文章编号：2096-5125 (2023) 03-0240-18 中图分类号：TM732；TP181 文献标志码：A
DOI：10.19705/j.cnki.issn2096-5125.2023.03.003

基于深度强化学习的微网优化运行综述

周翔[1*]，王继业[2]，陈盛[1]，王新迎[1]
（1．中国电力科学研究院有限公司，北京市 海淀区 100192；2．国网数字科技控股有限公司，北京市 西城区 100053）

Review of Microgrid Optimization Operation Based on Deep Reinforcement Learning

ZHOU Xiang[1*], WANG Jiye[2], CHEN Sheng[1], WANG Xinying[1]
(1. China Electric Power Research Institute Co., Ltd., Haidian District, Beijing 100192, China;
2. State Grid Digital Technology Holding Co., Ltd., Xicheng District, Beijing 100053, China)

Abstract: Microgrid is of great significance for the local consumption of new energy sources such as wind and solar, optimization of load levels, and improvement of energy utilization efficiency. However, the coupling of multiple energy systems, the intermittency in the output of new energy power, the randomness of load-side power demand and behavior have resulted in the microgrid becoming a dynamic and complex system, which has brought challenge to the optimization of the microgrid. Deep reinforcement learning (DRL) finds the optimal strategy through trial and error interaction with the environment, which can avoid accurate modeling of uncertainty, it is also suitable for solving sequential decision problems, therefore, it has advantages in solving dynamic microgrid optimization operation problems with a large number of uncertainties. This paper mainly reviews and analyzes the application of deep reinforcement learning in microgrid optimization operation. Finally, it discusses interpretability, reward function settings, user privacy, transferability, combining model and model-free algorithms, multi-objective function weights which faced by the application of reinforcement learning in the optimization operation of microgrid.

Keywords: microgrid; deep reinforcement learning; complex system; optimal operation

摘 要：微网在分布式新能源消纳、负荷优化、提高能源利用效率等方面具有重要作用。但新能源出力的间歇性、负荷侧用电行为的随机性导致微网成为一个动态的复杂系统，难以通过准确的物理模型刻画，给微网优化运行带来巨大挑战。深度强化学习（deep reinforcement learning，DRL）通过与环境交互试错寻找最优策略，不依赖于新能源出力和负荷的精确建模，适用于解决序贯决策问题，在求解含有大量不确定性的微网优化运行难题时具有优势。为此，从DRL原理、DRL在单个微网以及微网群优化运行中的应用进行了综述与分析，最后对应用中所面临的算法可解释性、奖励函数设置、用户隐私性等方面进行了展望。

关键词：微网；深度强化学习；复杂系统；优化运行

0 引言

构建新型电力系统，大力发展新能源，提高能源利用效率是必然趋势[1-3]。在电源侧，将分布式新能源就地消纳，可以减少输电损耗和对电网的冲击，提高供电可靠性[4]。在负荷侧，用户用电需求的多样性和随机性逐渐提升，通过协调利用多种能源，可以在满足用户负荷需求的同时提高能源利用效率。微网作为一种管理分布式电源[5]，协调多种能源运行的小型配电系统[6]，通常可分为单个微网或多个微网组成的微网群（如图1和图2所示），通过加强微网的能量优化运行管理，能够显著提升电能以及气、热等多种能源利用效率，减少对环境的污染，降低碳排放。

尽管微网有以上诸多优势，但是也面临一系列挑战。首先，大规模分布式能源设备的接入、多种能源系统的相互耦合增加了微网系统运行的复杂性，通过精确的物理模型刻画微网系统的特性变得愈发困难。其次，新能源出力的间歇性、用户用能行为的随机性导致微网中包含大量不确定性。最后，电动汽车、蓄电池等柔性负荷和电力电子设备增加了微网的动态特性，多种能源负荷以及大量不同类型能源设备的相互配合使得微网优化成为一个时变的多变量问题，因此

基金项目：国家电网有限公司总部科技项目（基于边云协同的微网群优化运行智能技术研究及应用，5700-202130263A-0-0-00）。
Science and Technology Foundation of SGCC(5700-202130263A-0-0-00).

需要实时在线的能量优化管理策略[7-8]。

现有对于调度与控制的研究方法多为传统优化算法，然而线性规划、非线性规划、混合整数线性规划、随机优化与鲁棒优化等算法依赖于精确的数学模型[9-12]，通过特定的概率分布来刻画不确定量偏差较大，且在鲁棒优化中也会存在一定程度上结果保守的问题。同时微网优化调度与控制是一个高维变量问题，在用传统方法求解时，复杂度非常大，求解过程非常耗时，无法满足实时在线的策略生成需求。

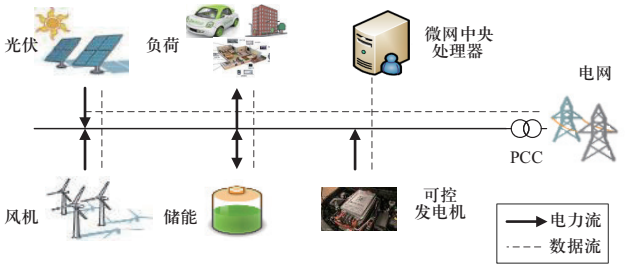

图 1　单个微网示意图

Fig. 1　Schematic diagram of single microgrid

近年来，强化学习（reinforcement learning，RL）成为人工智能领域的一个研究热点，其具有在动态不确定环境中通过交互试错进行自我学习的优势，被广泛应用于游戏博弈、生产调度、工业制造等诸多领域[13-17]。通过智能体与环境进行自主的交互试错，不断改进优化策略，最大化智能体从环境中获得的累积奖励值[18-19]。在强化学习的发展历史中，已经有较多学者进行了强化学习与神经网络结合的研究，但是在复杂序贯决策问题中并没有太大的进展。随着深度学习在大规模复杂数据驱动任务中表现出卓越性能，产生了深度学习和强化学习相结合的深度强化学习（deep reinforcement learning，DRL）。深度强化学习将感知、学习和决策整合到同一个框架内，能够实现“端到端”的感知和决策。在深度强化学习中，智能体通过交互经验来学习和改进策略，可以学习到在不同系统状态下的动作策略，不需要通过精确的数学公式刻画物理模型信息，因此基于深度强化学习的方法可以捕捉到难以精确建模系统的动态特性，在一些高维复杂任务中的表现优于基于模型的算法。

目前深度强化学习算法在包括微网优化运行的电力系统运行和控制中的应用也引起广大学者的关注。深度强化学习算法在解决微网优化运行问题时主要具有以下优势。

1）通过与微网系统的源、荷、储等设备环境状态信息进行交互试错，不断改善运行优化策略，不需要构建微网的复杂显式物理模型。

2）通过深度强化学习算法中的深度神经网络可以拟合系统不同状态下的优化动作策略，自适应风/光新能源出力和负荷等不确定性。

3）通过离线历史数据进行模型训练，能够应用于在线的策略生成，极大缩短策略生成时间，将已有模型应用在其他类似场景时，可以进行模型自主进化。

本文在相关研究成果与综述的基础上，对深度强

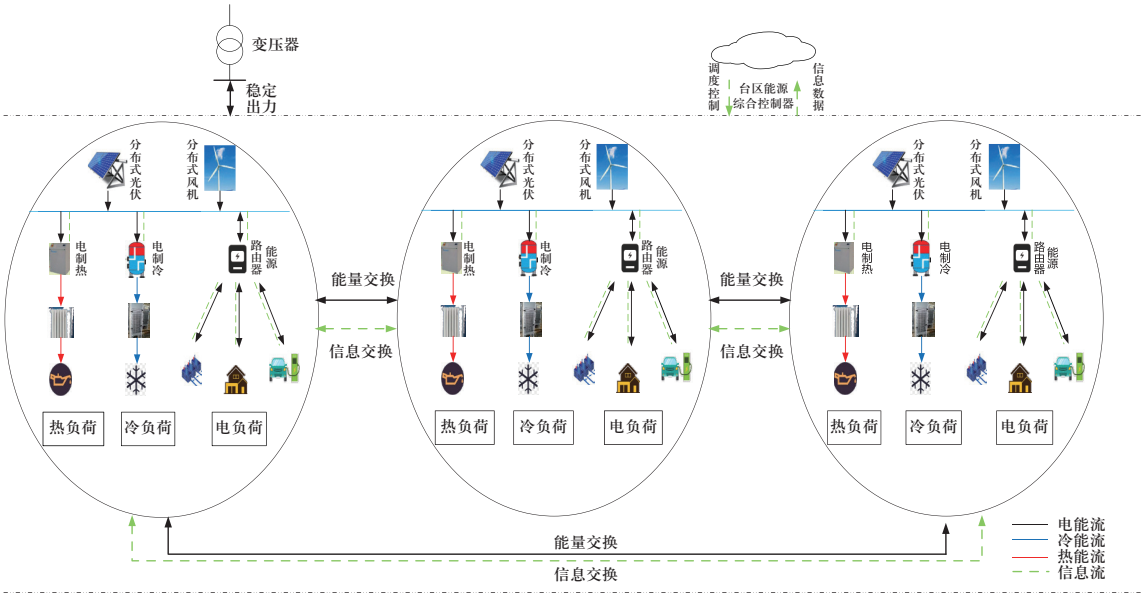

图 2　微网群示意图

Fig. 2　Schematic diagram of microgrid group

化学习在微网优化运行中的应用进行了回顾与总结，第1章主要介绍强化学习的定义、分类和基本原理；第2章总结深度强化学习在微网优化运行中的应用，将微网优化调度分为单个微网以及微网群的调度问题，重点讨论以电能为单一能源的微网和电/气/热/冷多种综合能源的微网，以及针对分布式电源、储能充放电、电动汽车充放电、家庭能量管理、楼宇能量管理等具体场景的应用，还包括在微网优化控制中的应用，主要针对电压和频率的稳定进行讨论，算法方面对比分析各种问题中所运用的深度强化学习算法；第3章对深度强化学习在未来微网应用中可能会面对的问题和挑战进行总结和展望；第4章对本文工作进行总结。

1 深度强化学习基础

强化学习是一种从环境状态映射到动作的学习方式，其学习过程可描述为：①在每一个时刻通过智能体与环境交互得到全部状态或者部分可观测状态；②针对观察状态，基于价值函数对各个动作进行评价，采用某种策略将当前状态映射为相应动作；③环境对动作作出回应，得到下一时刻的状态[20]。强化学习过程如图3所示。

根据是否依赖模型，强化学习分为基于模型的强化学习算法和无模型的强化学习算法，前者基于模型进行序贯决策，后者利用与环境交互数据改善策略优劣。根据奖励函数是否已知，强化学习分为正向强化学习与逆向强化学习，前者奖励通过人为设定，后者奖励通过机器学习算法所得。根据学习方式的不同，强化学习可以分为同策略和异策略，前者是指生成样本的策略与网络更新参数时使用的策略相同，即与环境互动和网络参数更新同时进行，后者是指生成样本的策略与网络更新参数时使用的策略不同，采用先采样后集中更新的方式进行学习[21]。根据策略更新方式不同，强化学习可分为基于值函数的方法和基于策略梯度的方法。强化学习算法一般分类方式如图4所示。本文所综述内容在算法方面多为无模型算法，因此以下关于强化学习算法介绍主要是基于无模型算法。

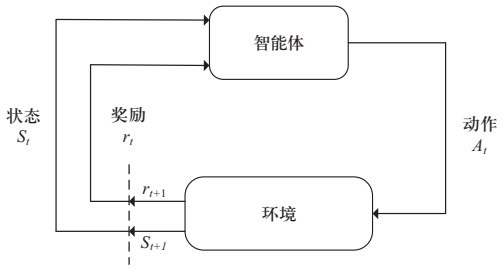

图 3 强化学习过程示意图

Fig. 3 Schematic diagram of reinforcement learning process

1.1 马尔可夫决策过程

强化学习中智能体与环境交互的过程可以通过马尔可夫决策过程（Markov decision process，MDP）

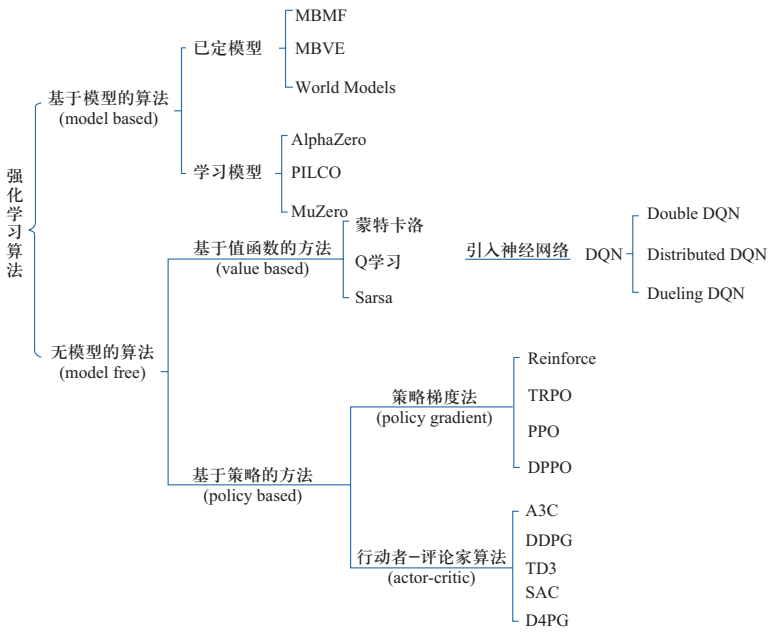

图 4 强化学习算法分类

Fig. 4 The classification of reinforcement learning algorithms

描述。一般将MDP定义为一个五元组（S、A、P、r、γ），其中S是状态空间，是智能体所能感知到外部环境状态的集合；A是动作空间，是智能体通过策略作出的所有动作集合，可以是离散或者连续的；P是状态转移概率，表示智能体在当前状态s_t执行动作a_t转移到下一状态s_{t+1}的概率；r是奖励函数，表示智能体在当前状态s_t执行动作a_t获得的即时奖励；γ是折扣因子，表示对将来回报的衰减系数[22]。当该转移过程结束时，奖励之和定义为

$$G_t = R_t + \gamma R_{t+1} + \gamma^2 R_{t+2} + \cdots = \sum_{k=0} \gamma^k R_{t+k+1} \tag{1}$$

式中：G_t表示在训练轨迹中从时刻t开始的回报奖励值；γ是折扣因子；R_t是即时奖励值。

强化学习的目标是最大化获得的奖励，主要挑战之一就是利用和探索之间的平衡，利用是作出当前信息下的最佳决定，探索则是尝试不同的行为继而收集更多的信息。

1.2　强化学习

1.2.1　基于值函数算法

值函数包括状态值函数和状态–动作值函数。状态值函数是指从当前状态到终止状态能够获得的累积回报的期望[23]，表示为

$$V_\pi(s) = E\left[G_t \middle| s_t = s\right] \tag{2}$$

式中：$V_\pi(s)$表示从当前状态s开始按照策略π所能获得累积回报奖励值。

状态–动作值函数是指在某一状态下，采取某一行动后，未来能够获得的累积回报期望值，表示为

$$Q_\pi(s,a) = E\left[G_t \middle| s_t = s, a_t = a\right] \tag{3}$$

式中：$Q_\pi(s, a)$表示从当前状态s开始，采取动作a，按照策略π所能获得累积回报奖励值。

值函数算法通过迭代更新值函数来间接得到智能体的策略，当值函数迭代达到最优时，智能体的最优策略π^*通过最优值函数得到：

$$\pi^* = \arg\max_\pi V_\pi(s) \tag{4}$$

为正确估计值函数，通常采取以下方式。

1）蒙特卡洛法（MC）。

蒙特卡洛法通过生成大量的训练轨迹，记录在每个状态或每个状态–动作对的平均奖励值来拟合值函数[24]，其计算方法如下：

$$V_\pi^{\mathrm{MC}}(s) = \lim_{i\to+\infty} E\left[G^i(s) \middle| s, \pi\right] \tag{5}$$

式中：$G^i(s)$表示第i个训练轨迹中，在状态s下的回报奖励值。

2）时间差分法（TD）。

时间差分法和蒙特卡洛法都是基于无模型的算法，不同的是时间差分法不需要从完整的训练轨迹中去学习，对于某个给定的策略，在每一个时间步上在线进行学习和更新，用当前时刻得到的奖励回报值更新上一个时刻值[25]，其计算方法如下：

$$v(S_t) \leftarrow v(S_t) + \alpha\left(R_{t+1} + \gamma v(S_{t+1}) - v(S_t)\right) \tag{6}$$

式中：α表示更新速率，满足$0<\alpha<1$。TD的这种方法也称为自举法（bootstrapping）。SARSA算法与Q-learning算法都是采用TD方法进行状态–动作值函数计算的。SARSA算法每次更新值函数需要知道当前的状态（state）、当前的动作（action）、奖励（reward）、下一步的状态（state）、下一步的动作（action），因此取名为SARSA算法。该算法为一种同策略（on-policy）算法，即实际执行策略与评价策略是同一个策略，其计算方法如下：

$$\begin{aligned} &Q(S_t, A_t) \leftarrow Q(S_t, A_t) + \\ &\alpha\left[R_{t+1} + \gamma Q(S_{t+1}, A_{t+1}) - Q(S_t, A_t)\right] \end{aligned} \tag{7}$$

与SARSA算法不同，Q-learning算法是一种异策略（off-policy）算法，即实际执行策略与评价策略不同，其计算方法如下：

$$\begin{aligned} &Q(S_t, A_t) \leftarrow Q(S_t, A_t) + \\ &\alpha\left[R_{t+1} + \gamma \max_a Q(S_{t+1}, a) - Q(S_t, A_t)\right] \end{aligned} \tag{8}$$

由于Q-learning算法是异策略算法，可以通过大胆探索得到经验轨迹来优化策略，从而有可能得到最优策略。

1.2.2　基于策略梯度算法

基于策略梯度算法主要应用于连续状态动作空间的强化学习问题，不依赖值函数，直接将策略Π参数化为$\Pi(s|\theta)$，然后计算出关于策略性能指标的梯度，沿着梯度方向，如式（9）所示，不断调整动作策略，使强化学习的目标累积回报期望值最大[26]。常见基于策略梯度算法有REINFORCE算法和行动者–评论家（actor-critic，AC）算法。

$$\theta \leftarrow \theta + \alpha\gamma^t g \nabla \log \pi_\theta(a|s) \tag{9}$$

式中：θ表示策略网络的参数；$\gamma^t g \nabla \log \pi_\theta(a|s)$表示策略朝着变化最快的方向进行更新。

策略梯度算法又分为随机性策略梯度和确定性策略梯度[27]。随机性策略梯度是在当前状态s下，经过参数化θ策略得到的动作a符合某个概率分布，因此在此状态下有不同概率对应于不同的动作选择，随机性策略梯度表达式如式（10）所示。而确定性策略梯度

是在当前状态s下，对于策略μ_θ来说，对应于唯一的动作，确定性策略梯度表达式如式（11）所示[28]。

$$\nabla_\theta J(\theta)=E_{s,a\sim\pi}\left[\nabla_\theta \ln\pi(s|\ \theta)Q_\pi(s,a)\right] \tag{10}$$

$$\nabla_\theta J(\theta)=E\left[\nabla_\theta \mu(s|\ \theta)\nabla_a Q_\mu(s,a)\big|_{a=\mu(s|\theta)}\right] \tag{11}$$

式中：$J(\theta)$表示目标函数。

总结以上内容，强化学习算法特性描述如表1所示。

表 1 强化学习算法特性

Table 1 Characteristics of reinforcement learning algorithms

种类	优点	缺点
基于模型算法	算法效率较高	需要对真实世界进行建模，泛化性较差
无模型算法	通用性和泛化性强	需要与外部环境进行不断交互
值函数算法	不需要显式制定策略；适用于离散空间	求解连续空间问题时，解值函数困难
策略梯度算法	策略参数化简单，易收敛；适用于连续空间	易收敛到局部最小值；评估单个策略不充分，方差较大
同策略	行为策略与目标策略一致，处理较简单	无法较好地平衡探索与利用关系
异策略	保持探索的同时，能更好地求到最优值	行为策略与目标策略不一致，优化复杂

1.3 深度强化学习

传统的强化学习算法会通过表格的形式存储状态值函数$V_\pi(s)$和状态–动作值函数$Q_\pi(a, s)$，其中每一个状态–动作对都有一个与之相关的离散估计值。当采用离散的方式表示估计值时，容易引起“维度诅咒”问题。因为在实际应用场景中，强化学习任务所面临的状态空间往往是连续的，存在无穷多个状态，在这种情况下，就不能再使用表格对值函数进行存储。值函数近似利用函数直接拟合状态值函数或状态–动作值函数，减少了对存储空间的要求，有效地解决了这个问题[29]。通过深度学习（deep learning，DL）实现函数逼近是一个切实有效的方式。深度学习起源于人工神经网络（artificial neural network，ANN），其核心思想是以数据驱动的方式，通过构建的多层网络结构，采用非线性变换实现对输入信息的分级表达，可以从原始数据中理解和学习到复杂的表示内容，拟合数据之间的关系，进而实现特征提取、分类和预测等功能。常见的深度学习结构包括深度神经网络（deep neural network，DNN）、卷积神经网络（convolutional neural network，CNN）、循环神经网络（recurrent neural network，RNN）、深度残差网络（deep residual network，DRN）等[30]。

近年来，深度学习与强化学习广泛结合形成了一个研究热点领域——深度强化学习。由于深度学习强大的感知能力，适用于处理高维、复杂的环境特征，结合强化学习适用于进行策略学习的能力，深度强化学习同时具备复杂环境感知与决策能力。深度强化学习算法在复杂任务（如Atari游戏和棋盘游戏Go）上取得了人类水平的性能（或更高）。表2列出了常见的深度强化学习算法的特性描述。

表 2 常见深度强化学习算法特性描述

Table 2 Description of common deep reinforcement learning algorithms

DRL算法	分类	算法描述	应用场景
DQN	无模型，基于值函数，异策略，单智能体	应用于离散的动作空间	机器人控制、路径规划、车间生产优化调度等
DDPG	无模型，基于策略梯度，异策略，单智能体	应用于连续的动作空间	无人驾驶决策、能源动态、经济调度等
TRPO	无模型，基于策略梯度，同策略，单智能体	保证策略向优化方向更新，计算开销大	网络资源分配、机械臂控制、自动驾驶等
PPO	无模型，基于策略梯度，同策略，单智能体	较TRPO更易实现，所需调节参数少	交通流优化控制、无人机避障、三维路径规划等
A3C	无模型，基于策略梯度，同策略，单智能体	多线程提高学习效率，降低样本相关性	搜索导航、服务集成、电梯调度等
SAC	无模型，基于策略梯度，异策略，单智能体	提高算法鲁棒性，加快训练速度	运动控制、电动汽车优化调度、安全自动驾驶等
MADDPG	无模型，基于策略梯度，异策略，多智能体	集中式训练，分布式执行	多无人机协同、边缘网络任务卸载等

1.3.1 深度Q网络

第一个深度强化学习算法——深度Q网络（DQN）算法，为了在连续的状态空间中计算值函数$Q^{\pi}(s,a)$，可以用一个函数$Q_{\phi}(s,a)$来表示近似计算，称为价值函数近似，如式（12）所示，DQN采用梯度下降的方法最小化代价函数式（13）。

$$Q_{\phi}(s,a)\approx Q^{\pi}(s,a) \tag{12}$$

$$L(w)=E\left[\left(r+\gamma\max_{a'}Q\left(s',a'\middle|w\right)-Q(s,a|w)\right)^2\right] \tag{13}$$

式中：a和s是当前时刻动作和状态；a'和s'是下一时刻动作和状态；r是即时奖励。

DQN算法为了训练的稳定性，采用了以下技术[31]。

1）目标网络。

Q网络每次迭代优化的目标值由另外一个单独的目标网络产生，目标网络与Q网络结构相同，且目标网络的参数每经过N次迭代用Q网络的参数进行更新，其余时刻保持不变。通过设置目标网络，DQN训练过程变得稳定。

2）经验回放。

将智能体在每个时间点上与环境交互形成的经验$e_t=(s_t,a_t,r_t,s_{t+1})$存入经验池，形成回放记忆序列$D=(e_1,e_2,e_3,\cdots)$。训练时，每次从$D$中随机抽取批量样本，并基于梯度规则更新网络参数。经验回放机制增加了历史数据的使用效率，减少了数据之间的相关性，有助于解决动作值函数不稳定甚至不收敛的问题。

在DQN算法问世以后，众多学者在DQN算法基础上对神经网络等部分进行了改进，衍生出了诸如Double DQN算法、Dueling DQN算法、深度递归Q网络算法等，提高了DQN算法的性能和应用范围。

1.3.2 深度确定性策略梯度

深度确定性策略梯度（deep deterministic policy gradient，DDPG）算法将强化学习算法中的策略梯度与值函数相结合[32]，通过深度神经网络去拟合状态与动作空间，在策略搜索中将值函数大小作为动作优劣的评判标准，能够有效减少方差，避免陷入局部最优解，适用于解决连续状态与动作空间的问题。

在DDPG算法中，分别使用θ^{π}和θ^{Q}的深度神经网络拟合确定性策略和动作状态值函数，并设置对应的结构相同的目标神经网络提升算法的稳定性和收敛性。为提高学习过程的稳定性，目标网络参数采用软更新方式：

$$\theta^{Q'}\leftarrow\tau\theta^{Q}+(1-\tau)\theta^{Q'} \tag{14}$$

$$\theta^{\pi'}\leftarrow\tau\theta^{\pi}+(1-\tau)\theta^{\pi'} \tag{15}$$

式中：τ表示更新率，且值远小于1。DDPG算法因其简洁性而容易应用到复杂问题中，但是其局限性在于收敛时间很长，在后续对基于策略梯度与值函数相结合的算法研究中，陆续诞生了TD3、SAC等算法，提高了算法稳定性和收敛性。

1.3.3 多智能体深度强化学习

在一些简单场景如轨迹追踪或控制任务的情况下，智能体不需要互相交互，但是在复杂的任务或场景下，其他智能体的状态或动作会影响当前智能体的动作选择，这就引出了多智能体系统。当多智能体系统与强化学习结合时，就产生了多智能体深度强化学习领域。

多智能体强化学习（multi-agent reinforcement learning，MARL）是指让多个智能体处于相同的环境中，每个智能体独立与环境交互，利用环境反馈的奖励改进自己的策略，以获得更高的回报。在多智能体系统中，一个智能体的策略不能简单依赖于自身的观测、动作，还需要考虑到其他智能体的观测、动作[33]。多智能体系统有4种常见设定：完全合作关系，即智能体的利益一致，获得的奖励相同，有共同的目标；完全竞争关系，即一方的收益是另一方的损失；合作竞争的混合，即智能体分成多个群组，组内的智能体是合作关系，它们的奖励相同，组间是竞争关系，两组的奖励是负相关的；利己主义，即系统内有多个智能体，一个智能体的动作会改变环境状态，从而让别的智能体受益或者受损[34]。

MARL算法的学习方式包括独立式学习和集中式学习2种。独立式学习方法对每个智能体分别使用强化学习算法，而将其他智能体看作环境的一部分，在学习过程中，每个智能体获得其局部观测，并且向着最大化整体奖励值的方向调整每个智能体的策略；集中式学习方法将所有智能体的状态和动作集合在一起，构成一个总的状态和动作空间，并直接使用单智能体的强化学习算法。“集中式训练–分布式执行”是当前常用的一种多智能体强化学习算法结构。所有的智能体采用集中式结构进行训练，每个智能体可以获得其他智能体的状态动作信息。在训练结束之后则执行分布式策略，每个智能体只能通过自身的观测和其他智能体的信息进行动作的选择，典型算法包括多智能体深度确定性策略梯度（multi-agent deep deterministic policy gradient，MADDPG）算法。

2 基于深度强化学习的微网优化运行

随着分布式新能源的不断接入，电动汽车等新兴

负荷大量涌现，新能源发电的不确定性、用户用能的多样性、用电行为的随机性、各种用电装置之间的互动配合、电力电子设备的非线性都给微网优化调度和控制带来了难题，因此协调微网内分布式发电机出力、多种能量转化装置之间互相配合、源荷储协同、控制电压与频率稳定具有重要研究意义。本文从单个微网中的分布式电源侧优化、储能侧优化、用户需求侧优化、电动汽车充放电优化，微网群能量优化调度，微网的优化控制等方面介绍基于强化学习算法的微网优化运行。微网系统包含元素众多，优化控制目标逐渐变成非凸非线性问题，同时由于存在许多不确定性和随机性的系统参数，通过精确的数学方法显式构建系统模型存在困难，因此通过无模型的强化学习算法，以数据驱动的方式对微网优化运行问题进行研究成为热点。

2.1 单个微网能量优化调度

微网优化调度问题是一个具有多目标、多维变量、多约束条件的非凸非线性问题，优化目标包括经济性、可靠性以及环保性，约束条件包括电力系统约束、天然气系统约束以及热能系统约束等。

1）经济性。

经济性主要是指微网运行成本，源侧成本包括购买天然气等能源成本、发电机发电成本、启停成本、运行维护成本、通过大电网购售电成本等；储能侧成本包括储电、储热、储气等储能装置运行损耗成本；负荷侧成本包括负荷削减、平移等参与需求响应成本。

2）可靠性。

由于分布式新能源出力与用户用电负荷的随机性，有功和无功变化较大会引起电压波动，同时微网会将多余电能上送到配网中，也可能会对配网产生冲击。因此负荷峰谷差、负荷尖峰以及联络线功率等指标影响着微网的可靠性。

3）环保性。

传统发电机发电会排放CO_2等温室气体，通过微网优化调度，可提高新能源利用率，减少传统发电机出力，降低碳排放。

传统算法对微网能量管理经济性、可靠性和环保性的研究已有大量成果，主要利用随机优化和鲁棒优化方法构建模型，然后运用混合整数线性规划、遗传算法、粒子群算法等进行模型求解。

文献[35]针对风电的随机性与间歇性，通过基于拉丁超立方采样的场景法处理风光出力的不确定性，建立了基于实时电价的虚拟电厂随机优化模型。文献[36]考虑到新能源出力与新兴负荷的随机性，提出了一个两阶段鲁棒随机规划模型应用于微网经济调度。随机优化利用随机变量概率分布刻画不确定性会存在一定误差，同时利用场景生成描述研究对象的不确定性会随着问题规模的变大而导致求解困难。鲁棒优化主要针对最恶劣场景下的优化问题，所得结果可能过于保守。

在模型求解方面，主要是通过解析法和人工智能方法[37]。解析法主要是通过线性化或者凸松弛技术将原问题转化为线性问题或者混合整数规划问题。文献[38]对电、热、气进行系统性统一建模，运用混合整数线性规划求解综合能源优化调度问题，对非线性关系进行分段线性化处理。此类方法适用于精度要求不高但是需要快速计算的场景。文献[39]综合考虑微网的成本、排污及风险，尤其是分布式电源出力的波动性影响，通过小生境多目标粒子群优化算法进行求解。启发式算法在求解高维问题时效率较低，求解不稳定，实时性方面有待提升。

在考虑通过深度强化学习优化以上目标时，主要从微网源侧、储能侧以及负荷侧三个方面进行分析。微网源侧主要是对分布式发电机出力进行优化，支撑新能源出力的不确定性；储能侧主要是对通过优化储电、储氢、储热等装置动作，平滑负荷曲线，消纳新能源；负荷侧主要是通过控制柔性负荷平移、截断或者电价制定的方式，保证负荷曲线平滑。

2.1.1 分布式机组出力优化

面对分布式新能源的大规模普及，通过整合微网中的分布式发电资源，可以有效应对新能源的随机性与波动性，提高电力系统的稳定性。文献[40]通过利用深度强化学习算法（A3C）实现分布式发电机组的优化调度，降低发电成本的同时，提高新能源消纳率，而且设计了基于边缘计算体系结构，能够更好满足优化调度所需的实时通信要求。文献[41]在考虑微网源侧光伏不确定性的条件下，通过贝叶斯网络对光伏出力进行建模，构建双深度期望Q网络，有效提升了模型训练速度和收敛性。文献[42]在含有风光等多个随机性因素的微网中，利用深度确定性策略梯度算法优化发电机出力和备用，避免对风光不确定性进行建模。

面对此类问题，在构建深度强化学习要素时，对于状态空间的设置通常包括分布式新能源风光的出

力、负荷功率、上一时段可控发电机组出力情况；动作空间的设置包括燃气轮机等机组的出力、机组备用预留；奖励函数的设置包括机组出力成本、系统备用成本、微网其他运行成本、机组爬坡约束、电力系统约束等。在实际训练过程中为了加速训练学习过程，需要将状态、动作进行归一化。

利用深度强化学习与环境交互可以学习到环境信息，不依赖精确模型的特性，因而适用于应对新能源出力的不确定性，可以通过训练完成的模型有效应对不同风光出力下可控发电机出力和备用的优化调度。但以上研究均是可以应对不同新能源场景下的日前调度，如果第二天实际场景发生变化，日前调度结果能否以一定概率满足需求需要进行进一步的研究。

2.1.2 储能装置充放电优化

由于微网中源侧与负荷侧波动较大，通过增加储能装置能够更好地进行电力电量的平衡，提高新能源的利用效率。当新能源出力较多时，通过储能装置可以将多余能量存储，当新能源出力不足时，储能装置将所储能量释放。文献[43]在包含光伏、储氢以及蓄电池储能的微网场景下，结合长短期记忆网络和卷积神经网络，利用深度Q学习方法离散化储能系统动作，能够有效处理微网中的源荷侧随机性，对于不同时刻、天气、季节的场景也均适用。文献[44]在微网并网模式下，以最小化从外部电网购电为目标，通过双重深度Q学习算法，以储能实时充放电为动作，训练得到储能优化策略。

面对此类问题，在构建深度强化学习要素时，对于状态空间的设置通常包括风光新能源出力、负荷功率、蓄电池等储能装置的状态；动作空间设置包括储能装置的动作，主要分为充电、放电以及无操作3种状态；奖励函数设置包括微网运行成本、新能源消纳率、储能荷电状态约束以及充放电功率约束等。

目前对储能优化的研究中，仅考虑单一储能装置的情况较多，虽然有涉及两种储能装置的优化，但是微网场景构建较简单，没有考虑氢气和电能传输时间尺度特性，且并没有涉及在需求响应参与情况下储能的优化策略。

2.1.3 用户负荷侧需求管理

负荷侧用户选择能源形式的多样性以及用能行为的随机性都可能导致负荷尖峰化效应和负荷峰谷差不断拉大的问题，通过直接控制柔性负荷的状态以及制定实时电价的方式可以有效调节负荷水平，保证用电可靠性和用户满意度。

文献[45]利用双深度Q网络控制可中断负荷的状态，能够实现需求侧的快速响应，降低负荷峰谷差和运行成本。文献[46]针对分布式电采暖的可时移特性，利用深度学习提取用户采暖单元温变–功率关系特征，通过改进的深度确定性策略梯度算法优化实时电价下的电采暖负荷，提高了需求侧响应能力，降低了用电成本。文献[47]在引入时间–价格弹性系数基础上，完善用户需求响应模型，通过DQN进行模型求解，实现需求响应最优策略。文献[48]通过聚类方法构建电加热器的需求响应模型，通过深度Q网络为负荷聚合生成动态补贴价格。

面对此类问题，在构建深度强化学习要素时，对于状态空间的设置通常包括微网系统运行状态（系统节点电压和功率）、柔性负荷状态（中断或转移的负荷量以及补偿价格）；动作空间设置包括柔性负荷控制（1表示可以进行负荷柔性控制，0表示不进行柔性控制）、补贴价格；奖励函数设置包括柔性负荷控制动作次数、微网运行成本、负荷节点电压以及负荷峰谷差水平、用户不满意度等。

当前随着电力市场发展不断深入推进，大量负荷侧资源存在巨大调控潜力，通过价格型和激励型两种方式引导用户负荷行为，构建需求响应模型，并利用深度强化学习不依赖环境的特性，能够兼具实时性和灵活性，有效提升用户和售电商的综合效益。

2.1.4 电动汽车充放电管理

随着大规模电动汽车群的接入，其大功率与时空不确定性会改变电网负荷峰谷差，形成负荷尖峰效应，对电网安全稳定产生冲击。但是电动汽车群也具备成为重要灵活可调度资源的潜力，通过有序充电，能够有效平抑新能源与负荷波动。由于充放电存在随机性，电动汽车充放电的优化管理具有挑战性。文献[49]通过双延迟深度确定性策略梯度算法对单个电动汽车充放电进行建模，再将训练得到模型进行分布式部署，实现对规模化的电动汽车群进行实时充放电优化。文献[50]在考虑用户参与电动汽车负荷优化的意愿以及用户可调性识别条件下，从电动汽车集群角度出发，利用soft actor-critic（SAC）算法对电动汽车充换电进行实时优化调度。文献[51]将电动汽车充放电调度问题建模为一个约束型马尔可夫决策过程，并提出一种基于安全深度强化学习的无模型方法，该方法不需要任何关于随机性的先验知识，且不需要手动设计惩罚项或调整惩罚系数。

此类问题中，在构建深度强化学习要素时，对于

状态空间的设置通常包括电动汽车基础负荷状态、分时电价、充电时段及离开时段等；动作空间的设置包括电动汽车充放电功率；奖励函数设置包括电动汽车聚合负荷的平抑程度、充电成本、换电站行为约束和电池充放电约束。

2.1.5 典型微网形式的协同优化

2.1.1—2.1.4主要聚焦于微网中的源侧、储能侧以及荷侧等各环节的优化问题，并未对微网进行整体协同优化调度的研究。微网作为小型配电系统，具有多种组织形式，本节选取较为典型的3种微网形式进行系统优化调度研究。

1）家庭能量系统优化。

随着分布式光伏系统、家用储能装置以及电动汽车越来越多进入到家庭中，电、热、冷、气等多种能源需要协调，同时需要考虑空调、热水器、厨具等家庭电器动态投退，这些不确定性和复杂性需要家庭能量管理系统（home energy management system，HEMS）进行管理，HEMS结构如图5所示。

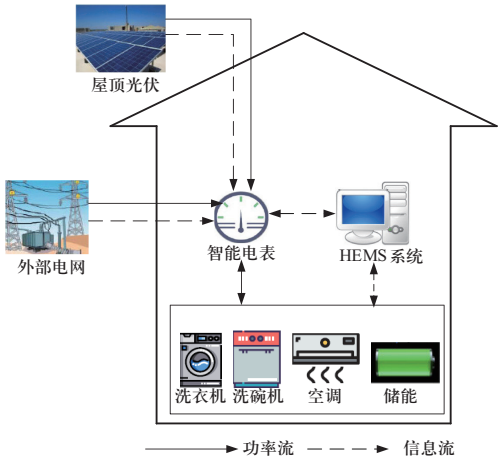

图 5 家庭能量管理系统结构

Fig. 5 The structure of home energy management system

为了应对用户端日益增加的复杂性和不确定性，文献[52]提出了基于双深度Q学习的家庭能量优化管理策略，降低了家庭用能成本。文献[53]以热水器、空调等设备作为优化对象，建立家庭综合能源需求响应的强化学习模型。文献[54]首先构建了基于状态概率分布的居民用电调度优化模型，然后通过异步强化学习算法进行家庭能量管理系统的在线优化，实现居民用电经济性目标。文献[55]在分析负荷特性并分类的基础上，构建双层HEMS框架，然后将Rainbow算法应用于HEMS的优化调度。

面对此类问题，在构建深度强化学习要素时，对于状态空间的设置通常包括电价、天然气价格、家庭用具的投退状态、热水需求量、室内温度、光伏出力等；动作空间的设置包括热水器、空调、洗衣机、厨具等设备的投退；奖励函数的设置包括用户用能成本、用户舒适度和系统约束越限惩罚等。

家庭能量系统的优化是一个具有高维决策变量，充满随机性和不确定性的优化调度问题，通过深度强化学习与各种环境交互学习最优策略，能够有效解决传统方法耗时且容易陷入局部最优的困境，实现家庭用能的实时在线调度，合理规划用能行为，降低家庭能源使用成本。

2）智能楼宇能量系统优化。

2019年，全球建筑物耗能占比达30%，产生了28%的碳排放量，未来建筑物的能源需求将进一步增加，其中大部分能耗用于加热、通风和空调等装置，因此智能楼宇的概念越来越受到关注，通过机器学习、大数据等先进技术，开发楼宇能量管理系统，能够降低能源消耗、减少碳排放、提高用户舒适度。

文献[56]通过基于DQN算法的商业楼宇优化调度策略，降低了能源消耗成本，同时保证室内温度满足用户舒适要求。文献[57]通过DQN和DDPG算法在线优化住宅楼宇的能源成本和负荷峰谷差。文献[58]在考虑楼宇用户热舒适度的条件下，首先设计了一个用户热舒适度预测神经网络，然后提出了基于DDPG的热舒适度优化控制策略，在提高用户舒适度的同时，暖通空调等能耗降低了4.13%。当楼宇中的空调、采暖、通风、冷水机组等设备较多时，通过采用多智能体深度强化学习算法能取得更好的优化效果。文献[59]采用多智能体深度强化学习算法，在不牺牲用户舒适度的前提下，调整空调、采暖等装置的设定值，每一个基于DDQN算法的智能体与其他智能体相互协调，以学习到最优控制策略。文献[60]提出基于深度强化学习的框架，用于控制建筑楼宇中的暖通空调系统、照明系统、遮阳系统和窗户系统，在保证所有子系统所消耗能量最低的同时，保持用户舒适度。

面对此类问题，在构建深度强化学习要素时，对于状态空间的设置通常包括楼宇内部和外部温度、湿度；动作空间的设置包括温度湿度的设定、进风/出风量等；奖励函数的设置包括楼宇用能成本最低、负荷峰谷差降低、电力设备约束等。

智能楼宇能量优化调度管理的研究对于推进节能环保、降低碳排放具有重要意义和必要性。但是也存

在诸多挑战，主要是构建精确高效的建筑物热动态模型十分困难；不确定性因素较多，例如分布式新能源发电、电价、温度、居住人数等；供暖系统、照明系统、遮阳系统等多个子系统相互耦合，决策变量众多，传统方法难以有效实时在线优化。而深度强化学习在解决以上问题时具有优势。

3）园区综合能源系统优化。

综合能源系统（integrated energy system，IES）是一种含有电、热、气等多种能源耦合互补的能源系统，通过不同能源之间整合转化的优化调度，能够在满足用户用能需求的同时，提高能源利用效率，促进风、光等新能源的就地消纳。

文献[61]提出双层强化学习模型，对含有电、热、气的园区IES进行优化调度，该模型上层将储能视为强化学习智能体，下层在上层给出动作后通过整数线性规划进行即时优化，最终实现IES的实时调度。文献[62]在将IES动态调度问题建模为马尔可夫决策过程后，通过改进的深度确定性策略梯度算法进行模型求解，提高了模型训练的质量和速度，同时降低了IES运行成本。文献[63]首先对IES中源荷储以及电、气、热等子系统进行多智能体划分，针对不同智能体诉求，分别构建目标函数，提出基于多智能体深度强化学习IES调度方法。文献[64]提出通过柔性行动器–评判器的深度强化学习算法，能够实现多种能源系统和装置的连续动作控制，对多种新能源出力场景下的电–气综合能源系统进行优化决策。

面对此类问题，在构建深度强化学习要素时，对于状态空间的设置通常包括电价、天然气价格、电/热/冷负荷、新能源出力及储能状态；动作空间的设置包括储能装置动作、常规机组出力、天然气供气量、电转气单元等；奖励函数设置包括电力系统运行成本、天然气成本、储能和能源转换设备维护成本、约束越限惩罚等。

面对IES系统中多个能源子系统耦合，多种能源转化装置相互关联，同时包含新能源出力不确定性与用户用能随机性的复杂情况下，无模型深度强化学习算法求解此类变量维度高且难以精确建模的序列决策问题具有优势。

2.2 微网群能量优化调度

微网以可靠、高效、经济的方式解决分布式新能源消纳并网、负荷优化等一系列问题，但是单个微网由于其体量较小，对新能源消纳以及负荷整合优化能力有限[65]，因此对于微网群技术的研究也成为热点，相较于单个微网，微网群更加灵活可靠，且更有优势。首先，由于微网群是地理位置相邻近的微网构成，通过微网之间的能量互济共享，能够提高新能源利用率，降低传统发电机发电成本和碳排放；其次，当配电网某个区域发生故障时，该处微网可通过解列方式与故障部分隔离，并通过其他微网提供辅助服务；最后，微网群能够更加有效地进行能量转移和协调[66]，作为大电网的灵活可调度资源，有巨大潜力实现负荷优化和削峰填谷，有助于提高电网的稳定性和可靠性[67]。虽然微网群系统具有巨大技术优势，但是其结构复杂，因此针对微网群的能量管理与优化研究成为学者关注的焦点。

多个微网之间的能量互济以及微网与配网之间能量互动过程复杂，涉及备用发电、储能装置以及用户需求响应等许多决策变量，且随着微网数量的增加，非线性等约束条件也会成倍增加，导致很大的计算成本。同时由于存在诸多不确定性与随机性，对于微网群的优化调度也是一个挑战。而且在电力市场持续深入发展情况下，每个微网或者每个能源运营商作为独立主体，需要在完全信息或者不完全信息条件下与其他独立主体进行博弈，寻找最优能源交易策略以争取自身利益最大化。面对以上各种复杂问题交织耦合，为了避免构建精确复杂的模型，在引入多智能体的合作或竞争机制条件下，深度强化学习和多智能体深度强化学习是解决此类问题的有效手段。

文献[68]设计了一种混合迭代双重拍卖的方式，通过微网内部能源管理与微网间能源交易对新能源及负荷波动进行平抑，增强了对需求侧的管理，最终通过多智能体深度强化学习进行策略优化。文献[69]通过将合作机制引入基于Q学习的深度强化学习算法中，对微网群进行分布式经济优化调度，避免对微网中存在的随机性进行建模，同时降低了计算复杂度。文献[70]为保护每个微网能源交易策略的隐私性，结合自动学习机和强化学习来寻找在不完全信息下博弈的纳什均衡点，以实现每个微网的利益最大化。文献[71]提出了两种不同的储能市场模型（微网内部储能市场模型和外部储能市场模型），并通过模拟退火Q学习算法寻找储能参与微网能源市场管理的最佳策略。文献[72]提出了一种基于深度强化学习的能源交易策略，用于优化调度微网群之间的能量互动，减少对发电厂的依赖，提高了新能源利用率。文献[73]构建了多园区微网能量市场交易模型，将目标问题建模为马尔可

夫博弈模型，通过多智能体深度确定性策略梯度算法进行模型训练，在降低园区运行成本的同时，也能保护各园区主体隐私。

面对此类问题，在构建强化学习要素时，对于状态空间的设置通常包括电能/天然气等功率–价格曲线截距、节点负荷和功率、储能状态、用户需求侧响应；动作空间的设置包括多种能源的实时价格、可控机组出力等；奖励函数的设置包括各微网运营商主体利益、总体新能源消纳率、系统约束等。

表3记录了基于深度强化学习的微网能量管理优化的相关文献，分为微网场景、解决问题、求解算法、性能提升四个方面进行总结。

表 3 基于强化学习的微网能量管理优化

Table 3 Microgrid management based on reinforcement learning

文献	微网场景	解决问题	求解算法	性能提升
[43]	孤岛微网	储能协调控制	DDQN	收敛速度较DQN快
[41]	微网	光伏不确定性	DDEQN	相对于随机优化算法光伏利用率提高5.07%
[42]	微网	源荷双侧不确定性	DDPG	调度成本较蒙特卡洛法减少1.6%
[47]	微网	需求响应	DQN	累积总收益提高约10%
[49]	微网	电动汽车有序充放电	TD3	相较于快速和均匀充电节约30%开销
[50]	微网	电动汽车充放电优化	SAC	相比于AC和PG算法降低41%的负荷波动
[46]	电–热能源系统	需求响应	DDPG	在线耗时0.35 s，成本相较于传统方式下降9.2%
[54]	综合能源系统	多主体经济调度	Q学习	在线耗时相较于遗传算法降低了29.8%
[52]	综合能源系统	实时经济调度	DQN	总成本相对于粒子群优化算法降低70元/d
[53]	综合能源系统	动态经济调度	DDPG	日平均成本相较于DQN降低5.1%
[55]	综合能源系统	多场景优化调度	SAC	运行成本相较于鲁棒优化降低2.05%
[57]	家庭能源系统	综合需求响应	DQN	与遗传算法相比降低能源成本29.4%
[58]	家庭能源系统	能耗优化调度	A3C	用电费用相比于DQN降低30.7%
[61]	楼宇能量系统	在线能量优化	DDPG	相比于DQN负荷降低16.9%，成本降低23.2%
[62]	楼宇能量系统	用户用热舒适度	DDPG	相比于DQN能耗降低4.31%，舒适度提升13%
[73]	微网群	多能系统协同优化	MADDPG	相比于单独优化系统总成本降低15.44%
[69]	微网群	多个微网分布式经济调度	DQN	相比于模糊Q学习算法成本降低21.9%
[70]	微网群	能源交易	LA-Q学习	买方和卖方收益最大化
[72]	微网群	能源交易	Q学习	与轻量化RL相比发电厂发电计划降低12.7%

2.3 基于深度强化学习的微网优化控制

微网中集成了各种分布式发电机，其中分布式新能源占比较大，带来了频率失稳的风险，且微网中存在的不确定性、逆变器等电力电子装置的非线性以及复杂性，导致无法通过精确的模型进行刻画，因此传统的基于模型的方法可能无法保证控制的性能。通过利用基于无模型强化学习算法，能够在自适应微网不确定性因素和应对非线性的基础上，对微网中出现扰动和故障时的电压和频率进行控制，保持电压和频率的稳定性。

集中式的控制方式通常存在网络通信过载的问题，且单一的集中式控制方案难以承受较高的计算负担，容易发生单点故障。文献[74]为解决直流微网母线电压调节问题，将多智能体深度确定性策略梯度算法应用于二次控制器中，所提出的控制方案是完全分布式的，每个智能体只与相邻的智能体交换信息。文献[75]将深度强化学习与分布式一致性算法相结合，适用于孤岛直流微网的二次电压恢复和均流。文献[76]设计了一个基于深度确定性策略梯度的辅助控制器，通过在线学习生成辅助控制动作，能够有效应对新能源和负荷的波动，以提高二次负荷频率控制器的性能。文献[77]在具有非线性动态特性，各种状态相互耦合以及通信网络拓扑信息不明确的条件下，提出了一个

基于无模型深度强化学习算法的二次电压控制系统。功率缓冲器是一种大容量的电力电子变换器，可保护直流微网免受负荷突变的影响，文献[78]提出基于强化学习与禁忌搜索算法相结合的数据驱动方式处理功率缓冲器等电力电子器件带来的非线性优化控制问题。

面对此类问题，在构建深度强化学习要素时，对于状态空间的设置通常包括分布式发电机的电压、输出电流；动作空间的设置包括可控发电机出力；奖励函数设置包括分布式发电机分流比以及端口电压和参考电压差值等。

在以上对单个微网和微网群优化调度与控制的综述内容中可以看出，深度强化学习在应对微网中存在的源荷双侧不确定性、电力电子设备非线性等动态特性问题，多种能源系统相互耦合、大量能量转化装置相互配合等复杂系统问题时具有优势，通过基于无模型的单智能体和多智能体强化学习算法，能够在不依赖于环境精确建模的前提下，处理机组出力、投切和负荷优化等连续或者离散问题，控制系统的电压和频率稳定，满足用户用能需求，降低微网运行成本，最大化各主体的收益，同时提高新能源消纳率，保证电网的安全稳定。

2.4 深度强化学习应用场景总结分析

对深度强化学习算法在以上不同微网优化运行场景中的应用进行总结，如表4所示。DQN算法用于离散动作场景，例如在对微网中的储能系统进行优化控制中，可以对离散的储能充放电动作进行优化，在用户需求侧管理制定激励或补贴价格中，实际应用多采用DDQN来减少Q值的过估计或采用优先回放DQN算法来提高智能体学习速率。

目前微网优化运行遇到的问题多需要采用连续的动作来解决，例如发电机出力和备用调节、微网用户需求响应实时电价制定等，DDPG算法适用于此类连续动作问题，但是DDPG算法对超参数较为敏感，且存在Q值过高估计问题，通过使用TD3算法能够有效持续抑制Q值的过高估计，延迟Actor网络的更新，使算法训练更加稳定，在电动车集群的有序充电中有相关算法应用。SAC算法相较于DDPG确定性策略算法，是一种随机策略，可以增加算法的鲁棒性和探索能力，是目前非常流行的一种无模型强化学习算法，在电动汽车充电管理、综合能源系统优化调度等场景中都有相关应用。A3C算法提供了一种异步并发的强化学习框架，多个并行的Actor网络训练有助于提高探索能力，收敛性能更好，有相关文献将A3C算法应用于家庭能量管理的在线实时决策中。

在用多智能体强化学习对微网群进行优化运行研究中，一种是利用多智能体深度强化学习对微网群中的分布式电源进行协同控制，通过分布式控制的方式实现系统的协同自治和经济运行，MADDPG算法是集中式训练分布式执行架构，在实际应用场景中只需要知道局部信息就能生成最优动作策略，已经应用在微网分布式电源出力协调控制或电压二次控制中；另一种是利用多智能体强化学习求解微网多主体博弈问题，例如通过多智能体Nash-Q学习算法求解综合能源系统中的能源定价问题，但是Nash-Q学习算法只能收敛到纳什均衡解，不一定满足合理性，Wolf-PHC算法能够在其他智能体采用固定策略情况下，收敛到一个相对于对手策略的最优策略，具备一定合理性，且不需要线性规划或二次规划来求解纳什均衡，提高了求解速度，有相关研究将算法应用在发电商报价中。

表 4 代表性深度强化学习算法在微网中的应用

Table 4 Representative DRL algorithms for microgrid

描述	智能体通过与环境交互学习生成最优策略					
优点	不需要对环境进行精确建模					
代表性算法	DQN	DDPG	A3C	SAC	MADDPG	Nash-Q
特点	用于离散动作空间，存在Q值过估计问题，可选用DDQN、Dueling-DQN	用于连续动作空间，对超参数较为敏感，且存在Q值过估计问题，可选用TD3	多个并行的Actor网络训练有助于进行探索，收敛性能更好，但容易陷入局部最优	生成随机策略，可以增加算法的鲁棒性和探索能力，训练速度更快	集中式训练分布式执行的架构解决多智能体之间的通信问题	多人一般和博弈，在合作或竞争下能收敛到纳什均衡
应用场景	微网储能控制、需求侧电价制定等	电动车集群有序充放电、综合能源系统经济调度等	家庭能量系统的实时在线管理等	电动汽车充放电管理、电–气多能系统优化调度等	多园区系统经济调度、微电网二次电压控制	多能源系统博弈、多主体运营商博弈

3 应用中面临的问题和展望

在构建新型电力系统过程中，微网作为终端网络状态，将发挥终端用能交互与自治协调的重要作用。未来微网的组成将是多种能源子系统相互耦合、新能源占比高或者完全由新能源进行电能供应的小型配电系统，大量储电、储氢、储热、储冷等能源转化装置需要协调配合，同时逆变器等电力电子设备、新兴负荷接入，增加了微网的非线性、不确定性和复杂性，难以建立精确的系统模型。深度强化学习作为一种解决序贯决策问题的有效方式，因其通过与环境交互来进行学习，可以在没有先验知识和精确模型条件下寻找最优策略，同时深度强化学习中的深度神经网络也可以解决微网优化控制面临的变量维数高的问题。结合本文所述深度强化学习在微网优化运行中的研究现状和进展，对以下问题进行讨论和展望。

1）深度学习的可解释性。

目前深度强化学习等人工智能算法与微网优化调度和控制的结合还在研究阶段，并没有在实际工程中进行大规模推广，一个重要因素是因为深度学习等人工智能算法的可解释性与强功能性并没有做到完全兼容。安全可靠往往是工程应用的先决条件，因此对于人工智能黑盒模型的研究探索，提高对于深度学习的物理可解释性，有助于提高深度强化学习在优化调度与控制领域的可靠性和安全性。在微网一些优化场景中，需要保证电压、频率的稳定性，因此确保控制器的策略安全至关重要，可以采用有约束强化学习算法、鲁棒的对抗式强化学习算法与微网的电压二次控制和频率稳定相关问题进行结合，保证电网运行在合理的物理条件约束下。

2）深度强化学习奖励函数设置。

深度强化学习算法的学习性能很大程度上依赖于奖励函数的设置。由于在微网中存在诸多不确定性，各种能源设备动作特性不一，多个利益主体诉求不同，因此设计出符合要求的奖励函数是一个需要面对的挑战。在优化调度方面，经常以微网运行成本作为奖励函数来设置，在一些简单微网构成场景下具有一定的物理意义，但是在多能源耦合、多种能源设备动态投退复杂多约束场景下，需要将各种约束条件转化为奖励函数的一部分，大量动态约束使无约束决策域大幅减小，导致智能体无法学习到有效经验或者偏离优化目标。因此建立有效的探索机制非常重要，随着智能体动作维度、约束条件增加，仅依靠随机性质的探索方法很难满足要求，需要对不同状态进行有效评估，基于好奇心机制的强化学习，着力于解决高效探索采样的问题。

3）用户隐私保护。

在未来微网发展过程中，各种不同能源主体、各个微网运营商、售电商、负荷聚合商甚至各个用户之间都会涉及能量互动与交易，因此用户的隐私问题受到广泛关注。用户隐私问题是联邦学习的主要特点之一，联邦学习是一种旨在保证数据不出域的前提下，实现各参与方公平且无损地联合训练机器学习模型的方法[79]。通过强化学习与联邦学习结合，能够有效保护各主体和用户的信息安全和隐私问题。

4）迁移学习。

在未来微网运行中将会出现大量新场景，同时对于用户数据的隐私保护导致数据量可能会较少。迁移学习是通过对现有模型进行微调，将新模型应用到新的场景和领域的技术，能够充分利用原有的源数据，将通用知识用于新的目标任务中，加快模型训练速度，降低对目标任务数据的依赖性[80]。因此将迁移学习应用到基于深度强化学习的优化运行模型中，能够在快速训练模型的同时克服数据依赖性，避免模型从零开始训练，此外还可以将微网运行过程中的领域知识、专家经验编码到控制器中，能够有效提高训练效率。

5）基于模型算法和无模型算法的融合。

自从提出微网的概念以来，对于微网的研究便成为学者关注的热点，因此有一些构建成熟的模型成果可以利用。单纯使用无模型的深度强化学习算法可能会造成策略安全性等问题，且目前基于模型的算法也有一定的研究基础，未来在微网优化运行问题中可以进行基于模型算法和无模型算法的结合，优势互补，基于模型的算法可以作为基础，无模型深度强化学习算法可以用来调整基于模型控制器的参数。

6）多目标问题权重参数设计。

在微网优化运行求解过程中，由于兼具经济性、环保性和可靠性要求，面对的是一个多目标问题，目标之间可能存在冲突，现有基于深度强化学习算法在设计奖励函数时只是将多个目标进行加权求和，但不同目标的权重参数确定是一个难题。因此为了避免对此类多目标问题确定加权参数，可以采用多目标元强化学习算法来进行微网的多目标优化。

4 结论

随着电力系统中新能源的占比逐步提高，大量新兴负荷的随机接入，用户用能选择和用电行为的随机性都会对电网的安全稳定产生影响，微网的出现为应对这一趋势提供了解决方法。本文从深度强化学习在微网优化运行中的应用角度出发，对微网优化调度问题分为单个微网优化调度与微网群优化调度进行阐述。单个微网调度综述主要分为分布式电源侧、储能侧、负荷侧以及电动汽车充放电等几个方面，并选取了三种代表性的微网形式。深度强化学习在处理含有不确定性、多维变量的复杂优化问题时具有独特优势，在处理微网群能量互动与交易方面，通过多智能体深度强化学习与博弈相结合，能够在保证微网内部自治平衡的同时，与其他微网或者配网进行能量互动互济，实现各主体利益最大化。在对直流微网或者交直流微网进行优化控制方面，主要通过多智能体深度强化学习或者深度强化学习与分布式算法相结合的方式，对微网的电压和频率进行有效调节，有效应对了系统建模困难以及系统的非线性和不确定性问题。最后对深度强化学习在微网优化运行领域的应用前景进行了讨论和展望，主要从深度学习的可解释性、深度强化学习算法的奖励设置、深度强化学习与联邦学习、迁移学习的结合、有模型和无模型算法结合、多目标函数权重结合等方面进行了阐述。

参考文献

[1] MISHRA S, PETERSON K, HILIMON T, et al. Resiliency oriented control of a smart microgrid with photovoltaic modules[J]. Global Energy Interconnection, 2021, 4(5): 441-452.

[2] 任冲，柯贤波，王吉利，等. 高比例新能源电网新能源功率优化分配方法[J]. 电力工程技术，2022，41(3)：110-117.
REN Chong, KE Xianbo, WANG Jili, et al. New energy power optimal distribution method for high proportion new energy power grid[J]. Electric Power Engineering Technology, 2022, 41(3): 110-117(in Chinese).

[3] 王睿，孙秋野，胡伟，等. 面向“碳达峰、碳中和”的新型电力系统的潮流计算[J]. 全球能源互联网，2022，5(5)：439-446.
WANG Rui, SUN Qiuye, HU Wei, et al. Power flow calculation of new power system for carbon peaking and carbon neutralization[J]. Journal of Global Energy Interconnection, 2022, 5(5): 439-446(in Chinese).

[4] 孙海霞，朱立位，韩钰倩，等. 基于非合作博弈的微网混合储能容量配置方法[J]. 全球能源互联网，2021，4(5)：454-463.
SUN Haixia, ZHU Liwei, HAN Yuqian, et al. Capacity configuration method of hybrid energy storage system in microgrids based on a non-cooperative game model[J]. Journal of Global Energy Interconnection, 2021, 4(5): 454-463(in Chinese).

[5] MORSTYN T, HREDZAK B, AGELIDIS V G. Control strategies for microgrids with distributed energy storage systems: an overview[J]. IEEE Transactions on Smart Grid, 2018, 9(4): 3652-3666.

[6] ANDERSON A A, SURYANARAYANAN S. Review of energy management and planning of islanded microgrids[J]. CSEE Journal of Power and Energy Systems, 2020, 6(2): 329-343.

[7] 刘俊峰，罗燕，侯媛媛，等. 考虑广义储能的微电网主动能量管理优化算法研究[J]. 电网技术，2023，47(1)：245-255.
LIU Junfeng, LUO Yan, HOU Yuanyuan, et al. Research on optimization algorithm of active microgrid energy management considering generalized energy storage[J]. Power System Technology, 2023, 47(1): 245-255 (in Chinese).

[8] NASSER N, FAZELI M. Buffered-microgrid structure for future power networks; a seamless microgrid control[J]. IEEE Transactions on Smart Grid, 2021, 12(1): 131-140.

[9] WU X, ZHAO W C, WANG X L, et al. An MILP-based planning model of a photovoltaic/diesel/battery stand-alone microgrid considering the reliability[J]. IEEE Transactions on Smart Grid, 2021, 12(5): 3809-3818.

[10] QIU H F, GU W, XU Y L, et al. Tri-level mixed-integer optimization for two-stage microgrid dispatch with multi-uncertainties[J]. IEEE Transactions on Power Systems, 2020, 35(5): 3636-3647.

[11] 苏磊，李振坤，张智泉，等. 基于机会约束规划的综合能源微网群协调运行策略研究[J]. 电力系统保护与控制，2021，49(14)：123-131.
SU Lei, LI Zhenkun, ZHANG Zhiquan, et al. A coordinated operation strategy for integrated energy microgrid clusters based on chance-constrained programming[J]. Power System Protection and Control, 2021, 49(14): 123-131(in Chinese).

[12] 许书伟，吴文传，朱涛，等. 机会约束随机动态经济调度的凸松弛迭代求解法[J]. 电力系统自动化，2020，44(17)：43-51.
XU Shuwei, WU Wenchuan, ZHU Tao, et al. Convex relaxation based iterative solution method for stochastic dynamic economic dispatch with chance constrains[J]. Automation of Electric Power Systems, 2020, 44(17): 43-51(in Chinese).

[13] LI G L, GOMEZ R, NAKAMURA K, et al. Human-centered reinforcement learning: a survey[J]. IEEE Transactions on Human-Machine Systems, 2019, 49(4): 337-349.

[14] 徐博涵，向月，潘力，等. 基于深度强化学习的含高比例可再生能源配电网就地分散式电压管控方法[J]. 电力系统

保护与控制，2022，50(22)：100-109.
XU Bohan, XIANG Yue, PAN Li, et al. Local decentralized voltage management of a distribution network with a high proportion of renewable energy based on deep reinforcement learning[J]. Power System Protection and Control, 2022, 50(22): 100-109(in Chinese).

[15] HUANG Y M, XU C M, ZHANG C, et al. An overview of intelligent wireless communications using deep reinforcement learning[J]. Journal of Communications and Information Networks, 2019, 4(2): 15-29.

[16] BRUCE J, SUENDERHAUF N, MIROWSKI P, et al. One-shot reinforcement learning for robot navigation with interactive replay[EB/OL]. 2017: arXiv: 1711.10137[cs.AI]. https://arxiv.org/abs/1711.10137.

[17] MNIH V, KAVUKCUOGLU K, SILVER D, et al. Playing Atari with deep reinforcement learning[EB/OL]. 2013: arXiv:1312.5602 [cs.LG]. https://arxiv.org/abs/1312.5602.

[18] PRASAD A, DUSPARIC I. Multi-agent deep reinforcement learning for zero energy communities[EB/OL]. 2018: arXiv: 1810.03679[cs.LG]. https://arxiv.org/abs/1810.03679.

[19] CEUSTERS G, RODRIGUEZ R C, GARCIA A B, et al. Model-predictive control and reinforcement learning in multi-energy system case studies[J]. Applied Energy, 2021, 303: 117634.

[20] 刘全，翟建伟，章宗长，等. 深度强化学习综述[J]. 计算机学报，2018，41(1)：1-27.
LIU Quan, ZHAI Jianwei, ZHANG Zongzhang, et al. A survey on deep reinforcement learning[J]. Chinese Journal of Computers, 2018, 41(1): 1-27(in Chinese).

[21] SUTTON R S, BARTO A G. Reinforcement learning: an introduction[M]. MIT Press, 2018.

[22] NIE W, HUANG K, YANG J, et al. A deep reinforcement learning-based framework for PolSAR imagery classification [J]. IEEE Transactions on Geoscience and Remote Sensing, 2022, 60: 1-15.

[23] 杜威，丁世飞. 多智能体强化学习综述[J]. 计算机科学，2019，46(8)：1-8.
DU Wei, DING Shifei. Overview on multi-agent reinforcement learning[J]. Computer Science, 2019, 46(8): 1-8(in Chinese).

[24] 孙长银，穆朝絜. 多智能体深度强化学习的若干关键科学问题[J]. 自动化学报，2020，46(7)：1301-1312.
SUN Changyin, MU Chaoxu. Important scientific problems of multi-agent deep reinforcement learning[J]. Acta Automatica Sinica, 2020, 46(7): 1301-1312(in Chinese).

[25] 马骋乾，谢伟，孙伟杰. 强化学习研究综述[J]. 指挥控制与仿真，2018，40(6)：68-72.
MA Chengqian, XIE Wei, SUN Weijie. Research on reinforcement learning technology: a review[J]. Command Control & Simulation, 2018, 40(6): 68-72(in Chinese).

[26] 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]. 模式识别与人工智能，2019，32(1)：67-81.
WAN Lipeng, LAN Xuguang, ZHANG Hanbo, et al. A review of deep reinforcement learning theory and application[J]. Pattern Recognition and Artificial Intelligence, 2019, 32(1): 67-81(in Chinese).

[27] 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(in Chinese).

[28] 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，2016，33(6)：701-717.
ZHAO Dongbin, SHAO Kun, ZHU Yuanheng, et al. Review of deep reinforcement learning and discussions on the development of computer Go[J]. Control Theory & Applications, 2016, 33(6): 701-717(in Chinese).

[29] MAHMUD M, KAISER M S, HUSSAIN A, et al. Applications of deep learning and reinforcement learning to biological data[J]. IEEE Transactions on Neural Networks and Learning Systems, 2018, 29(6): 2063-2079.

[30] 张怡，张恒旭，李常刚，等. 深度学习在电力系统频率分析与控制中的应用综述[J]. 中国电机工程学报，2021，41(10)：3392-3406.
ZHANG Yi, ZHANG Hengxu, LI Changgang, et al. Review on deep learning applications in power system frequency analysis and control[J]. Proceedings of the CSEE, 2021, 41(10): 3392-3406(in Chinese).

[31] KIRAN B R, SOBH I, TALPAERT V, et al. Deep reinforcement learning for autonomous driving: a survey[J]. IEEE Transactions on Intelligent Transportation Systems, 2022, 23(6): 4909-4926.

[32] ARADI S. Survey of deep reinforcement learning for motion planning of autonomous vehicles[J]. IEEE Transactions on Intelligent Transportation Systems, 2022, 23(2): 740-759.

[33] 梁星星，冯旸赫，马扬，等. 多Agent深度强化学习综述[J]. 自动化学报，2020，46(12)：2537-2557.
LIANG Xingxing, FENG Yanghe, MA Yang, et al. Deep multi-agent reinforcement learning: a survey[J]. Acta Automatica Sinica, 2020, 46(12): 2537-2557(in Chinese).

[34] NGUYEN T T, NGUYEN N D, NAHAVANDI S. Deep reinforcement learning for multiagent systems: a review of challenges, solutions, and applications[J]. IEEE Transactions on Cybernetics, 2020, 50(9): 3826-3839.

[35] 袁桂丽，贾新潮，房方，等. 虚拟电厂源荷双侧热电联合随机优化调度[J]. 电网技术，2020，44(8)：2932-2940.
YUAN Guili, JIA Xinchao, FANG Fang, et al. Joint stochastic optimal scheduling of heat and power considering source and load sides of virtual power plant[J]. Power System Technology, 2020, 44(8): 2932-2940(in Chinese).

[36] DANESHVAR M, MOHAMMADI-IVATLOO B, ZARE K, et al. Two-stage robust stochastic model scheduling for

transactive energy based renewable microgrids[J]. IEEE Transactions on Industrial Informatics, 2020, 16(11): 6857-6867.

[37] 黎静华，朱梦姝，陆悦江，等. 综合能源系统优化调度综述[J]. 电网技术，2021，45(6)：2256-2272.
LI Jinghua, ZHU Mengshu, LU Yuejiang, et al. Review on optimal scheduling of integrated energy systems[J]. Power System Technology, 2021, 45(6): 2256-2272(in Chinese).

[38] TIAN L T, CHENG L, GUO J B, et al. System modeling and optimal dispatching of multi-energy microgrid with energy storage[J]. Journal of Modern Power Systems and Clean Energy, 2020, 8(5): 809-819.

[39] 李存斌，张建业，李鹏. 考虑成本、排污及风险的微电网运营多目标优化模型[J]. 中国电机工程学报，2015，35(5)：1051-1058.
LI Cunbin, ZHANG Jianye, LI Peng. Multi-objective optimization model of micro-grid operation considering cost, pollution discharge and risk[J]. Proceedings of the CSEE, 2015, 35(5): 1051-1058(in Chinese).

[40] LIN L, GUAN X, PENG Y, et al. Deep reinforcement learning for economic dispatch of virtual power plant in Internet of energy[J]. IEEE Internet of Things Journal, 2020, 7(7): 6288-6301.

[41] 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 Changsen, ZHANG Yu, WEN Fushuan, 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(in Chinese).

[42] 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.
PENG Liuyang, SUN Yuanzhang, XU Jian, et al. Self-adaptive uncertainty economic dispatch based on deep reinforcement learning[J]. Automation of Electric Power Systems, 2020, 44(9): 33-42(in Chinese).

[43] 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，2019，43(6)：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, 43(6): 1914-1921(in Chinese).

[44] 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.
LIANG Hong, LI Hongxin, ZHANG Huaying, et al. Control strategy of microgrid energy storage system based on deep reinforcement learning[J]. Power System Technology, 2021, 45(10): 3869-3877(in Chinese).

[45] WANG B, LI Y, MING W Y, et al. Deep reinforcement learning method for demand response management of interruptible load[J]. IEEE Transactions on Smart Grid, 2020, 11(4): 3146-3155.

[46] 严干贵，阚天洋，杨玉龙，等. 基于深度强化学习的分布式电采暖参与需求响应优化调度[J]. 电网技术，2020，44(11)：4140-4149.
YAN Gangui, KAN Tianyang, YANG Yulong, et al. Demand response optimal scheduling for distributed electric heating based on deep reinforcement learning[J]. Power System Technology, 2020, 44(11): 4140-4149(in Chinese).

[47] 徐弘升，陆继翔，杨志宏，等. 基于深度强化学习的激励型需求响应决策优化模型[J]. 电力系统自动化，2021，45(14)：97-103.
XU Hongsheng, LU Jixiang, YANG Zhihong, et al. Decision optimization model of incentive demand response based on deep reinforcement learning[J]. Automation of Electric Power Systems, 2021, 45(14): 97-103(in Chinese).

[48] ZHONG S Y, WANG X Y, ZHAO J, et al. Deep reinforcement learning framework for dynamic pricing demand response of regenerative electric heating[J]. Applied Energy, 2021, 288: 116623.

[49] 赵星宇，胡俊杰. 集群电动汽车充电行为的深度强化学习优化方法[J]. 电网技术，2021，45(6)：2319-2327.
ZHAO Xingyu, HU Junjie. Deep reinforcement learning based optimization for charging of aggregated electric vehicles[J]. Power System Technology, 2021, 45(6): 2319-2327(in Chinese).

[50] 刘敦楠，王玲湘，汪伟业，等. 基于深度强化学习的大规模电动汽车充换电负荷优化调度[J]. 电力系统自动化，2022，46(4)：36-46.
LIU Dunnan, WANG Lingxiang, WANG Weiye, et al. Optimal scheduling of electric vehicle load for large-scale battery charging and swapping based on deep reinforcement learning[J]. Automation of Electric Power Systems, 2022, 46(4): 36-46(in Chinese).

[51] LI H P, WAN Z Q, HE H B. Constrained EV charging scheduling based on safe deep reinforcement learning[J]. IEEE Transactions on Smart Grid, 2020, 11(3): 2427-2439.

[52] LIU Y K, ZHANG D X, GOOI H B. Optimization strategy based on deep reinforcement learning for home energy management[J]. CSEE Journal of Power and Energy Systems, 2020, 6(3): 572-582.

[53] 苏永新，吴泽旋，谭貌，等. 基于深度强化学习的家庭综合需求响应在线优化[J]. 中国电机工程学报，2021，41(16)：5581-5593.
SU Yongxin, WU Zexuan, TAN Mao, et al. Online optimization for home integrated demand response based on deep reinforcement learning[J]. Proceedings of the CSEE, 2021, 41(16): 5581-5593(in Chinese).

[54] 张虹，申鑫，穆昊源，等. 基于Multi-Agent异步深度强化学习的居民住宅能耗在线优化调度研究[J]. 中国电机工程学报，2020，40(1)：117-127.

ZHANG Hong, SHEN Xin, MU Haoyuan, et al. Research on online optimal dispatching of residential energy consumption based on multi-agent asynchronous deep reinforcement learning[J]. Proceedings of the CSEE, 2020, 40(1): 117-127(in Chinese).

[55] 张甜，赵奇，陈中，等. 基于深度强化学习的家庭能量管理分层优化策略[J]. 电力系统自动化，2021，45(21)：149-158.
ZHANG Tian, ZHAO Qi, CHEN Zhong, et al. Hierarchical optimization strategy for home energy management based on deep reinforcement learning[J]. Automation of Electric Power Systems, 2021, 45(21): 149-158(in Chinese).

[56] WEI T S, WANG Y Z, ZHU Q. Deep reinforcement learning for building HVAC control[C]//2017 54th ACM/EDAC/IEEE Design Automation Conference. Austin, TX, USA. IEEE: 1-6.

[57] 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, 2019, 10(4): 3698-3708.

[58] GAO G Y, LI J, WEN Y G. DeepComfort: energy-efficient thermal comfort control in buildings via reinforcement learning[J]. IEEE Internet of Things Journal, 2020, 7(9): 8472-8484.

[59] NAGARATHINAM S, MENON V, VASAN A, et al. MARCO - Multi-Agent Reinforcement learning based COntrol of building HVAC systems[C]//Proceedings of the Eleventh ACM International Conference on Future Energy Systems. Virtual Event Australia. New York, NY, USA: ACM, 2020.

[60] DING X Z, DU W, CERPA A. OCTOPUS: deep reinforcement learning for holistic smart building control[C]//Proceedings of the 6th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation. New York, NY, USA: ACM, 2019.

[61] 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.
NIE Huanhuan, ZHANG Jiaqi, CHEN Ying, et al. Real-time economic dispatch of community integrated energy system based on a double-layer reinforcement learning method[J]. Power System Technology, 2021, 45(4): 1330-1336(in Chinese).

[62] YANG T. Dynamic energy dispatch strategy for integrated energy system based on improved deep reinforcement learning[J]. Energy, 2021, 235: 121377.

[63] 刘洪，李吉峰，葛少云，等. 基于多主体博弈与强化学习的并网型综合能源微网协调调度[J]. 电力系统自动化，2019，43(1)：40-48.
LIU Hong, LI Jifeng, GE Shaoyun, et al. Coordinated scheduling of grid-connected integrated energy microgrid based on multi-agent game and reinforcement learning[J]. Automation of Electric Power Systems, 2019, 43(1): 40-48(in Chinese).

[64] 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(in Chinese).

[65] 刘迎澍，陈曦，李斌，等. 多微网系统关键技术综述[J]. 电网技术，2020，44(10)：3804-3820.
LIU Yingshu, CHEN Xi, LI Bin, et al. State of art of the key technologies of multiple microgrids system[J]. Power System Technology, 2020, 44(10): 3804-3820(in Chinese).

[66] 李鹏，吴迪凡，李雨薇，等. 基于谈判博弈的多微网综合能源系统多目标联合优化配置[J]. 电网技术，2020，44(10)：3680-3690.
LI Peng, WU Difan, LI Yuwei, et al. Multi-objective union optimal configuration strategy for multi-microgrid integrated energy system considering bargaining games[J]. Power System Technology, 2020, 44(10): 3680-3690(in Chinese).

[67] ZOU H L, MAO S W, WANG Y, et al. A survey of energy management in interconnected multi-microgrids[J]. IEEE Access, 2019, 7: 72158-72169.

[68] LATIFI M, RASTEGARNIA A, KHALILI A, et al. A self-governed online energy management and trading for smart micro/nano-grids[J]. IEEE Transactions on Industrial Electronics, 2020, 67(9): 7484-7498.

[69] LIU W R, ZHUANG P, LIANG H, et al. Distributed economic dispatch in microgrids based on cooperative reinforcement learning[J]. IEEE Transactions on Neural Networks and Learning Systems, 2018, 29(6): 2192-2203.

[70] WANG H W, HUANG T W, LIAO X F, et al. Reinforcement learning in energy trading game among smart microgrids[J]. IEEE Transactions on Industrial Electronics, 2016, 63(8): 5109-5119.

[71] SESETTI A, NUNNA H S V S K, DOOLLA S, et al. Multi-agent based energy trading platform for energy storage systems in distribution systems with inter-connected microgrids[C]//2018 IEEE Industry Applications Society Annual Meeting. Portland, OR, USA. IEEE: 1-8.

[72] LU X Z, XIAO X Y, XIAO L, et al. Reinforcement learning-based microgrid energy trading with a reduced power plant schedule[J]. IEEE Internet of Things Journal, 2019, 6(6): 10728-10737.

[73] 杨照，黄少伟，陈颖. 基于多智能体强化学习的多园区综合能源系统协同优化运行研究[J]. 电工电能新技术，2021，40(8)：1-10.
YANG Zhao, HUANG Shaowei, CHEN Ying. Research on cooperative optimal operation of multi-park integrated energy system based on multi agent reinforcement learning[J]. Advanced Technology of Electrical Engineering and Energy, 2021, 40(8): 1-10(in Chinese).

[74] XIA Y, XU Y, WANG Y, et al. A distributed control in

islanded DC microgrid based on multi-agent deep reinforcement learning[C]//IECON 2020 The 46th Annual Conference of the IEEE Industrial Electronics Society. Singapore. IEEE: 2359-2363.

[75] LIU Z F, LUO Y, ZHUO R Q, et al. Distributed reinforcement learning to coordinate current sharing and voltage restoration for islanded DC microgrid[J]. Journal of Modern Power Systems and Clean Energy, 2018, 6(2): 364-374.

[76] KHOOBAN M H, GHEISARNEJAD M. A novel deep reinforcement learning controller based type-II fuzzy system: frequency regulation in microgrids[J]. IEEE Transactions on Emerging Topics in Computational Intelligence, 2021, 5(4): 689-699.

[77] SMITH E, ROBINSON D A, AGALGAONKAR A. Cooperative secondary voltage control of static converters in a microgrid using model-free reinforcement learning[C]//2019 21st European Conference on Power Electronics and Applications (EPE '19 ECCE Europe). September 3-5, 2019. Genova, Italy. IEEE, 2019.

[78] MASSENIO P R, NASO D, LEWIS F L, et al. Data-driven sparsity-promoting optimal control of power buffers in DC microgrids[J]. IEEE Transactions on Energy Conversion, 2021, 36(3): 1919-1930.

[79] 李少波，杨磊，李传江，等. 联邦学习概述：技术、应用及未来[J]. 计算机集成制造系统，2022，28(7)：2119-2138.
LI Shaobo, YANG Lei, LI Chuanjiang, et al. Overview of federated learning: technology, applications and future[J]. Computer Integrated Manufacturing Systems, 2022, 28(7): 2119-2138(in Chinese).

[80] 张博，史忠植，赵晓非，等. 一种基于跨领域典型相关性分析的迁移学习方法[J]. 计算机学报，2015，38(7)：1326-1336.
ZHANG Bo, SHI Zhongzhi, ZHAO Xiaofei, et al. A transfer learning based on canonical correlation analysis across different domains[J]. Chinese Journal of Computers, 2015, 38(7): 1326-1336(in Chinese).

收稿日期：2022-04-17；修回日期：2022-08-15。


周翔

作者简介：

周翔（1991），男，博士研究生，研究方向为人工智能在电力系统中的应用。通信作者，E-mail：878585460@qq.com。

王继业（1964），男，教授级高级工程师，博士生导师，研究方向为能源互联网、智能电网、电力信息通信技术等。

陈盛（1989），男，高级工程师，博士研究生，研究方向为人工智能在电力系统中的应用。

王新迎（1987），男，高级工程师，博士，研究方向为人工智能在电力系统中的应用。

（责任编辑　李锡）