

# 基于深度强化学习的微网储能系统控制策略研究

梁宏<sup>1</sup>, 李鸿鑫<sup>2</sup>, 张华赢<sup>2</sup>, 胡子珩<sup>2</sup>, 秦兆铭<sup>1</sup>, 曹军威<sup>3</sup>

- (1. 清华大学自动化系, 北京市 海淀区 100084;  
2. 南方电网公司新型智慧城市高品质供电联合实验室(深圳供电局有限公司), 广东 深圳 518020;  
3. 清华大学北京信息科学与技术国家研究中心, 北京市 海淀区 100084)

## Research on Control Strategy of Microgrid Energy Storage System based on Deep Reinforcement Learning

LIANG Hong<sup>1</sup>, LI Hongxin<sup>2</sup>, ZHANG Huaying<sup>2</sup>, Hu Ziheng<sup>2</sup>, QIN Zhaoming<sup>1</sup>, CAO Junwei<sup>3</sup>

- (1. Department of Automation, Tsinghua University, Haidian District, Beijing, 100084;  
2. New Smart City High-quality Power Supply Joint Laboratory of China Southern Power Grid (Shenzhen Power Supply CO., LTD), Shenzhen, Guangdong, 518020;  
3. Beijing National Research Center For Information Science And Technology, Haidian District, Beijing, 100084)

**ABSTRACT:** As a new form of energy management, microgrid has developed rapidly in recent years. In order to ensure the safe, stable and economical operation of microgrid system, it is important to provide reasonable energy scheduling strategy for it. According to different operating modes, micro-grids can be divided into two categories: grid-connected microgrids and island microgrids. This paper took the grid-connected microgrid as the object, applied Simulink simulation technology to build a microgrid system including external power supply, photovoltaic power generation, energy storage, and load according to the principle of constant power control (PQ control). Then based on this simulation system, combined with the double deep Q network algorithm, trained an optimization strategy of energy storage control problem, which goal was to minimize the 24-hour electricity cost while meet the voltage stability of the microgrid, power balance and the constraint of state of charge of the energy storage. Through experimental verification, the rationality of the energy storage strategy was analyzed from a qualitative perspective, and the effectiveness of the method proposed in this paper was demonstrated from a quantitative perspective.

**KEY WORDS:** microgrid; deep reinforcement learning; Simulink; PQ control; energy storage control; optimization strategy

基金项目: 南方电网公司科技项目(090000KK52190169/ SZKJXM2019669).  
Project Supported by Science and Technology Project of China Southern Power Grid(090000KK52190169/ SZKJXM2019669).

**摘要:** 微网作为新兴的能源管理形式, 近年来发展迅速, 为保障微网系统能安全、稳定、经济的运行, 为其提供合理的能量调度策略是关键。微网根据运行模式的不同, 可分为并网微网和孤岛微网两大类。本文以并网微网为研究对象, 应用 Simulink 仿真技术, 按照恒功率控制(PQ控制)原理, 搭建了一个包含外部电源、光伏发电、储能, 以及负荷的微网系统。然后以此仿真系统为基础, 结合深度强化学习中的双重深度Q学习算法, 以最小化微网24小时从外电网取电费用为目标, 在满足微网系统电压稳定性、功率平衡以及储能的荷电状态等约束下, 以储能的实时充放电功率为控制变量, 训练得到储能控制的优化策略。并通过实验验证与传统方法进行对比, 分别从定性的角度分析了储能充放电策略的合理性, 和从定量的角度展示了本文所提方法在优化购电费用上的有效性。

**关键词:** 微网; 深度强化学习; Simulink 仿真; PQ控制; 储能控制; 优化策略

**DOI:**

## 0 引言

微网是为解决光伏发电、风力发电等分布式能源并网问题而快速发展起来的小型发配电系统<sup>[1]</sup>。为了维持微网的功率平衡, 降低分布式能源出力不确定性对微网的影响, 配置储能装置十分必要。因此典型的微网通常包含各种可再生能源发电装置、储能装置、负荷等组件, 以及能量转换器、控制器等辅助设备<sup>[2]</sup>。对微网的研究包含众多方面, 如对微网体系架构的研究、电力电子技术研究、投资和

运营成本研究、安全与运维研究、能量调度控制策略研究等等。其中储能控制策略作为微网能量调度的研究重点被广泛关注。

经典优化方法<sup>[3,4]</sup>、基于规划的方法<sup>[5,6]</sup>、启发式算法<sup>[7,8,9]</sup>等都可以用来处理微网的能量调度控制问题,但是当实际场景较为复杂难以用明确数学模型描述,且数据量巨大时,上述算法将很难适用。近年来随着人工智能的发展而兴起的深度强化学习算法则为处理复杂微网控制问题提供了新的思路。深度强化学习是实现人工智能的关键技术,随着 2016 年谷歌 Deep Mind 创建的 AlphaGo 的问世,深度强化学习作为处理序列控制问题的有效手段被研究者们广泛应用并发展<sup>[10,11,12]</sup>。

微网能量控制问题作为典型的时序控制问题,符合深度强化学习求解框架,目前已有很多突出的工作。文献[13]给出了基于异步优势的深度强化学习算法在求解无模型、多目标、多微网的能量调度控制问题中的应用,并通过具体数值算例,说明深度神经网络的搭建依据。文献[14]以维护孤岛微网系统稳定性和运行经济效益为目标,通过深度强化学习算法给出微网的复合储能协调控制策略。文献[15],建立了基于能量总线的微网框架模型,对比了深度 Q 学习算法相较于启发式算法在能量调度控制问题上的优势。文献[16]针对微网光伏出力的随机性,采用神经网络预测的数据进行强化学习训练,降低微网用电成本。文献[17]综合考虑负荷、光伏、实时电价的波动和潮流计算,提出了一种基于模型的近似动态规划算法,然后用卷积神经网络来近似深度强化学习中的值函数。文献[18]从电网运行安全角度出发,提出一种基于深度强化学习的电网切机控制策略。文献[19]从分销系统运营商的角度出发,以降低需求侧峰值比和保护用户隐私为目标,通过蒙特卡洛强化学习算法给出零售定价策略。文献[20]在智能电网环境下探索了使用深度强化学习对建筑能源管理系统的进度进行在线优化的好处,并在大型 Pecan Street Inc. 数据库上验证了方法的有效性。

与已有工作不同,本文应用深度强化学习算法求解储能的充放电控制策略时,采用的训练数据是通过 Simulink 搭建的微网仿真环境运行出来的实时数据,在仿真环境中,可以考虑微网各个组成模块的有功功率和无功功率,并且应用 PQ 控制策略保证有功和无功的可控性。用仿真环境运行得到的信息进行强化学习训练,可以添加电压、电流、相角等更多状态信息,从而更全面的考虑各种状态信

息对目标函数的影响,相比于数学公式逼近的状态数据更加真实可靠。本文结构安排如下,首先在第一部分给出所研究的微网对象的组成和各个模块的 Simulink 仿真设计,然后在第二节重点介绍所用深度强化学习算法框架以及其在储能控制问题上的应用流程。在第三节实验验证中,通过与已有算法对比来说明本文算法流程的有效性。最后给出总结与下一步工作展望。

## 1 微网组成与 Simulink 仿真设计

本文考虑并网运行的微网的能量调度问题,给出典型微网的拓扑结构如图 1 所示。

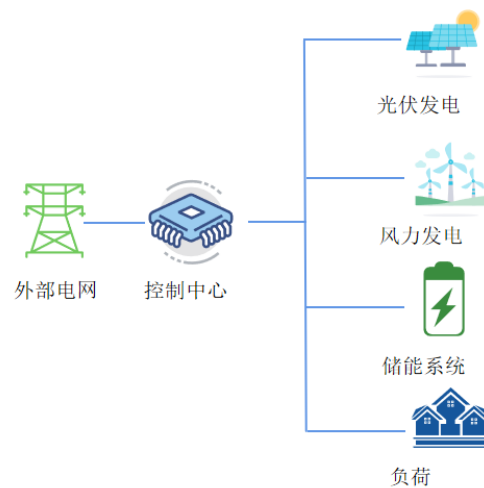


图 1 微网结构图

Fig.1 Microgrid structure diagram

下面给出各个组成模块的详细介绍。

### 1.1 微网组成模块介绍

(1) 光伏发电: 得益于太阳能发电技术的成熟,光伏发电成为微网系统分布式能源的重要组成部分,因此本文以光伏发电作为分布式发电的代表进行详细讨论。光伏发电功率可由下述公式给出:

$$P_{PV}(t) = R(t) \times S \times \eta_{PV} \quad (1)$$

其中  $R(t)$  表示时间  $t$  时刻的太阳辐射强度,单位为  $W/m^2$ ,  $S$  表示布置的太阳能电池板的面积。太阳辐射强度与电池板面积的乘积则为太阳能辐射功率,再乘以光伏发电的转换效率  $\eta_{PV}$ ,就得到了  $t$  时刻光伏发电的功率。

(2) 储能系统: 由于光伏和风力等分布式发电模式具有间歇性等不稳定的特点,因此需要添加储能系统来维护微网稳定运行和平滑系统功率。蓄电池作为高效的储能设备被广泛应用,蓄电池中能

量状态满足下述公式:

$$E_{bat}(t) = E_{bat}(t-1) + P_{bat}(t) \times \Delta t \quad (2)$$

其中  $P_{bat}(t)$  表示  $t$  时刻蓄电池的充放电功率,  $\Delta t$  表示两次充放电动作之间的时间间隔。在微网系统中, 储能系统的充放电功率通常作为重要的控制变量参与微网能量调度, 也是本文的控制重点。

(3) 负荷: 负荷是微网系统中消耗电能部分的总称, 针对固定微网系统, 负荷需求由当地气候环境和微网性质决定, 因此通常不可调节, 在本文能量调度问题中, 负荷曲线作为固定量输入到微网系统。

### 1.2 Simulink 仿真设计

本节详细介绍微网系统各组成模块的仿真模型搭建原理<sup>[21]</sup>。与图 1 微网结构相对应, 图 2 是利用 Matlab/Simulink 工具箱, 搭建的微网系统的仿真模型。下面对各个仿真模块进行详细说明。

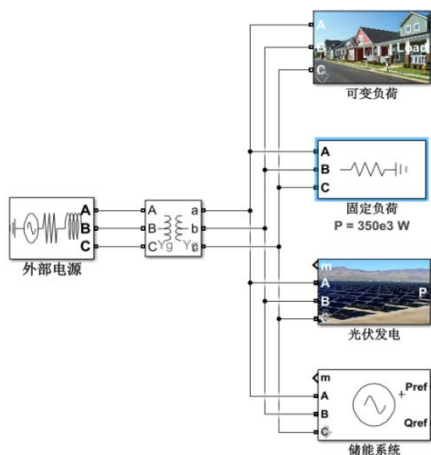


图 2 微网仿真结构图

Fig. 2 Microgrid simulation structure diagram

(1) 外部电源: 本文所用外部电源由 Simulink 自带的三相交流电源模块替代, 模块的参数设置需要根据实际仿真需求给出, 具体包括线电压有效值、A 相初相相角、频率、内部连接方式等。然后通过变阻器模块与微网系统内部模块相连, 为系统内部提供所需电能。

(2) 光伏发电: 本文考虑的微网系统是在并网模式下工作, 为了保证电网发电和用电的平衡, 要求光伏发电系统的有功功率与无功功率可调节, 因此针对光伏发电采用 PQ 控制策略<sup>[22]</sup>。具体模型如图 3 所示。其中输入端口 1 的光伏功率由上一节中的光伏功率计算公式给出, 然后通过一个三相动态负荷模块实现 PQ 控制, 最后在输出端口 1 输出有功功率数据。

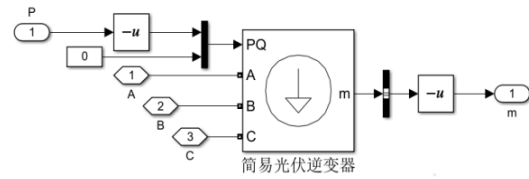


图 3 光伏系统模型

Fig. 3 Photovoltaic system model

(3) 储能系统: 储能系统的充放电功率是本文的主要控制对象, 因此要求搭建的储能系统模型可调节充放电功率, 所以同样用三相动态负荷来达到 PQ 控制效果。如图 4 所示, 从输入端口输入外部控制指令, 然后通过一个计算模块, 检查电池是否过度充放电, 并且根据公式 (2) 计算电池的有功功率  $P$  和无功功率  $Q$ 。最后将得到的  $P$ 、 $Q$  值带入三相动态负荷模块, 得到储能系统最终输出的有功功率和无功功率。

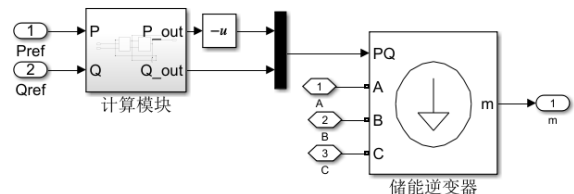


图 4 储能系统模型

Fig. 4 Energy storage system model

(4) 负荷: 在仿真实验中, 由于负荷的随机性, 微网负荷通常由两部分组成, 其中一部分是可变负荷, 模型如图 5 所示。从输入端口输入负荷功率数据, 通过计算式  $Q^2 + P^2 = P^2 \mu^2$  计算得到  $Q$  值, 其中  $\mu$  表示功率因数。最后通过三相动态负荷模块得到输出结果。另一部分是固定负荷, 在仿真模型中直接由 Simulink 中的电阻模块充当。

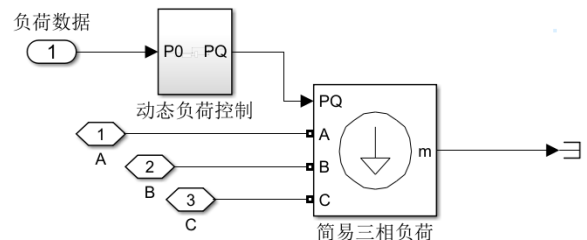


图 5 可变负荷模型

Fig. 5 Variable load model

综上, 本文利用 Simulink 工具箱为并网运行的微网系统搭建了一套完整的仿真模型。为了保证微网系统与外界电网交互的稳定性与可控性, 各个分布式能源的仿真模块均采用 PQ 控制策略进行控制。本文在此仿真模型的基础上进行深度强化学习训练, 可以综合考虑系统电压、各个模块的有功功

率、无功功率、电池的 SOC 状态等各个状态变量对微网系统从外电网购电费用的影响,从而得到优化的储能系统充放电控制策略,相较于传统的以数学公式为基础的强化学习训练,可以综合考虑更加丰富的状态变量对控制目标的影响。

## 2 深度强化学习算法

强化学习是以最大化预期回报为目标,通过智能体(Agent)和环境(Environment)的不断交互,得到环境的状态变量(State)与智能体的动作变量(Action)之间的映射关系,为智能体提供优化的动作策略(Policy)。深度强化学习则是在传统强化学习的基础之上,利用深度神经网络来创建状态变量与动作变量之间的对应关系。由于深度神经网络强大的表达能力,深度强化学习可以处理更加复杂、更加接近实际的策略决策问题。近年来,深度强化学习在优化控制、机器人控制、自动驾驶等领域的应用已经取得了突破性的进展。

深度强化学习通常是基于马尔可夫决策过程的(Markov Decision Process, MDP),即智能体与环境在下一时刻的交互结果只与当前的环境状态有关,而与之之前的环境状态无关。传统的 MDP 过程包含四个要素,由  $(S, A, R, \pi)$  四元组给出<sup>[23]</sup>。其中  $S$  表示环境状态的集合,  $A$  表示表示智能体可以采取的动作的集合,  $R$  表示奖励函数,即智能体在某一状态下采取某种动作后得到的奖励,  $\pi$  表示策略集合,是状态和动作之间的对应关系。

强化学习优化策略的求解算法可以分为基于值函数求解、基于策略梯度求解、基于搜索与监督求解三大类<sup>[24]</sup>。本文重点介绍一下基于价值函数的求解方法,典型的有动态规划算法、蒙特卡洛算法、时序差分算法等,其中动态规划算法适合求解有模型且状态空间维数较低的情况。蒙特卡洛算法是一种不基于模型的简易算法,但是缺点是需要完整的状态序列信息,这在很多非周期的过程中是很难获取的。时序差分方法则不需要完整的状态序列就可以逼近价值函数,经典的时序差分方法包括在线差分算法 SARSA 和离线差分算法 Q-Learning 算法,这两种方法都是通过维护一个 Q 表的方式求解小型强化学习问题。当状态和动作空间是连续的或者是规模非常大的离散的情况下,需要维护的 Q 表异常巨大,会给存储带来困难。但是神经网络的发展,使得这一问题得到解决,用深度神经网络代替 Q 表,得到更加适合于复杂问题的深度强化学习算法,一类典型的算法为深度 Q-Learning (Deep Q

Network, DQN) 算法。

### 2.1 算法简介

本文的微网控制问题采用的是 DQN 的改进版本,即双重深度 Q-Learning (Double Deep Q Network, DDQN) 算法<sup>[25]</sup>。

Q-Learning 算法是通过下面的更新公式来更新定义的 Q 函数。

$$Q(S, A) = Q(S, A) + \alpha(R + \dots \gamma \max Q(S', a) - Q(S, A)) \quad (3)$$

其中  $Q(S, A)$  表示动作价值函数,  $\alpha$  表示学习率,当上述更新公式收敛时,可以得到强化学习的最优控制策略。

DQN 则是用深度神经网络  $Q(S, A | \omega)$  来代替 Q-Learning 中的 Q 函数,如图:

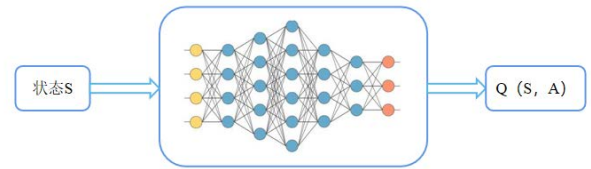


图 6 神经网络估计 DQN 价值函数

Fig. 6 Estimation of value function for DQN

然后由 (3) 式计算出当前目标 Q 值,根据当前目标 Q 值与 Q 网络给出的 Q 值之间的均方误差来更新神经网络的参数  $\omega$ 。

DDQN 算法则是在 DQN 的基础上进行两点改进。首先是引入两个网络,一个是目标网络  $Q'(S, A | \omega')$  用来计算目标 Q 值,另一个是更新网络  $Q(S, A | \omega)$  用来更新 Q 值,这样可以减小目标 Q 值和更新网络参数之间的依赖关系。目标网络与 Q 值网络具有完全相同的结构,并且定时从 Q 值网络同步参数。

其次是解耦目标 (3) 式中目标 Q 值的动作选择和目标 Q 值的计算过程,从而减小因为贪婪算法导致的过度估计。具体做法是计算目标 Q 值时,不再从目标 Q 网络中找出动作对应的最大 Q 值,而是先在更新网络中找出最大 Q 值对应的动作:

$$\tilde{a}(S' | \omega) = \arg \max_{a \in A} Q(S, A | \omega) \quad (4)$$

然后利用这个动作  $\tilde{a}(S' | \omega)$  在目标网络  $Q'(S, A | \omega')$  中计算目标 Q 值:

$$Q' = R + \gamma Q'(S', \tilde{a}(S' | \omega) | \omega') \quad (5)$$

通过以上两点改进,使得 DDQN 算法解决了传



统 DQN 算法的强依赖与过度估计问题，其它算法流程与 DQN 相同。

### 2.2 深度强化学习在微网控制问题中的应用

本节中，首先给出微网控制问题的描述，然后给出如何将微网控制问题转化成适合深度强化学习求解的形式。

#### 2.2.1 控制问题描述：

(1) 控制目标：

根据微网系统的不同需求，可以设置不同的控制目标，比如最小化污染物排放量、最小化燃料发电成本、最小化电压偏移、最小化网络有功损耗，或者做大化电压稳定性等等。通常可以设置单一控制目标，或者以上多种控制目标的组合，本文所考虑的算例以最小化外电网取电费用为目标。具体形式如公式 (6)：

$$C_{ost} = \int_{t=0}^T C_{out}(t) \cdot P_{out}(t) dt \quad (6)$$

其中  $C_{ost}$  表示在时间  $[0, T]$  内，整个微网从外部获取电能的总费用， $C_{out}(t)$  和  $P_{out}(t)$  分别表示  $t$  时刻的实时电价和外电网发电功率，且  $P_{out}(t)$  满足下面功率平衡公式：

$$P_{PV}(t) + P_{bat}(t) + P_{out}(t) = P_{load}(t) \quad (7)$$

左侧皆为电源组件功率，右侧为负荷功率。

(2) 控制变量：

本文算例研究的是对微网储能系统的控制，因此控制变量选择为储能系统的充放电功率  $P_{bat}$ ，其它变量如  $P_{PV}$ 、 $P_{out}$ 、 $P_{load}$  以及仿真系统的电压  $V$ ，储能的荷电状态  $SOC$  等都是状态变量。这些状态变量的值由 1.2 节中搭建的仿真模型动态给出。

(3) 约束条件：

结合实际情况，要求储能控制问题满足功率约束和荷电状态约束，即：

$$\begin{aligned} P_{bat}^{\min} &\leq P_{bat}(t) \leq P_{bat}^{\max} \\ SOC^{\min} &\leq SOC(t) \leq SOC^{\max} \end{aligned} \quad (8)$$

综上，微网储能控制问题可以表述为：在 1.2 节搭建的仿真环境中，通过控制储能充放电功率  $P_{bat}(t)$ ，在满足约束条件 (8) 的情况下，使得从外电网取电费用  $C_{ost}$  尽可能小。

#### 2.2.2 算法流程

将上述控制问题转化为强化学习算法 DDQN 的求解形式，1.2 节搭建的微网仿真模型充当强化学习问题的环境，储能充放电功率控制器则为智能体，强化学习的目标是通过智能体和环境的不断交

互来得到优化的控制策略。与此对应，需要确定此问题的马尔可夫状态序列四元组  $(S, A, R, \pi)$  的具体表达。

(1) 状态变量  $S = (P_{PV}, P_{out}, P_{load}, V, SOC, C_{out})$ ，是仿真环境实时给出的动态结果。

(2) 动作变量  $A = (P_{bat})$  是智能体的决策变量。

(3) 奖励函数  $R$  的构造需要结合目标函数 (6) 和约束条件 (8)。从当前状态  $s_t$  到下一状态  $s_{t+1}$  的奖励函数定义为：

$$R(s_t, t) = -[\int_t^{t+1} C_{out}(t) \cdot P_{out}(t) dt + \beta(t)] \quad (9)$$

其中第一部分是此时间段内电费消费情况，第二部分是为了刻画约束条件而设置的惩罚项，即当约束条件 (8) 满足时， $\beta(t)=0$ ，当约束条件不满足时，给  $\beta(t)$  赋值成一个量级非常大的常数。因为强化学习的目的是使奖励函数最大化，因此需要在这两部分之前加上负号。

(4) 策略函数  $\pi$  是微网的各个状态变量和储能的功率之间的对应关系，也即 DDQN 算法中所要求的深度神经网络。有关策略  $\pi$  满足的神经网络的设计本文将在数值算例中详细给出。

有了上面的一一对应关系之后，用深度强化学习算法来求解微网储能控制问题的思路就可以用图 7 清晰的给出。具体网络结构和参数设计将在下一节实验验证中详细给出。

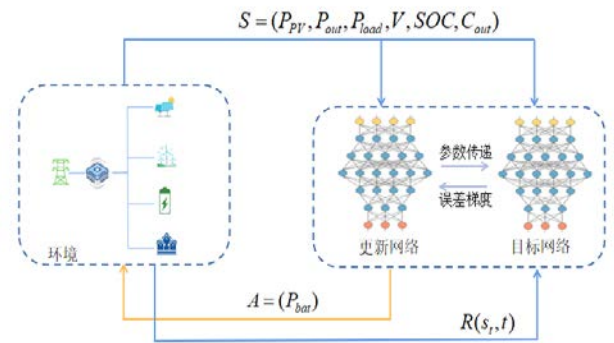


图 7 微网控制问题求解流程

Fig. 7 Solving process of Microgrid control problem

### 3 实验验证

本节通过具体数值算例来说明 DDQN 算法在求解微网储能控制问题中的应用，并与最大功率点追踪 (Maximum Power Point Tracking, MPPT) 控制进行比较，说明 DDQN 算法的有效性。

#### 3.1 实验设置

为了加速 Simulink 的仿真速度，在仿真实验验证中，本文不考虑风力发电等其它分布式发电模

块，在微网中只考虑光伏发电、储能、负荷和外电网相连。各个仿真模块的具体参数设置详见附录中的表 2。

约束条件 (8) 中的上下限约束设置为：

$$P_{bat}^{\min} = -4e5(w), P_{bat}^{\max} = 4e5(w),$$

$$SOC^{\min} = 0.19, SOC^{\max} = 0.85.$$

另外一个重要的设置就是 DDQN 算法中 Q 值估计神经网络的设置，根据经验以及不断的调试验证，本文所用神经网络为 4 层 24 节点的全连接层，并且以 relu 函数作为中间层激活函数，具体如图 8：

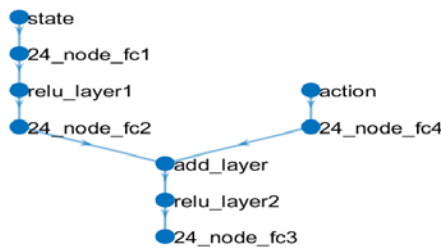


图 8 神经网络拓扑结构

Fig. 8 Topological structure of neural network

### 3.2 数值结果

根据以上设置，结合 Reinforcement Learning Toolbox 工具箱提供的深度强化学习算法与 Simulink 搭建的仿真环境完成 DDQN 算法的构建，然后在具有 Intel i7-8586U CPU 和 2G 内存独立显卡的设备上进行训练。

本文实验算例是在给定微网 24 小时的光伏发电功率曲线、负荷用电功率曲线与实时电价的前提下，通过调节储能充放电功率，来优化全天 24 小时从外电网取电费用。在进行 DDQN 算法训练时，需要将状态变量的各个分量进行归一化处理之后再带入 DDQN 的目标网络中进行训练，虑到本文的控制目标是 minimized 外电网购电费用，所以为了在训练过程中增大电价波动对控制目标的影响，本文在训练时给电价状态变量乘上 2 倍的放大系数。训练时间步长取为 300s，因此一个训练周期为 288 步。本实验在 100000 组训练周期下，得到的结果如下。

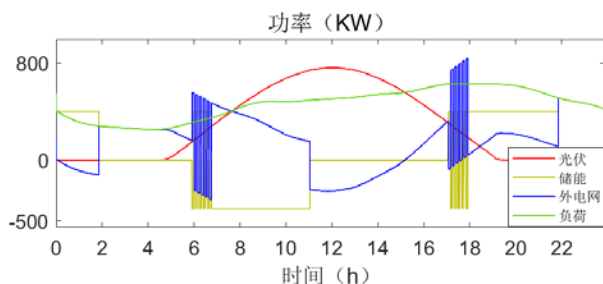


图 9 各模块有功功率波动曲线

Fig. 9 Active power fluctuation curve of each module

图 9 给出了光伏、储能、外电网、和负荷在全天 24 小时内的有功功率曲线。横坐标是时间轴，以小时 h 为单位，纵坐标表示功率，以 kW 为单位。其中光伏和负载曲线是事先给定的外部数据，储能功率是由 DDQN 训练得到的神经网络输出的策略结果，然后根据功率平衡得到外电网有功功率。

结合图 10，可以清楚地看出 DDQN 算法得到的储能充放电策略为，大约在 0h~2h 之间，此时光伏发电不工作，储能先放电，储能的 SOC 到达下限之后，此时由外电网发电，然后当光伏开始工作时，大约是 7h~11h，储能开始充电，将多余的光伏发电存储起来。在时间 11h~17h 之间储能充满电后停止工作，多余的光伏发电反馈给外电网。然后在 17h~22h 用电高峰时刻，此时实时电价升高，从外电网取电成本增高，储能在此时段放电供微电网负荷使用。综上此控制策略是在白天太阳能充足时将太阳能存储起来供晚上实时电价比较高的用电高峰时刻利用，以此来达到节约购电成本的目的。因此从逻辑上分析，此控制策略是合理的。相比于 DDQN 算法得到的控制策略，DQN 算法在相同训练周期内得到的控制策略为在 0h~1h 对储能进行放电，然后储能一直处于闲置状态直到 14h 左右对储能开始充电，然后在 18h 电价开始升高时储能进行放电。从后面的数据结果可以看出 DDQN 算法得到的控制策略的效果优于 DQN 算法得到的控制效果。

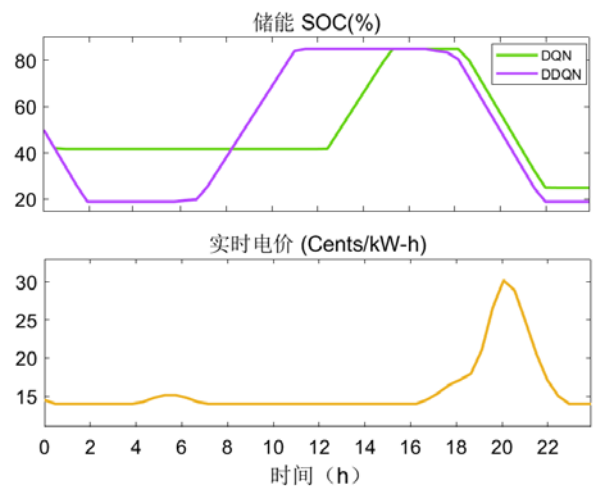


图 10 储能 SOC 与实时电价对照图

Fig. 10 Chart of SOC and real-time electricity price

前面结果给出了储能控制策略的定性分析，图 11 和表 1 则给出了定量的数据结果来说明 DDQN 算法训练得到的控制策略相比于传统控制方法和 DQN 算法更加优化。传统的基于光伏发电的 MPPT

控制,是简单的状态机控制方法,即判断微网中负荷和光伏的功率差的正负,在储能满足给定的SOC状态下,按照光伏最大功率追踪原理给出储能的充放电功率,具体算法流程见[26]。

表1 DDQN、DQN、MPPT、无储能四种情况数据结果

Tab.1 Data results of DDQN,DQN, MPPT, and no energy storage

方法	购电量(kW-h)	费用(\$)
无储能	4344.6	840.1
MPPT	4003.6	735.7
DQN	3833.5	626.9
DDQN	3762.4	613.7

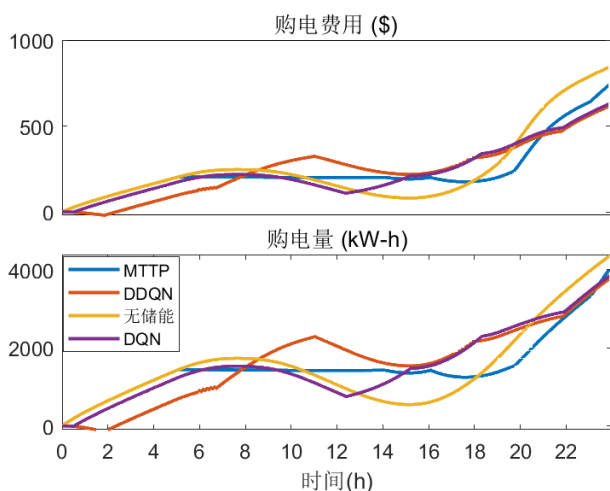


图11 DDQN、DQN、MPPT、无储能四种情况对比结果

Fig. 11 Comparison results of DDQN, DQN, MPPT, and no energy storage.

从图11和表格1的对比数据可以看出,相比于MPPT算法,强化学习算法训练得到控制策略更加节省购电费用,且在相同的训练周期下,DDQN算法比DQN算法费用更低。相比于不加储能的微网运行,DDQN算法将节省26.95%的购电费用,远大于MPPT算法的12.43%,这充分说明了DDQN算法的有效性。

## 4 总结

本文在深度强化学习算法框架下,以Simulink搭建的仿真模型作为环境向智能体提供状态信息,通过仿真模型与智能体的交互训练得到微网储能控制问题的优化策略。由于深度强化学习在处理无模型问题上的优势,本文研究不需要知道并网微网的控制目标、控制变量与状态信息之间的显示关系,而是通过训练由神经网络逼近的Q值函数来寻找控制目标与状态变量之间的作用关系,这对于处理由仿真环境提供状态数据的微网控制问题非常

有效。第3节数值验证的结果表明了本文求解思路的有效性。

由于篇幅有限,本文只讨论了一种应用实例,但是本文所提方法完全可以推广到多种微网能量控制场景中,比如在仿真模型中增删不同的分布式发电模块、改变不同的控制目标、设置不同的控制变量等等。在后续工作中,我们将从提高控制问题复杂性和提高深度强化学习的求解效率两方面入手。首先通过综合考虑多种控目标、控制变量以及约束条件使得构建的控制问题更加符合实际生产运作需求,然后通过结合压缩感知算法来进一步提高强化学习在微网控制问题中的状态采集效率,进而加速训练过程,提高算法效率。

## 参考文献

- [1] 杨新法, 苏剑, 吕志鹏, 等. 微电网技术综述[J]. 中国电机工程学报, 2014, 34(1): 57-70.  
Yang Xinfu, Su Jian, Lv Zhipeng, et al. Overview on Micro-grid Technology[J]. Proceedings of the CSEE, 2014, 34(1): 57-70(in Chinese).
- [2] 李霞林, 郭力, 王成山, 等. 直流微电网关键技术研究综述[J]. 中国电机工程学报, 2016, 36(1): 2-17.  
Li Xialin, Guo Li, Wang Chengshan, et al. Key Technologies of DC Microgrids: An Overview[J]. Proceedings of the CSEE, 2016, 36(1): 2-17(in Chinese).
- [3] Olama A, Mendes P R C, Camacho E F. Lyapunov-based hybrid model predictive control for energy management of microgrids[J]. IET Generation, Transmission & Distribution, 2018, 12(21): 5770-5780.
- [4] Shi W, Li N, Chu C C, et al. Real-time energy management in microgrids[J]. IEEE Transactions on Smart Grid, 2015, 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.  
Ma Tianxiang, Jia Boyan, Zhang zhiyuan, et al. Energy Optimal Dispatching Method of Micro-energy Internet Based on Bi-level Programming[J]. Automation of Electric Power Systems, 2019, 43(16): 34-43(in Chinese).
- [6] Tabar V S, Jirdehi M A, Hemmati R. Energy management in microgrid based on the multi objective stochastic programming incorporating portable renewable energy resource as demand response option[J]. Energy, 2017, 118: 827-83.
- [7] 张少明, 盛四清. 基于改进粒子群算法的微网优化运行[J]. 中国电力, 2020, 053(005): 24-31.  
Zhang Shaoming, Sheng Siqing. Optimal Operation of Microgrid Based on Improved Particle Swarm Optimization Algorithm[J]. Electric Power, 2020, 053(005): 24-31(in Chinese).
- [8] 禹威威, 刘世林, 陈其工, 等. 考虑电动汽车充电和需求侧响应的光伏微电网多目标优化调度[J]. 电力系统及其自动化学报, 2018, 30(01): 88-97.  
Yu Weiwei, Liu Shilin, Chen Qigong, et al. Multi-objective Optimization Scheduling for PV Microgrid Considering Electric Vehicle Charging and Demand Response[J]. Proceedings of the CSU-EPSA, 2018, 30(01): 88-97(in Chinese).
- [9] 周楠, 樊玮, 刘念, 等. 基于需求响应的光伏微网储能系统多目

- 标容量优化配置[J]. 电网技术, 2016(6): 1709-1716.  
Zhou Nan, Fan Wei, Liu Nian, et al. Battery Storage Multi-Objective Optimization for Capacity Configuration of PV-Based Microgrid Considering Demand Response[J]. Power System Technology, 2016(6): 1709-1716(in Chinese).
- [10] Nguyen, Thanh Thi, Ngoc Duy Nguyen, and Saeid Nahavandi. Deep reinforcement learning for multiagent systems: A review of challenges, solutions, and applications. IEEE transactions on cybernetics (2020).
- [11] 万里鹏, 兰旭光, 张翰博, 等. 深度强化学习理论及其应用综述[J]. 模式识别与人工智, 2019, 32(01): 67-81.  
Wan Lipeng, Lan Xunguang, Zhang Hanbo, et al. A Review of Deep Reinforcement Learning Theory and Application[J]. Pattern Recognition and Artificial Intelligence, 2019, 32(01): 67-81(in Chinese).
- [12] Luong N C, Hoang D T, Gong S, et al. Applications of deep reinforcement learning in communications and networking: A survey[J]. IEEE Communications Surveys & Tutorials, 2019, 21(4): 3133-3174.
- [13] Hua H, Qin Y, Hao C, et al. Optimal energy management strategies for energy Internet via deep reinforcement learning approach[J]. Applied Energy, 2019, 239: 598-609.
- [14] 张自东, 邱才明, 张东霞, 等. 基于深度强化学习的微电网复合储能协调控制方法[J]. 电网技术, 2019, 043(006): 1914-1921.  
Zhang Zidong, Qiu Caiming, Zhang Dongxia, et al. A Coordinated Control Method for Hybrid Energy Storage System in Microgrid Based on Deep Reinforcement Learning[J]. Power System Technology, 2019, 043(006): 1914-1921(in Chinese).
- [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. 电网技术: doi: 10.13335/j.1000-3673.pst.2020.0144.  
Liu Junfeng, Chen Jianlong, Wang Xiaosheng, et al. Energy Management and Optimization of Multi-energy Grid Based on Deep Reinforcement Learning[J]. Power System Technology: doi: 10.13335/j.1000-3673.pst.2020.0144.
- [16] François-Lavet V, Taralla D, Ernst D, et al. Deep reinforcement learning solutions for energy microgrids management[C]//European Workshop on Reinforcement Learning. 2016.
- [17] Zeng P, Li H, He H, et al. Dynamic energy management of a microgrid using approximate dynamic programming and deep recurrent neural network learning[J]. IEEE Transactions on Smart Grid, 2018, 10(4): 4435-4445.
- [18] 刘威, 张东霞, 王新迎, 等. 基于深度强化学习的电网紧急控制策略研究[J]. 中国电机工程学报, 2018, 38(001): 109-119.  
Liu Wei, Zhang Dongxia, Wang Xinying, et al. A Decision Making Strategy for Generating Unit Tripping Under Emergency Circumstances Based on Deep Reinforcement Learning[J]. Proceedings of the CSEE, 2018, 38(001): 109-119(in Chinese).
- [19] Du Y, Li F. Intelligent multi-microgrid energy management based on deep neural network and model-free reinforcement learning[J]. IEEE Transactions on Smart Grid, 2019, 11(2): 1066-1076.
- [20] Mocanu E, Mocanu D C, Nguyen P H, et al. On-line building energy optimization using deep reinforcement learning[J]. IEEE transactions on smart grid, 2018, 10(4): 3698-3708.
- [21] Jonathan LeSage (2020). Microgrid Energy Management System (EMS) using Optimization(<https://github.com/jonlesage/Microgrid-EMS-Optimization/releases/tag/v19.1.0>), GitHub.
- [22] 吕志鹏, 罗安, 荣飞, 等. 电网电压不平衡条件下微网 PQ 控制策略研究[J]. 电力电子技术, 2010(06): 75-78.  
Lv zhipeng, Luo An, Rong Fei, et al. Micro-grid PQ Control Strategy Analysis Under Utility Voltage Imbalance[J], 2010(06): 75-78(in Chinese).
- [23] Sutton R S, Barto A G. Reinforcement learning: An introduction[M]. MIT press, 2018.
- [24] 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, 042(006): 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, 042(006): 1406-1438(in Chinese).
- [25] Van Hasselt H, Guez A, Silver D. Deep reinforcement learning with double q-learning[J]. arXiv preprint: 1509.06461, 2015.
- [26] Liu X, Wang P, Loh P C. A hybrid AC/DC microgrid and its coordination control[J]. IEEE Transactions on smart grid, 2011, 2(2): 278-286.



## 附录：

表 2 仿真模型参数设置

Tab.2 Parameter setting of simulation model		
模块	参数	数值
三相交流外部电源模块	线电压有效值 (Vrms)	1.38e4
	A 相初相角 (degrees)	0.071
	频率 (Hz)	60
光伏模块三相动态负载	初始有功功率 Po(W)	-2.5e4
	初始无功功率 Qo(var)	-3.39e-13
	初始正序电压	[0.999 -0.0039]
储能模块三相动态负载	初始有功功率 Po(W)	-1e5
	初始无功功率 Qo(var)	-3.39e-13
	初始正序电压	[0.999 -0.0039]
负荷模块三相动态固定负荷	初始有功功率 Po(W)	2e5
	初始无功功率 Qo(var)	4.59e4
	有功功率 P(W)	3.50e5