学基金(面向跨境互联的多能互补新能源系统关键技术研究,2018YFE0208400)。
作者简介:罗建勋(1996—),男,硕士研究生,研究方向为微电网优化调度,1944571225@qq.com;
张 玮(1973—),女,博士,副教授,研究方向为电力系统继电保护、分布式发电,zhangwei_jn@126.com;
王 辉(1974—),男,博士,教授,研究方向为新能源与分布式发电、电力电子变换与控制技术,sddlwh@sdu.edu.cn;
邓立军(1988—),男,硕士研究生,高级工程师,研究方向为电力系统及其自动化,769139380@qq.com。
引文格式:罗建勋,张玮,王辉,等. 基于深度强化学习的微电网优化调度研究[J]. 电力学报,2023,38(01):54-63. DOI:10.13357/j.dlxb.2023.006.

nism of DQN algorithm is improved from uniform sampling to importance sampling, and is added to the training of A3C algorithm. The DQN, AC and improved A3C algorithms are trained and simulated respectively and compared. Simulation shows that the utilization rate of samples is improved and the training time is reduced by improved A3C algorithm, whose training time is 1.46 minutes. When the wind power fluctuates, the model trained by the improved A3C algorithm controls the charging and discharging of the energy storage device according to the real-time electricity price of the main grid. The scheduling strategy given in the model improves the economic benefits and effectively reduces the impact of wind power fluctuations on the microgrid. The proposed scheme can provide reference for intelligent dispatching of microgrdid.

Key words: micro-grdid; optimization; scheduling strategies; deep reinforcement learning; importance sampling

0　引言

为了解决化石能源日益枯竭以及其燃烧所引起的环境问题，清洁能源作为可再生能源已得到了广泛的使用。但可再生能源如风电、光电有很大的波动性和间歇性，给大规模的新能源并网带来了挑战[1]。微电网是由分布式电源、储能装置和负荷等组成的小型发配电系统，能促进分布式电源与可再生能源的大规模接入[2]，已得到广泛应用。基于新能源预测制定的日内调度计划有时会与实际发电量有较大的出入，会造成供需不平衡，同时，分布式电源的接入也可能使母线电压波动，对微电网运行的经济性和安全性产生影响[3]。因此，对微电网优化调度的研究有着十分重要的意义。

深度强化学习（Deep Reinforcement Learning，DRL）借鉴人类学习的过程，通过与环境的不断交互进行试错来寻求最优的策略，给学者在微电网优化调度研究上提供了许多启发[4-5]。近年来，许多学者采用深度强化学习算法来解决微电网优化问题。文献[6]提出了一种基于Q学习的微电网经济调度算法，在应对风能、光能的不确定时，通过调整储能系统的功率，保证微电网总的充放电功率保持稳定。文献[7-8]提出了一种基于Q学习的控制器，使风电和储能组成的混合系统参与电力交易来减轻风电的不确定影响。文献[9]将深度Q-Learning应用到电池的能量使用管理问题中，在历史电价下学习最优的充放电计划。文献[10]提出了一种基于DQN算法的实时能量管理方法，联合调度备用发动机的使用，管理储能系统和电网运行。文献[11]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，对可再生能源的随机波动具有一定的鲁棒性，可以有效地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。

以上研究多采用基于值函数的强化学习算法，常见的有Q-Learning[12]，Sarsa[13]和DQN[14]（Deep Q-Network）算法，基于策略函数方面上的研究还较少。在保证微电网安全运行的前提下，以最低运行成本为目标，本文提出了一种改进的A3C（Asynchronous Advantage Actor Critic）算法，采用多个智能体共同探索实现异步训练，加快模型的收敛，同时减少样本的关联性。为了提高样本的利用率，避免无用的训练，A3C算法采用改进的经验回放机制，由均匀采样改为重要性采样。

1　微电网优化模型

微电网模型的结构图如图1所示，其是由一个风力发电的分布式电源，一个公共的储能系统（Energy Storage System，ESS），一组恒温控制负荷（Thermostatically Controlled Loads，TCLS）如空调，冰箱或热水器等，以及一组住宅负荷组成的新型微电网模型，同时微电网与主电网相连，可以在电力系统中进行能源的买卖。基于深度强化学习算法构建的微电网能量管理系统（Energy Management System，EMS）接收各系统反馈的信息，对TCLS进行直接控制，对电力进行定价，在能量不足时控制与主电网进行电力购买和储能系统放电，能量过剩时控制储能系统充电和其与主电网进行电力出售。

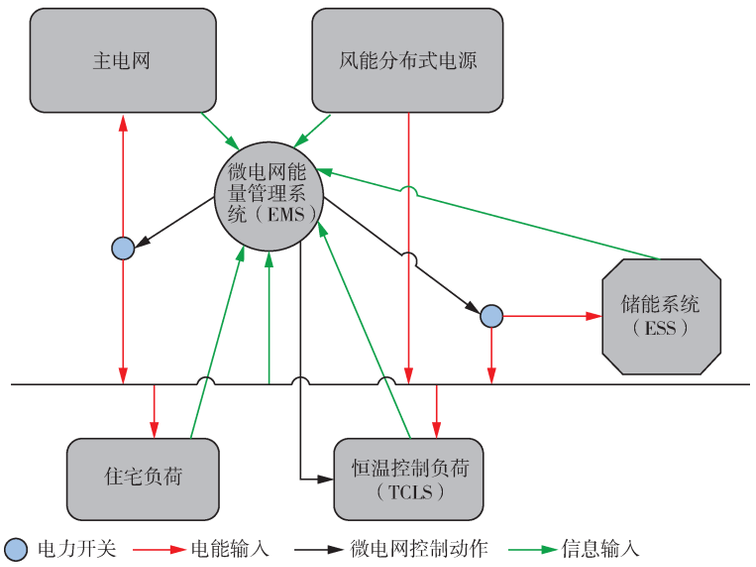

图 1 微电网模型结构图

Fig. 1 Diagram of microgrid model structure

1.1 优化目标

微电网优化调度是根据能量状况对储能装置和其与主电网的交互进行控制，同时对恒温控制负载和住宅负载进行控制。在满足各种约束条件下，对能量进行调度以实现微电网运行经济化。目标函数可表示为：

$$F=\min C\left(\sum_{t=0}^{T}R_T-C_T\right). \tag{1}$$

$$R_T=P_T\sum_{i=1}^{w_{\text{loads}}}L_{\text{load}}^{(i)}+C_{\text{gen}}\sum_{j=1}^{w_{\text{TCLS}}}L_{\text{tcl}}^{(j)}u_{\text{b,t}}^{(j)}+P_T^{(\text{D})}E_T^{(\text{S})}. \tag{2}$$

$$w_{\text{AC},T}=C_{\text{gen}}G_T+\left(P_T^{(\text{U})}+C_{\text{tr imp}}\right)E_T^{(\text{P})}+C_{\text{tr exp}}E_T^{(\text{S})}. \tag{3}$$

式(1)中，R_T表示在T时刻微电网中出售电力给用户和主电网所获得的收入；C_T表示在T时刻微电网发电，及从主电网购电和输电所花的成本。式(2)中，P_T表示微电网卖给用户电力的价格；$L_{\text{load}}^{(i)}$表示第i组住宅负荷所消耗的功率；w_{loads}表示微电网中的住宅负荷总数；C_{gen}是发电成本；$L_{\text{tcl}}^{(j)}$表示第j组TCL所消耗的功率；$u_{\text{b,t}}^{(j)}$表示第j组TCL备份控制器决定的开关动作；w_{TCLS}表示微电网中的恒温控制负荷总数；$P_T^{(\text{D})}$表示下调价格，即向主电网出售电力的价格；$E_T^{(\text{S})}$表示出售给主电网的电量。式(3)中，G_T和$E_T^{(\text{P})}$分别为风力发出的电量和从主电网购买的电量；$P_T^{(\text{U})}$表示上调价格，即向主电网购买电力的价格；$C_{\text{tr imp}}$和$C_{\text{tr exp}}$分别表示从外网输入和输送到外网的输电成本。

1.2 约束条件

微电网调度优化调度需要满足下列约束条件。

1)功率平衡约束条件：

$$G_T+E_T^{(\text{P})}+D_T-P_T^{(\text{loads})}-P_T^{(\text{tcls})}-E_T^{(\text{S})}-C_T=0. \tag{4}$$

式(4)中，G_T是风机输出的有功功率，$E_T^{(\text{P})}$是从主电网购买的有功功率，D_T是储能装置发出的有功功率，$P_T^{(\text{loads})}$是微电网中所有住宅负荷消耗的有功功率，$P_T^{(\text{tcls})}$是微电网中所有TCLS消耗的有功功率，$E_T^{(\text{S})}$是出售给主电网的有功功率，C_T是储能系统充入的有功功率。

2)储能系统运行约束条件：

$$w_{\text{SOC,min}}\leqslant w_{\text{SOC},T}\leqslant w_{\text{SOC,max}}. \tag{5}$$

$$w_{\text{SOC},T+1}=w_{\text{SOC},T}+\eta_c C_T-\frac{D_T}{\eta_{\text{d}}}. \tag{6}$$

$$0\leqslant C_T\leqslant C_{\max}. \tag{7}$$

$$0 \leqslant D_T \leqslant D_{\max}\ . \tag{8}$$

式(5)中，$w_{\mathrm{SOC},T}$表示 T时刻该储能系统的荷电状态，$w_{\mathrm{SOC},\min}$、$w_{\mathrm{SOC},\max}$为其上下限约束。式(6)中，η_c、η_d为储能系统的充放电系数，C_T、D_T是储能系统的充放电功率。式(7)和式(8)中，$C_{\max}$、$D_{\max}$是根据微电网调度控制的充放电速率得出的储能系统最大充放电功率。

3)与主电网交互约束条件：

$$-G_T \leqslant P_{\mathrm{ex}} \leqslant P_{\mathrm{ex},\max}\ . \tag{9}$$

式(9)中，P_{ex}为与主电网交换的功率，$-G_T$表示 T时刻向主电网可以出售的最大功率为风电发出的功率，$P_{\mathrm{ex},\max}$表示可以从主电网购买的最大功率。

4)恒温控制负荷约束条件：

$$u_{\mathrm{b,t}}^{(j)} = \begin{cases} 0, & T_T^{(j)} > T_{\max}^{(j)} \\ u_T^{(j)}, & T_{\min}^{(j)} \leqslant T_T^{(j)} \leqslant T_{\max}^{(j)} \\ 0, & T_T^{(j)} < T_{\min}^{(j)} \end{cases}\ . \tag{10}$$

式(10)中，$u_{\mathrm{b,t}}^{(j)}$为第 j组 TCLS 备份控制器决定的开关动作，$T_T^{(j)}$为第 j组 TCLS 设备在 T时刻的运行温度，$T_{\max}^{(j)}$、$T_{\min}^{(j)}$为用户设置的温度上下限。

2　马尔科夫决策过程

马尔科夫(Markov)定义：下一个状态的产生只和当前的状态有关，即：

$$P[S_{T+1}|S_T] = P[S_{T+1}|S_1, \cdots, S_T]\ . \tag{11}$$

式(11)右边所示说明了下一个时刻状态 S_{T+1}是与所有的历史状态 $S_1, \cdots, S_T$有关，但是 Markov 定义则是忽略了 $S_1, \cdots, S_{T-1}$历史信息，只保留了当前时刻状态 S_T的信息来预测下一个状态。

强化学习中智能体(Agent)在与环境进行交互时，首先初始化环境获得当前时刻 T的状态空间 S_0(State)，智能体基于这一时刻的状态 S_0给出动作 A_0(Action)作用于环境上，环境返回一个奖励值 R_1(Reward)和下一个时刻 $T+1$的状态信息 S_1。这样智能体与微电网环境之间的交互就产生了一个序列：$S_0, A_0, R_1, S_1, A_1, R_2, \cdots$，这个即为序列决策过程，而马尔科夫决策过程就是一个典型的序列决策过程的公式化。有了马尔科夫的假设，在解决这个序列决策过程才比较方便实用。

在真实的环境转化中，转化到下一个状态 S_{T+1}的概率和当前状态 S_T有关，还和上一个状态以及之前的每一个状态有关，这样我们的环境转化模型非常复杂，复杂到难以建模。因此，可以将微电网优化调度问题建模成一个马尔科夫决策过程(Markov Decision Process，MDP)，使转化到下一个状态 S_{T+1}的概率仅和当前状态 S_T有关，与之前状态无关。可以用一个五元组(S，A，P，$R()$，γ)来表示，其中 S是系统的状态集，A是系统的动作集，P是状态转移概率，$R()$是奖励函数，$\gamma \in [0,1)$是折扣因子。

MDP 的关键是通过不断交互试错寻找到一个可以获取最大累积奖励值的策略。奖励函数 $R()$描述的是智能体在状态 S_T下采取动作 A_T转换到下一个状态 S_{T+1}时获得的即时奖励。智能体一个具体的迭代获得的奖励可以表示为：

$$R_T = R_{T+1} + \gamma R_{T+2} + \cdots = \sum_{k=0}^{\infty} \gamma^k R_{k+T+1}\ . \tag{12}$$

式(12)中，R_T表示智能体某次迭代所获得的奖励值；R_{T+1}表示转换到下一个状态 S_{T+1}获得的奖励值；R_{T+2}表示由状态 S_{T+1}转换为 S_{T+2}获得的奖励值；$\gamma \in [0,1)$是折扣因子，表示后面所经历的状态受当前状态影响的衰减性。

下面我们将定义本文所研究问题的状态空间 S、动作空间 A和奖励函数 $R()$。

2.1　状态空间

状态空间是由智能体在每个时间 T与微电网环境进行交互产生的。状态空间包括当前时间 T、温度 T_T、风力发电量 G_T、用户所需负载 $L_T^{(i)}$、当前市场所定电价 P_T、恒温控制负载组的平均电荷状态 $w_{\mathrm{SC},T}$及当前时刻储能系统的荷电状态 $w_{\mathrm{SOC},T}$。定义的状态由 S_T表示如式(13)：

$$S_T=\left(T_T, w_{\mathrm{SC},T}, G_T, L_T^{(i)}, P_T, T, w_{\mathrm{SOC},T}\right). \tag{13}$$

2.2 动作空间

根据T时刻下的微电网状态S_T,系统对风力所发出的电量和与主电网的交互进行调度来实现微电网系统的能量分配。

$$A_T\in\left\{A_{\mathrm{tcl}}, A_{\mathrm{price}}, A_{\mathrm{D}}, A_{\mathrm{E}}\right\}. \tag{14}$$

式(14)中,A_{tcl}是对恒温控制负载进行动作,A_{price}是根据用户用电量进行价格动作,A_{D}是当前微电网电量不足进行与微电网交互动作,A_{E}是当前微电网电量过剩进行储能动作。

2.3 奖励函数

本文的目标是对微电网的能量进行调度,期望得到最小的运行成本。定义T时刻下的微电网总运行成本为奖励值。

$$R_T=P_T\sum_{i=1}^{w_{\mathrm{loads}}}L_{\mathrm{load}}^{(i)}+C_{\mathrm{gen}}\sum_{j=1}^{w_{\mathrm{TCLS}}}L_{\mathrm{tcl}}^{(j)}u_{\mathrm{b,t}}^{(j)}+P_T^{(\mathrm{D})}E_T^{(\mathrm{S})}-C_{\mathrm{gen}}G_T-\left(P_T^{(\mathrm{U})}+C_{\mathrm{tr\,imp}}\right)E_T^{(\mathrm{P})}-C_{\mathrm{tr\,exp}}E_T^{(\mathrm{S})}. \tag{15}$$

3 深度强化学习算法

3.1 深度强化学习算法分类介绍

深度强化学习算法分为Model-Free和Model-Based两种。Model-Free指无模型即不需学习和理解环境,Model-Based需要对环境进行建模,这种模型能够从它的观察角度描述环境是如何工作的,然后利用这个模型做出动作规划。Model-Free方法又可以分为基于值函数和基于策略函数两大类深度强化学习算法,图2展示了Model-Free方法的具体分类。

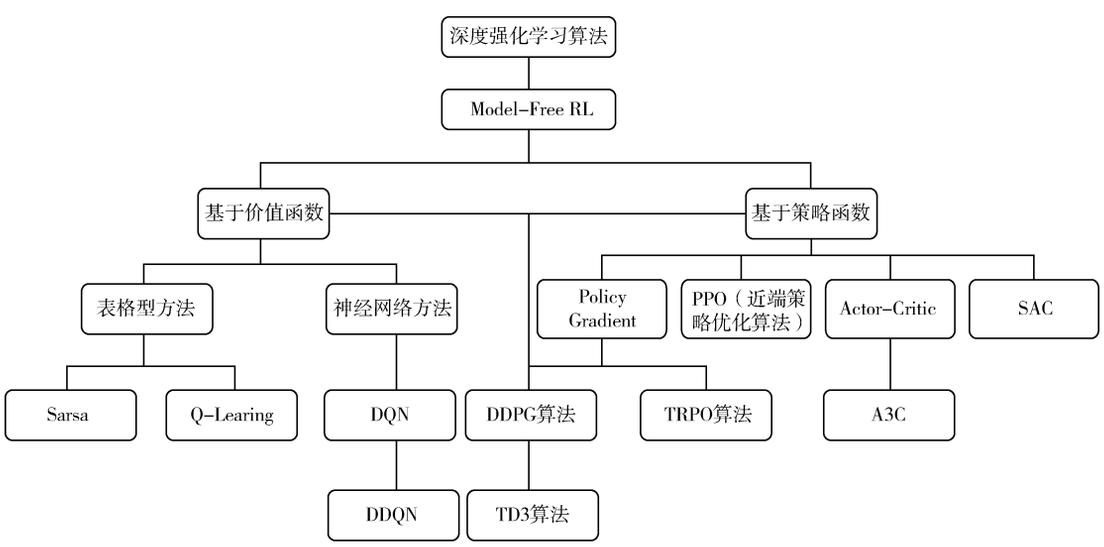

图2 Model-Free RL算法分类

Fig. 2 Diagram of Model-Free RL algorithm classification

3.2 DQN(Deep Q-Network)算法

DQN是基于值函数DRL算法的一个代表算法[15],其将当前时刻环境的状态State输入进神经网络中,通过神经网络获得近似值函数,输出一个关于动作价值的Q网络$Q(s,a)$。再通过ε-Greedy策略来输出当前状态下对应的动作Action。采用此动作继续与环境进行交互,得到当前动作的奖励值Reward以及动作后环境的状态值,这便是DQN算法与环境的一次交互。算法的核心便是通过不断交互得到的数据去优化神经网络的参数来使动作更加精确。

DQN算法有经验回放机制和固定Q目标网络两大创新处。

在与环境进行交互时,获得的数据具有很强的关联性,直接采用这些数据进行网络训练,可能导致模型

难以收敛。因此，DQN算法通过使用经验回放机制，将每一次交互得到的样本数据一起储存起来。在进行网络训练时，随机抽取一定量的样本数据进行训练，这样大大降低了样本之间的关联性，同时也提高了样本的利用率。

因为网络会不断地进行更新，所以相同状态和动作下的Q-Target网络和Q-Current网络的参数是不固定的，这样训练起来比较困难。DQN将Q-Target网络参数固定，这样问题就变成了一个回归问题：用Q-Current去逼近Q-Target。具体实现时，使用两个结构相同但参数不同的Q网络，Q-Current的网络参数是不断更新的；而Q-Target目标网络的参数则是相对固定的，每隔一段时间才会从Q-Current网络中复制更新参数。这种方式降低了Q-Target和Q-Current两个网络之间的运算量。DQN通过这两种网络提高了DQN算法的鲁棒性和准确性。

3.3　AC(Actor-Critic)算法

图3为AC算法的框架图。AC算法将值函数网络和策略网络结合使用。策略梯度(Policy Gradient，PG)可以在连续动作空间上进行动作，但其在每一个回合结束后才进行参数的训练，无法单步更新，学习效率比较慢，而Q-Learning可以进行单步更新，但动作只能局限在离散空间上，二者优势互补，将二者相结合形成了AC算法。

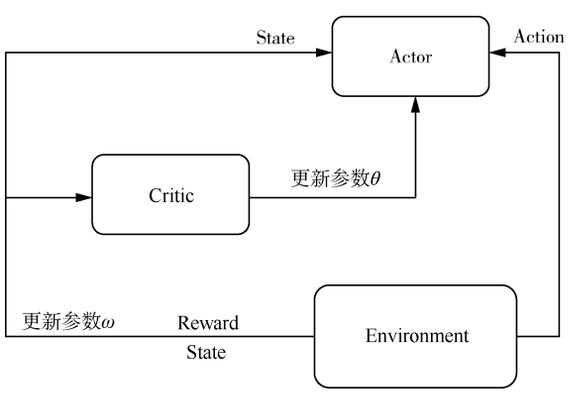

图3　AC算法的框架图

Fig. 3　Frame diagram of the AC algorithm

AC算法由两个神经网络组成，$\pi(a,s;\theta)$为策略网络Actor，θ为网络参数；$q(s,a;\omega)$为价值网络Critic，ω为网络参数。在与环境进行交互时，Actor策略网络根据环境状态输出动作。Critic价值网络需要输入状态和动作，该网络的作用是对Actor网络所选的动作进行评判。Actor网络参数的更新目标是为了获取更大的Critic网络输出值，使策略不断优化。Critic网络参数的更新目标是通过环境给出的奖励值来对自己的打分系统进行优化，使打分更加标准，从而不断优化策略网络的动作。

3.4　改进的A3C(Asynchronous Advantage Actor Critic)算法

A3C算法为异步的优势行动者评论家算法。前面提到的DQN算法使用到了经验回放机制，尽管随机提取样本在一定程度上降低了样本的相关性，但样本之间仍存在一定的相关性。A3C算法通过多线程方法与环境进行交互，将多个Actor学习到的经验集中起来用于共同学习，这样避免了经验回放相关性过强的问题，又做到了异步并发的学习模型。同时，在进行训练时，从样本空间抽样过的样本和新采样的样本对网络训练的重要性是不相同的，因此，改进的A3C算法由均匀性采样改进为重要性采样。样本根据TD Error进行降序处理，TD Error越大，被抽样的概率越大。新采样的样本具有最高的优先级，每采样一次后，更新其TD Error。图4为A3C算法的框架图。

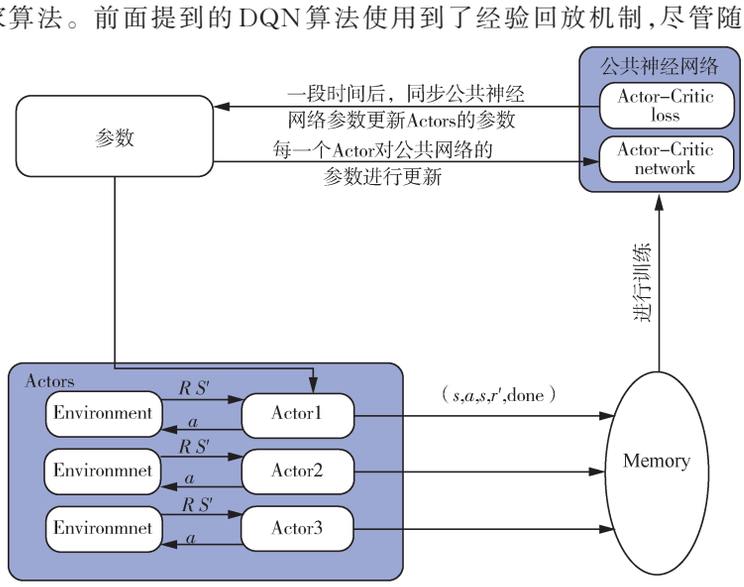

图4　A3C算法的框架图

Fig. 4　Frame diagram of the A3C algorithm

图4中Global Network是一个公共的神经网络，多个Actor线程都有和Global Network一样的网络结构，互相独立的每个线程都与环境进行交互获得经验数据，当数据积累到一定量时，每个线程通过计算自己

线程里的神经网络损失函数的梯度分别去更新公共神经网络的参数。一段时间后,独立的各个线程同步公共神经网络的参数来更新线程的网络参数。经过多次训练使模型收敛。表1显示了一个线程的训练流程。

表1 一个线程的训练流程

Tab. 1 Training flow of one thread

1)初始化环境,更新时间序列t=1,重置公共神经网络的梯度更新量d_θ,d_ω为0。

2)线程同步公共神经网络得到参数$\theta'=\theta$,$\omega'=\omega$;初始化环境s_t,$t_{\text{start}}=t$。

3)基于策略网络$\pi(a_t, s_t; \theta)$选择动作a_t。

4)与环境交互,输入动作a_t,得到奖励$R()$和下一状态s_{t+1}。

5)更新时间$t=t+1$。

6)如果交互后训练进入终止状态或达到最大训练步转入步骤7),否则回到步骤2)进行循环。

7)计算最后一个时间序列s_t的$Q(s,t)$:

$$Q(s,t)=V(s_t,\omega').$$

8)根据TD Error进行采样更新网络参数。

a)计算每一时刻的$Q(s,i)$:

$$\text{for } i\in(t-1,t-2,\cdots,t_{\text{start}}),$$
$$Q(s,i)=r_i+\gamma Q(s,i+1).$$

b)累计Actor的本地梯度更新参数θ:

$$d_\theta=d_\theta+\nabla_{\theta'}\log\pi(a_i,s_i;\theta')(Q(s,i)-V(s_i,\omega^i))+c\nabla_{\theta'}H(\pi(s_i;\theta')).$$

c)累计Critic的本地梯度更新ω:

$$d_\omega=d_\omega+\frac{\partial(Q(s,i)-V(s_i,\omega^i))^2}{\partial\omega'}.$$

9)更新全局神经网络的模型参数:

$$\theta=\theta-\alpha d_\theta,\ \omega=\omega-\alpha d_\omega.$$

10)如果达到最大训练时长,结束输出公共部分的神经网络参数θ、ω,否则进入步骤2)。

4 实验结果和分析

4.1 实验设置

对上述的3种算法进行验证,在Python环境下以芬兰的微电网系统为对象进行仿真,3种RL算法是由TensorFlow进行神经网络训练。微电网结构如图1所示,微电网中设备的一些运行参数如表2所示。

表2 微电网设备的运行参数

Tab. 2 Operating parameters of microgrid equipment

单元	参数设置	数值
ESS储能装置	充电系数η_c	0.9
	放电系数η_d	0.9
	最大充电功率C_t/kW	250
	最大放电功率D_t/kW	250
风力发电装置	风力发电量/(kW·h)	参考4.2节图6
	风力发电费用/[欧元·(MW·h)$^{-1}$]	32
输电装置	主电网输电输入费用/[欧元·(MW·h)$^{-1}$]	9.7
	微电网输电输出费用/[欧元·(MW·h)$^{-1}$]	0.9
恒温控制负载	机组数/台	100
购电售电价格	随环境变化	参考4.2节图8

4.2 训练时间和平均奖励值

在Python环境中分别对3种RL算法进行训练,表3记录了3种算法的训练时间和10天内的总奖励值以及平均每小时的奖励值。图5记录了改进的A3C算法、AC算法和DQN三种算法模型模拟微电网运行10天的奖励曲线图。

表3　3种算法的训练结果

Tab. 3　Training results of 3 algorithms

算法	DQN算法	AC算法	改进A3C算法
训练时间/min	7.38	4.24	1.46
10天总奖励值	−4.497 0	1.559 8	1.991 8
平均每小时奖励值	−0.018 7	0.006 5	0.008 1

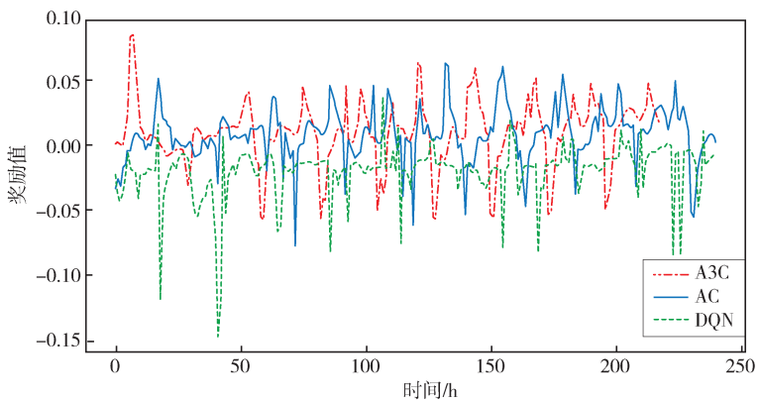

图5　三种算法10天的奖励值曲线图

Fig. 5　Diagram of 10-day reward value curve for three algorithms

3种算法模型训练过程中,损失值逐渐下降直至趋于稳定时,训练结束。表3展示了3种算法训练至收敛所用的时间,其中采用多线程训练的A3C算法所用时间最少。采用训练好的3种算法模型模拟微电网运行10天,所获得的奖励曲线如图5所示,图中DQN算法模型给出策略的奖励值曲线总体低于其他两种算法模型的。通过表3的数据对比看出,较其他两种算法,改进的A3C算法模型给出的策略的奖励值最大。

分析表明,改进的A3C算法所用的训练时间最短,训练模型给出的策略获得的平均奖励值最大。3种算法模型都可以有效地给出调度策略,但不同算法给出的策略优劣不同。下面我们分别采用3种算法模拟微电网的一天的运行情况,来分析一天内具体的调度策略。

4.3　3种模型模拟微电网一天的运行情况

使用训练好的模型,我们获取了该微电网某一天的运行情况,图6展示了微电网当日的环境温度、风力发电量和微电网预期所需电量,图7展示了改进A3C算法模型下储能装置的荷电状态。通过图6的风力发电曲线图看出,在15时风力发电出现波动,发电量正在逐渐减少。

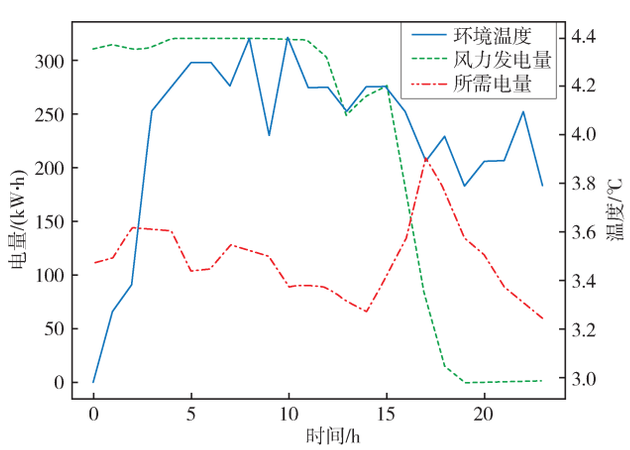

图6　环境温度,风力发电量和预期所需电量

Fig. 6　Diagram of ambient temperature, wind power generation and expected electricity demand

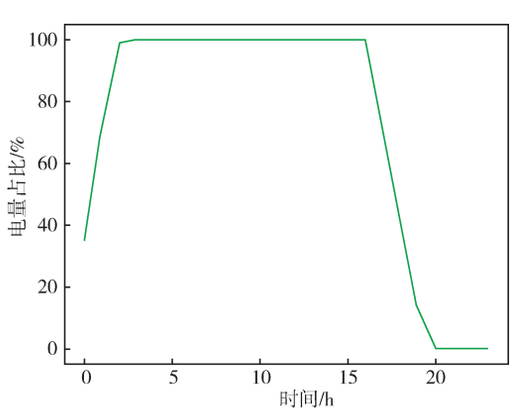

图7　改进A3C算法模型下储能装置的荷电状态

Fig. 7　Diagram of the state of charge of the energy storage device under improved A3C algorithm model

图 8—图 10 分别是在改进 A3C 算法、AC 算法和 DQN 算法模型给出的调度策略下，微电网与主电网的电力交互曲线图以及电力价格曲线图。在风力发电出现波动时，DQN 算法给出的调度策略在 17 时从主电网购入大量的电力，AC 算法给出的调度策略在 18 时购入大量电力，而在 17 时、18 时这两个时间段内电网售电价格高于其他时间段，这时大量购电会降低微电网的经济效益。相对比而言，改进的 A3C 算法给出的调度策略充分考虑了购电和售电的价格，图 8 表示在 17 时、18 时电价高时，调度策略控制储能装置放电 40% 来减少从主电网的购电量，而 20 时以后电价低时再大量购电。通过对微电网一天运行情况的模拟看出：改进的 A3C 算法模型在应对风电波动时可以给出更加有效的调度策略，并有效地减少风电波动对微电网的影响，保证微电网可以安全经济地运行。

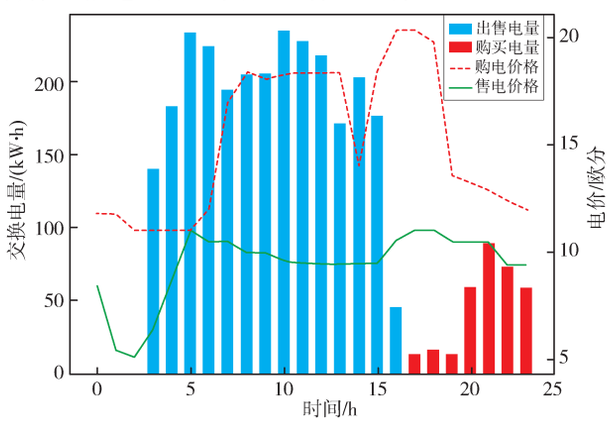

图 8 改进 A3C 算法模型微电网与主电网的电力交互曲线

Fig. 8 Power interaction curve between microgrid and main grid under the improved A3C algorithm model

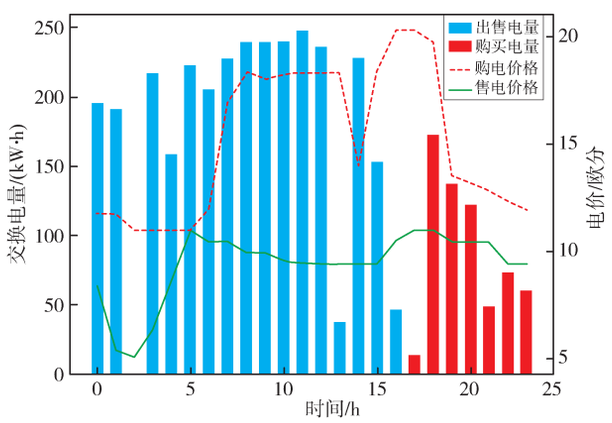

图 9 AC 算法模型微电网与主电网的电力交互曲线

Fig. 9 Power interaction curve between microgrid and main grid under the A3C algorithm model

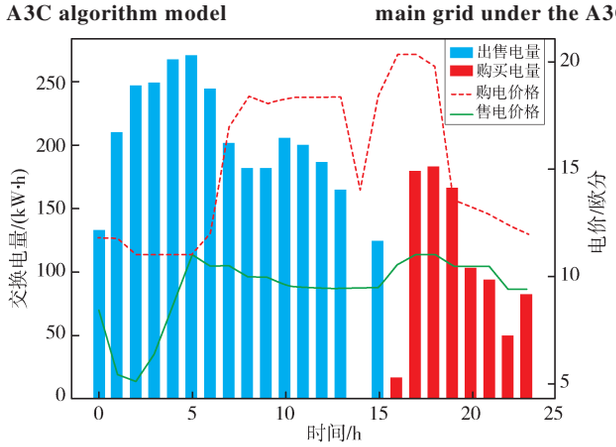

图 10 DQN 算法模型微电网与主电网的电力交互曲线

Fig. 10 Power interaction curve between the microgrid and the main grid under the DQN algorithm model

5 结束语

以微电网经济运行费用最小为优化目标，在基于 AC 算法的框架上，提出了一种改进的 A3C 算法，基于马尔科夫决策过程将微电网优化调度问题进行建模分析，通过不断与微电网环境进行交互试错获取大量的调度经验来训练神经网络。利用训练好的神经网络生成决策来控制储能装置的动作和其与主电网进行电力交互，实现微电网经济调度优化。分别用改进 A3C 算法、AC 算法和 DQN 算法模型对芬兰的微电网系统进行仿真分析，仿真结果表明，当微电网风力发电出现波动时，相对比而言，改进的 A3C 算法训练时间更短，算法模型给出的调度策略经济效益更高，可以有效地降低风电波动性对微电网安全经济运行的影响。

参考文献：

[1] 张丽，戚维燕，耿凯旋，等. 可再生能源的发展路线与影响研究[J]. 世界石油工业，2019，26(5)：29-34.
ZHANG Li，QI Weiyan，GENG Kaixuan，et al. Development Route and Influence of Renewable Energy[J]. World Petroleum Industry，2019，26(5)：29-34.

[2] 王成山，李鹏. 分布式发电、微网与智能配电网的发展与挑战[J]. 电力系统自动化，2010，34(2)：10-14，23.
WANG Chengshan，LI Peng. Development and Challenges of Distributed Generation，the Micro-Grid and Smart Distribution System[J]. Automation of Electric Power Systems，2010，34(2)：10-14，23.

[3] 于建成，迟福建，徐科，等. 分布式电源接入对电网的影响分析[J]. 电力系统及其自动化学报，2012，24(1)：138-141.
YU Jiancheng，CHI Fujian，XU Ke，et al. Analysis of the Impact of Distributed Generation on Power Grid[J]. Proceedings of the Chinese Society of Universities for Electric Power System and Its Automation，2012，24(1)：138-141.

[4] 刘朝阳，穆朝絜，孙长银. 深度强化学习算法与应用研究现状综述[J]. 智能科学与技术学报，2020，2(4)：314-326.
LIU Zhaoyang，MU Chaoxu，SUN Changyin. An Overview on Algorithms and Applications of Deep Reinforcement Learning[J]. Chinese Journal of Intelligent Science and Technology，2020，2(4)：314-326.

[5] 宋鹏飞，杨宁，崔承刚，等. 深度强化学习应用于电力系统控制研究综述[J]. 现代计算机，2021(1)：39-44.
SONG Pengfei，YANG Ning，CUI Chenggang，et al. Survey of the Application of Deep Reinforcement Learning in Power System Control[J]. Modern Computer，2021(1)：39-44.

[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 Jinhua，KE Zhongming，ZHOU Wenhui. 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] 刘国静，韩学山，王尚，等. 基于强化学习方法的风储合作决策[J]. 电网技术，2016，40(9)：2729-2736.
LIU Guojing，HAN Xueshan，WANG Shang，et al. Optimal Decision-Making in the Cooperation of Wind Power and Energy Storage Based on Reinforcement Learning Algorithm[J]. Power System Technology，2016，40(9)：2729-2736.

[8] 余宏晖，林声宏，朱建全，等. 基于深度强化学习的微电网在线优化[J/OL]. 电测与仪表：1-7(2021-10-22)[2023-02-14]. http://kns.cnki.net/kcms/detail/23.1202.TH.20211021.1651.007.html.

[9] 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.

[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] 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.
JI Ying，WANG Jianhui. Online Optimal Scheduling of a Microgrid Based on Deep Reinforcement Learning[J]. Control and Decision，2022，37(7)：1675-1684.

[12] WATKINS C J C H，DAYAN P. Q-Learning[J]. Machine Learning，1992，8(3/4)：279-292.

[13] EBELL N，HEINRICH F，SCHLUND J，et al. Reinforcement Learning Control Algorithm for a PV-Battery-System Providing Frequency Containment Reserve Power[C]//2018 IEEE International Conference on Communications，Control，and Computing Technologies for Smart Grids (SmartGridComm). October 29-31，2018，Aalborg，Denmark. IEEE，2018：1-6.

[14] FRANCOIS-LAVET V，TARALLA D，DERNSTet al. Deep Reinforcement Learning Solutions for Energy Microgrids Management [C/OL]//European Workshop on Reinforcement Learning EWRL 2016. (2016-11-28)[2023-02-14]. https://hdl.handle.net/2268/203831.

[15] 刘建伟，高峰，罗雄麟. 基于值函数和策略梯度的深度强化学习综述[J]. 计算机学报，2019，42(6)：1406-1438.
LIU Jianwei，GAO Feng，LUO Xionglin. Survey of Deep Reinforcement Learning Based on Value Function and Policy Gradient[J]. Chinese Journal of Computers，2019，42(6)：1406-1438.

[责任编辑：王　静]