

基于深度强化学习的微能源网 能量管理与优化策略研究

刘俊峰¹, 陈剑龙¹, 王晓生¹, 曾君², 黄倩颖¹

(1. 华南理工大学 自动化科学与工程学院, 广东省 广州市 510640;

2. 华南理工大学 电力学院, 广东省 广州市 510640)

Energy Management and Optimization of Multi-energy Grid Based on Deep Reinforcement Learning

LIU Junfeng¹, CHEN Jianlong¹, WANG Xiaosheng¹, ZENG Jun², HUANG Qianying¹

(1. School of Automation Science and Engineering, South China University of Technology,

Guangzhou 510640, Guangdong Province, China;

2. School of Electric Power, South China University of Technology, Guangzhou 510640, Guangdong Province, China)

ABSTRACT: A energy management and optimization method based on deep reinforcement learning (DRL) for a micro-energy grid connected with a variety of renewable energy sources is proposed in this paper. This method adopts deep Q network (DQN) to learn environmental information including load forecasting, renewable energy power output and time-of-use electricity price, and manages the energy of micro-energy grid through the acquired strategy set, which is a model-free and value-based intelligent algorithm. Firstly, an energy bus-based structure of multi-energy grid is introduced. Secondly, the frame of reinforcement learning, the basic theories of the Q-learning algorithm and the DQN algorithm are presented. The mechanism of experience replay and fixed network is introduced to enhance the algorithm performance. Furthermore, for the purpose of economy, the performances of Q-learning and DQN with different parameters in energy management of multi-energy grid are compared and the effectiveness of the algorithm is verified by simulation. It can be more effective to apply inherited strategy set on DQN, leading to better optimization results and faster iterations.

KEY WORDS: multi-energy grid; energy management; deep reinforcement learning; Q-learning; deep Q network

摘要: 面向多种可再生能源接入的微能源网, 提出一种基于深度强化学习(deep reinforcement learning, DRL)的微能源网能量管理与优化方法。该方法使用深度 Q 网络(deep Q network, DQN)对预测负荷、风/光等可再生能源功率输出和分时电价等环境信息进行学习, 通过习得的策略集对微能源

网进行能量管理, 是一种模型无关基于价值的智能算法。首先, 基于能量总线模型, 建立了微能源网研究框架及设备模型。在深入阐述强化学习的框架、Q 学习算法和 DQN 算法的基础理论的基础上, 分析了提升 DQN 性能的经验回放机制与冻结参数机制, 并以经济性为目标完成了微能源网能量管理与优化。通过对比不同参数的 DQN 算法及 Q 学习算法在微能源网能量管理中的表现, 仿真结果展示了继承策略集后算法性能的提升, 验证了深度强化学习相比启发式算法在微能源网能量管理应用的可行性和优越性。

关键词: 微能源网; 能量管理; 深度强化学习; Q 学习; 深度 Q 网络

DOI: 10.13335/j.1000-3673.pst.2020.0144

0 引言

建设综合能源系统(integrated energy systems, IESs)是提高能源利用率, 减少二氧化碳排放, 增加可再生能源比例的有效解决方案之一。微能源网是一种面向楼宇、社区、工业园区或城镇等终端能源用户的微型综合能源系统, 是未来人类社会能源的主要承载形式之一。微能源网接受多类型的能源输入(电、可再生能源、天然气等), 经过多类型的能量转换(气/电、气/热、电/热、电/冷、热/冷等)装置和储能装置, 满足多类型的能源(电、热、冷)消费需求^[1]。

微能源网通常以联产系统为核心, 接入多种可再生能源和储能系统, 通过恰当的系统配置和运行策略, 实现多能互补和能量的梯级利用, 最终实现能源的高效灵活利用。在微能源网中, 可再生能源的不稳

基金项目: 国家自然科学基金(61573155, 51877085)。

Project Support by National Natural Science Foundation of China (61573155, 51877085).

定性、能量流动的多向性,以及多种能量在不同时间和空间层面调度需求等问题,使得多能流耦合既是综合能源系统的特点,更是其能量管理面临的重大挑战之一^[2]。寻求有效的优化策略,实现系统优质高效的能量管理,对提高系统的性能十分重要。

根据微电网或微能源网是否与主电网相连接,可将其分为并网型和独立型2种。本文以并网型微能源网为研究对象,研究其并网运行的能量管理与优化问题。目前,针对微能源网的能量管理,从算法上来讲,多结合最优化算法或者启发式算法进行。文献[3]建立了微网混合整数非线性模型,通过将其分解为组合问题和最优潮流问题,避免直接求解混合整数非线性问题,加快了寻优速度。文献[4]通过概率约束对旋转备用储能的不确定性进行建模,提出了一种使用机会约束规划来最小化微网运营成本的最优调度方法。文献[5]在考虑可再生能源随机性和用户满意度的情况下,热电解耦的同时通过基于概率约束规划法优化,使运行成本最低。文献[6]提出一种基于记忆机制的遗传算法,对微网运行成本最小化问题进行求解。文献[7]提出了一种新的混合粒子群优化算法对含有可再生能源的微电网进行优化。文献[8]在考虑符合满意度的基础上,利用NSGA-II算法对微电网运行进行多目标优化。文献[9]基于Stone-Geary函数和双向拍卖机制构建了智慧能源系统能量优化模型,给出了通过收敛判定域引导的全局随机寻优与区域定向寻优策略。最优化算法的计算效率较高,但是其在处理非线性、非凸或非连续问题时难以逃出局部最优。启发式算法可以在给定的条件下求得相应的最优解或Pareto前沿,但限制多,计算耗时长且算法泛化学习能力不足。

随着人工智能技术的发展,强化学习(reinforcement learning, RL)受到研究者的关注,并开始应用于风储协作^[10]、风力发电控制^[11]、负荷响应^[12]等电力系统领域。文献[13]引入强化学习建立相应模型用于微网能量管理,取得了较好的效果。从这些研究来看,强化学习算法非常适用于解决能量调度及成本最小化问题。但是,这些研究多关注于微电网,而且基于传统强化学习算法,在面对复杂环境或连续动作时无法解决策略集的维数灾难,难以实际应用与推广,若将其推广到更为复杂的多能源网将面临更大的挑战。随着谷歌Deepmind团队在2015和2016年的《Nature》杂志上分别公布了基于Atari视频游戏的深度强化学习算法^[14]和计算机围棋机器人AlphaGo^[15-16],以深度强化学习为代表的新一代人工智能技术,迎来了蓬勃发展的新

时代。将具备感知能力的深度学习(deep learning, DL)引入具备决策能力的强化学习,形成的深度强化学习(deep reinforcement learning, DRL)算法展现出了超越以往大部分人工智能算法的性能^[17]。近年来,DRL在电力系统的应用也引起了一些学者的关注。文献[18]提出一种基于深度强化学习的微电网复合储能协调控制方法,依据电网运行环境信息通过数据分析得到控制策略。文献[19]根据采用深度神经网络预测的负荷及光伏出力,结合DL对微电网系统进行最优能量调度,显著降低了用电成本。文献[20]基于能源互联网的运行优化问题,应用无模型深度强化学习算法得到解决方案,并通过数值模拟评估了所提方法的可行性和性能。

本文以微能源网为对象,建立了基于能源总线模型的微能源网系统,提出利用深度强化学习算法对微能源网进行能量管理与优化策略研究。面向微能源网系统,采用经历回放机制和冻结网络参数机制提升深度强化学习算法的性能,并通过深度神经网络储存策略集解决传统强化学习的维数灾难,实现对微能源网的能量管理与优化,有效解决了具有随机性和间歇性的面向可再生能源的微能源网运行优化所面临的建模困难、传统算法运行收敛较慢难以满足实时优化要求以及系统开放性问题。更进一步,本文在验证继承已训练策略集的强化学习算法的优越性的基础上,对比了在离线数据集上训练完成的深度强化学习算法与启发式算法(以遗传算法为例)对同一新负荷场景的优化结果与计算时间,证明了深度强化学习的可行性与优越性。对比还证明了深度强化学习不仅可以实现对单一场景进行能量管理与策略优化,还具备优秀的泛化能力,实现对新场景能量管理的快速收敛并获得趋优解。

1 基于能量总线的微能源网系统结构及设备模型

以图1为基础,本文给出基于能量总线模型的微能源网系统的结构,如图2所示。本文按照能量传输介质分为电、热、冷3条能量总线(相同传输介质的能量总会由源流向总线,再由总线流向负荷)。忽略线路损耗及能量转换过程的损耗,根据微能源网的结构和总线模型,可列出微能源网系统的总线能量平衡如式(1)所示。

$$\sum_{i=1}^N P_i = \sum_{i=1}^M L_i \quad (1)$$

式中:假设总线上连接着 N 个能量源和 M 个负荷, P_i 为各能量源的功率,kW; L_i 为各负荷所需功率,

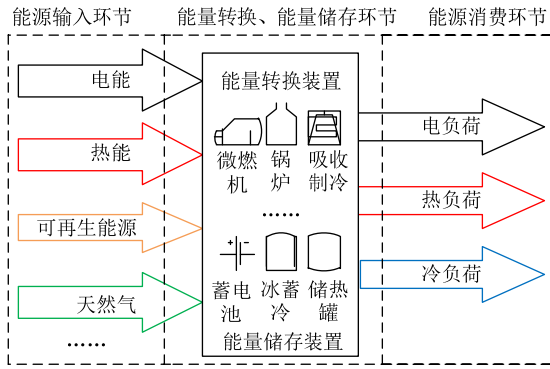


图1 微能源网系统的基本组成

Fig. 1 Basic elements of multi-energy grid

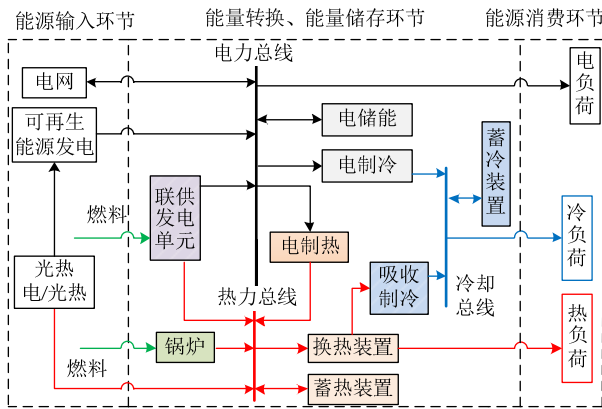


图2 基于能量总线的微能源网系统

Fig. 2 Energy bus-based structure of multi-energy grid system

kW。针对图2中的模型，可知电、热、冷均须实现能量平衡。

1.1 联供发电单元模型

联供发电单元是微能源网的核心设备。联供发电单元的发电功率为

$$P_{MT}(t) = V_{MT}(t)q_{NG}\eta_{MT} \quad (2)$$

式中： $P_{MT}(t)$ 为 t 时段联供发电单元的发电功率，kW； $V_{MT}(t)$ 为 t 时段联供发电单元的单位时间天然气消耗量， m^3/h ； q_{NG} 为天然气的低热值， $kW \cdot h/m^3$ ； η_{MT} 为联供发电单元的发电效率。

联供发电单元的发电功率满足约束条件

$$P_{MT,min} < P_{MT}(t) < P_{MT,max} \quad (3)$$

式中 $P_{MT,max}$ 、 $P_{MT,min}$ 分别为联供发电单元发电功率的上下限值。

联供发电单元除了将化学能转化为电能外，还会产生大量具有热量的烟气，其数学表达式为

$$Q_{MT}(t) = V_{MT}(t)q_{NG}(1 - \eta_{MT}) \quad (4)$$

式中 $Q_{MT}(t)$ 为 t 时段联供发电单元的输出热功率，kW。

联供发电单元输出热功率的约束为

$$Q_{MT,min} < Q_{MT}(t) < Q_{MT,max} \quad (5)$$

式中 $Q_{MT,max}$ 、 $Q_{MT,min}$ 分别为联供发电单元输出热功率的上下限值。

1.2 余热回收锅炉模型

余热回收锅炉收集来自联供发电单元的高温烟气供给系统的热负荷需求。它的输出热功率与输入热功率和余热回收锅炉换热效率有关。

$$Q_{HR}(t) = Q_{HR,in}(t)\eta_{HR} \quad (6)$$

式中： $Q_{HR}(t)$ 为 t 时段余热回收锅炉的输出热功率，kW； $Q_{HR,in}(t)$ 为 t 时段余热回收锅炉的输入热功率，kW； η_{HR} 为余热回收锅炉的换热效率。

余热回收锅炉的输出热功率满足约束

$$Q_{HR,min} < Q_{HR}(t) < Q_{HR,max} \quad (7)$$

式中 $Q_{HR,max}$ 、 $Q_{HR,min}$ 分别为余热回收锅炉输出热功率的上下限。

1.3 燃气锅炉模型

燃气锅炉是微能源网中产生热能的设备，其输出热功率为

$$Q_{SB}(t) = V_{SB}(t)q_{NG}\eta_{SB} \quad (8)$$

式中： $Q_{SB}(t)$ 为 t 时段燃气锅炉的输出热功率，kW； $V_{SB}(t)$ 为 t 时段燃气锅炉的单位时间天然气消耗量， m^3/h ； η_{SB} 为燃气锅炉的效率。

燃气锅炉输出热功率满足约束条件

$$Q_{SB,min} < Q_{SB}(t) < Q_{SB,max} \quad (9)$$

式中 $Q_{SB,max}$ 、 $Q_{SB,min}$ 分别为燃气锅炉输出热功率的上下限值。

1.4 换热装置模型

换热装置将联供发电单元和燃气锅炉产生的热能进行转化，供给微能源网的热负荷需求。换热装置的输出热功率为

$$Q_{HX}(t) = Q_{HX,in}(t)\eta_{HX} \quad (10)$$

式中： $Q_{HX}(t)$ 为 t 时段换热装置的输出热功率，kW； $Q_{HX,in}(t)$ 为 t 时段换热装置的输入热功率，kW； η_{HX} 为换热装置的换热效率。

换热装置输出热功率满足约束条件

$$Q_{HX,min} < Q_{HX}(t) < Q_{HX,max} \quad (11)$$

式中 $Q_{HX,max}$ 、 $Q_{HX,min}$ 分别为换热装置输出热功率的上下限值。

1.5 电制冷机模型

电制冷机是微能源网中唯一产生冷能的设备，消耗电能产生冷能来满足冷负荷的需求。电制冷机的制冷功率为

$$Q_{EC}(t) = P_{EC}(t)\eta_{EC} \quad (12)$$

式中： $Q_{EC}(t)$ 为 t 时段电制冷机的制冷功率，kW；

$P_{EC}(t)$ 为 t 时段电制冷机消耗的电功率, kW; η_{EC} 为电制冷机的性能系数。

电制冷机的制冷功率满足约束条件

$$Q_{EC,min} < Q_{EC}(t) < Q_{EC,max} \quad (13)$$

式中 $Q_{EC,max}$ 、 $Q_{EC,min}$ 分别为电制冷机制冷功率的上下限值。

1.6 电储能模型

电储能是微能源网的重要组成部分, 能在一定程度上缓解微能源网中可再生能源和负荷不确定性。本文以蓄电池为例建立电储能模型。蓄电池的荷电状态与上一时段的荷电状态和蓄电池的充电功率有关, 满足以下差分方程。

$$\Phi(t+1) = \Phi(t) + \frac{P_{BC}(t)\Delta t - P_{BD}(t)\Delta t}{W_B} \quad (14)$$

式中: $\Phi(t)$ 为蓄电池 t 时刻的荷电状态; $P_{BC}(t)$ 、 $P_{BD}(t)$ 分别为 t 时段蓄电池的充、放电功率, kW; W_B 为蓄电池的最大容量, kW·h; Δt 为时间步长。

为了尽量延长电池的使用寿命, 规定蓄电池储能约束和蓄电池充放电功率约束如下:

1) 蓄电池储能约束。

$$\eta_{B,min} < \Phi(t) < \eta_{B,max} \quad (15)$$

式中 $\eta_{B,max}$ 、 $\eta_{B,min}$ 分别为蓄电池的最大储能系数、最小储能系数。

2) 蓄电池充放电功率约束。

$$0 \leq P_{BC}(t)\Delta t \leq W_B\eta_{BC} \quad (16)$$

$$0 \leq P_{BD}(t)\Delta t \leq W_B\eta_{BD} \quad (17)$$

式中 η_{BC} 、 η_{BD} 分别为蓄电池的最大充、放电率。

1.7 约束条件

系统的运行约束条件包括各个设备自身的约束和各能量流的平衡约束条件, 设备的自身约束已在上文中给出。根据微能源网的结构和总线模型, 各总线能量平衡约束如下。

1) 电力总线能量平衡方程。

$$P_{MT}(t) + P_{BD}(t) + P_g(t) + P_{PV}(t) + P_{WT}(t) = L_e(t) + P_{EC}(t) + P_{BC}(t) \quad (18)$$

式中: $P_g(t)$ 、 $P_{PV}(t)$ 、 $P_{WT}(t)$ 分别为 t 时段电网流入微能源网的电功率、光伏发电功率、风机发电功率, kW; $L_e(t)$ 为电负荷, kW。

2) 热力总线能量平衡方程。

$$Q_{SB}(t) + Q_{HR}(t) = L_h(t) / \eta_{HX} \quad (19)$$

式中 $L_h(t)$ 为 t 时段的热负荷, kW。

3) 冷却总线能量平衡方程。

$$P_{EC}(t) = L_c(t) / C_{EC} \quad (20)$$

式中 $L_c(t)$ 为 t 时段的冷负荷, kW。

结合以上, 可以看出微能源网的特点:

1) 微能源网具有多种类型能源的输入和输出, 涉及多种能量转换单元。在能量输入和消费环节, 同时存在波动性和随机性。加之各个环节响应时间差异, 使得微能源网中的能源输入、转换、分配和消费环节存在不同空间层面和时间尺度的不平衡, 给微能源网的能量管理和优化带来了巨大的挑战。

2) 在该微能源网模型中, 输入的能源为天然气、电能和可再生能源, 输出的能源为电能、热能和冷能。微能源网的能量转换装置使得能量在转换过程中耦合。例如, 联供单元将天然气中的化学能转换为电能和热能的过程中, 无法控制电能与热能的输出比例, 造成电能或热能的输出与当前需求不符, 加剧了能源在时间层面的不平衡。通过设置储能装置, 可一定程度改善这个问题, 但其同时带来的能量流动的多向性增加了微能源网能量管理的难度。人工智能算法的融入为如何有效实现微能源网能量管理智能化提供了一种新的有效途径。

2 深度强化学习

2.1 强化学习及 Q 学习算法

强化学习是指从环境状态到动作策略集的学习, 以使得动作从环境中获得的累计奖励最大, 学习的过程, 如图 3 所示。本质上, 智能体与环境的互动是一种马尔科夫决策过程(markov decision process, MDP)。MDP 一般由一个四元组(S, A, R, π)定义, 其中:

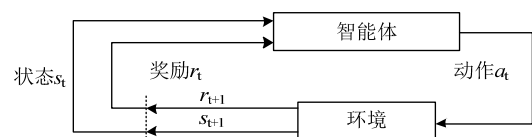


图3 马尔科夫决策过程

Fig. 3 Markov decision process

1) S 为所有环境状态的集合, $s_t \in S$ 表示智能体在 t 时段所处环境中感知到的状态。

2) A 为智能体所有可执行动作的集合, $a_t \in A$ 表示智能体在 t 时段采取的动作。

3) R 为奖励函数, $r_t \in R(s_t, a_t)$ 表示智能体在状态 s_t 执行动作 a_t 获得的即时奖励。

4) π 为智能体的策略集, 表示状态空间 S 到动作空间 A 的映射。

Q 学习^[21](Q-learning)是一种不基于环境模型、基于价值的强化学习算法。Q 学习的主要思路是定义状态动作价值函数, 即 Q 函数, 将观测数据代入到以下更新公式中对 Q 函数进行迭代学习。

$$Q_{t+1}(s_t, a_t) = Q_t(s_t, a_t) + \alpha[r_{t+1} + \gamma \max_{a'} Q(s_{t+1}, a') - Q_t(s_t, a_t)] \quad (21)$$

式中： Q 是状态动作价值函数，指在当前状态 s 下对环境采取动作 a 的价值； t 是当前时段； α 是学习率； γ 是未来的 Q 值在当前的衰减率； a' 是状态 s_{t+1} 可以执行的动作。该算法根据智能体与环境互动之后得到的下一个状态信息与奖励来逼近最优 Q 函数。 Q 学习被证明满足一定条件时可以在时间趋近于无穷时得到最优控制策略^[22]。

2.2 深度 Q 网络

在传统的 Q 学习算法中，由于维数灾难，强化学习难以解决大规模 MDP 问题或连续空间 MDP 问题。因此，价值函数近似方法被提出以解决这一问题。通过构造由参数 ω 构成的函数 \hat{Q} 对价值函数 Q 进行近似，即

$$\hat{Q}(s, a, \omega) \approx Q(s, a) \quad (22)$$

随着深度学习的发展，深度神经网络成为强化学习中最常用的函数近似方法之一。于是在 Q 学习算法和深度神经网络的基础上发展出了深度强化学习代表性的算法，深度 Q 网络(deep Q network, DQN)^[14]。DQN 相较于 Q 学习主要是把 Q 函数通过价值函数近似方法转换为一个深度神经网络，如图 4 所示。

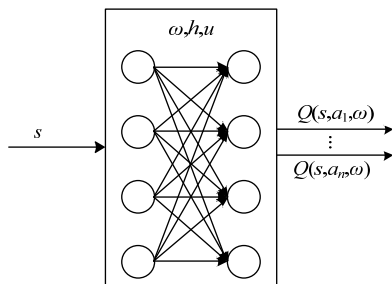


图 4 神经网络价值函数近似方法

Fig. 4 Neural network value function approximation

将深度神经网络应用 Q 学习之后，DQN 为了降低样本之间的相关性，提升训练的稳定性，引入了经验回放机制和冻结参数机制，如图 5 所示。

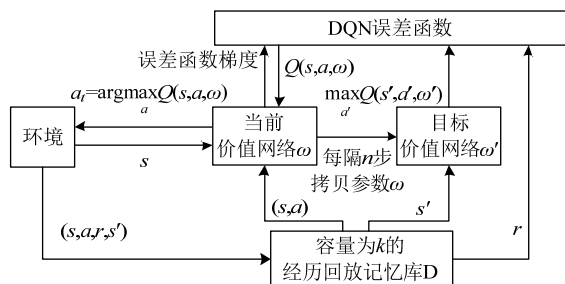


图 5 带有经验回放和冻结参数的 DQN 训练过程

Fig. 5 DQN training process with experience reply and fixed network

3 基于深度强化学习的微能源网能量管理方法

3.1 状态空间

对于微能源网模型，环境提供给智能体的信息一般为可再生能源出力、分时电价、负荷和电储能的荷电状态。因此微能源网模型的状态空间定义为

$$\mathbf{S} = [P_{RE,i}(t), L_i(t), \Gamma(t), \Phi(t)] \quad (23)$$

式中： $P_{RE,i}(t)$ 为可再生能源 i 在 t 时段的功率输出，kW； $L_i(t)$ 为负荷 i 在 t 时段的负荷需求，kW； $\Gamma(t)$ 为 t 时段的分时电价(time of use price, TOU)，元/(kW·h)； $\Phi(t)$ 为 t 时段电储能的荷电状态。

3.2 动作空间

智能体观测到环境的状态信息后，根据自身策略集 π 在动作空间 \mathbf{A} 中选择一个动作。微能源网模型中的设备较为复杂，可以分为能量转换设备和储能设备。对于能量转换设备 i ，其动作是 t 时段设备的出力水平 $P_i(t)$ ，kW；对于储能设备 i ，其动作是 t 时段设备的能量输出或输入量 $B_i(t)$ ，kW。因此微能源网的动作空间为

$$\mathbf{A} = [P_i(t), B_i(t)] \quad (24)$$

由于无法处理连续动作，对于 Q 学习算法，需要对动作空间进行离散化，因此离散化的动作空间为

$$\mathbf{A}_D = [P_i(t), B_i(t)|g] \quad (25)$$

式中： g 为离散化的粒度，粒度越大状态空间内包含的状态越少，粒度越小状态空间可以描述的状态越多。

3.3 奖励函数

在微能源网中，运行成本主要来自于向电网购电的成本和购买天然气的成本，设备启停成本和维护成本可以忽略不计，奖励函数只考虑购电成本和购气成本。以降低微能源网运行成本为目的，日前经济调度的运行成本为

$$C = C_e + C_f \quad (26)$$

式中： C_e 为微网向电网购电的成本，元； C_f 为购买天然气的成本，元。购电成本由下式定义

$$C_e = \sum_{i=1}^T \Gamma(t) P_{g,i} \Delta t \quad (27)$$

式中： T 为日前调度的时间段数； $P_{g,i}$ 为第 i 时间段内的平均购电功率，kW。购入天然气成本为

$$C_f = c_f \sum_{i=1}^T (F_{MT,i} + F_{SB,i}) \quad (28)$$

式中： c_f 为天然气的单位热量价格，元/(kW·h)； $F_{MT,i}$ 和 $F_{SB,i}$ 分别为 i 时段的供应给联供单元和燃气锅炉

的天然气的燃烧热, kW·h。

为了在迭代过程中使各变量满足电力总线的能量平衡和储能约束, 定义惩罚函数为

$$D=D_p+D_b=c_p\sum_{i=1}^T(|d_{p,i}|+|d_{b,i}|) \quad (29)$$

式中: D_p 和 D_b 分别为电力总线不平衡电量的惩罚和电池过放或过充电量的惩罚, 元; c_p 为单位差额电量的惩罚, 元/(kW·h); $d_{p,i}$ 和 $d_{b,i}$ 分别是 i 时段电力总线不平衡电量和电池过放或过充电量, kW·h。

由以上成本及惩罚公式, 定义强化学习的奖励函数为

$$R=-(C+D) \quad (30)$$

4 实验验证与分析

4.1 场景描述

本文采用的微能源网系统结构如图 2 所示。各设备简称及工作参数如表 1 所示。

表 1 各设备简称及工作参数 Tab. 1 Abbreviations and paramaters of devices			
设备	简称	参数	数值
联供发电单元	MT	发电效率 η_{MT}	0.3
余热回收锅炉	HR	换热效率 η_{HR}	0.73
换热装置	HX	换热效率 η_{HX}	0.9
燃气锅炉	SB	效率 η_{SB}	0.9
电制冷机	EC	性能系数 η_{EC}	4
电储能	B/BD/BC	最大充电率 η_{BC}	0.2
		最大放电率 η_{BD}	0.4
		最大储能系数 $\eta_{B,max}$	0.9
		最小储能系数 $\eta_{B,min}$	0.2
风机	WT	最大功率 $P_{WT,max}$	30
光伏	PV	最大功率 $P_{PV,max}$	35

其中, 在微能源网系统的热力部分, 外购的天然供气供应联供发电单元运转, 最大发电功率 60 kW。联供单元排出的高温烟气通向最大输出功率 120 kW 的余热回收装置。余热回收装置与最大功率 100 kW 的燃气锅炉共同向蒸汽热水换热装置供热, 以满足热负荷。电力总线上接入电网、光伏、风机和储能容量为 200 kW·h 的蓄电池, 同时电能供应电负荷和最大功率 300 kW 的电制冷机。该系统运行过程中, PV 与 WT 将按照实时最大发电功率出力, 并且系统与大电网间单向传输, 仅微能源网从电网购电。本文取天然气价格 3.45 元/m³, 折合为单位热值为 0.349 元/(kW·h)。设置储能设备的初始荷电状态为 20%。图 6 为一典型日微能源网的预测日负荷、预测可再生能源出力和分时电价曲线。图 6 中: LD_e、LD_h、LD_c 分别为电负荷、热负荷和冷负荷; PV、WT 分别为光伏和风机的出力。

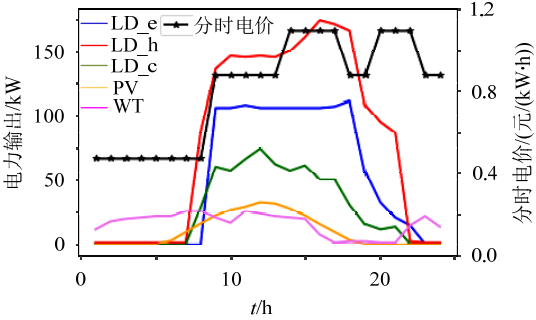


图 6 微能源网负荷和可再生能源预测曲线(场景 1)

Fig. 6 Predicted curve of loads, renewable energy output (scene 1)

4.2 微能源网状态空间及动作空间

由图 2 的微能源网结构及式(23), 可以得出该微能源网的状态空间为

$$S=[P_{PV}(t), P_{WT}(t), L_e(t), L_h(t), L_c(t), \Gamma(t), \Phi(t)] \quad (31)$$

式中 $P_{PV}(t)$ 、 $P_{WT}(t)$ 分别为光伏和风机在 t 时段的功率输出, kW。

本算例中可再生能源出力、负荷及分时电价都基于图 6 中的预测数据, 因此式(31)状态空间中的前 6 个环境状态变量为时段 t 的相关变量, 是可以使用时段 t 进行统一表征静态量。而 $\Phi(t)$ 与该变量在前一时段的状态以及电池的充放电功率有关, 不能单纯使用时段 t 表示。则状态空间可简化为使用时段 t 与当前时段电池的荷电状态 $\Phi(t)$ 2 个变量构成

$$S=[t, \Phi(t)] \quad (32)$$

微能源网中能量转换设备较多, 每台设备的出力共同构成了微能源网的动作空间:

$$A=[P_{MT}(t), P_B(t), P_g(t), P_{EC}(t), Q_{HR}(t), Q_{SB}(t)] \quad (33)$$

动作空间 A 较为复杂, 需要对其进行分析来简化变量以减轻计算负荷。 $P_{MT}(t)$ 与 $Q_{HR}(t)$ 存在耦合关系, 因此 $Q_{HR}(t)$ 可以由 $P_{MT}(t)$ 进行表征, 无须在 A 中体现。由式(19)的热力总线平衡能量, 在热负荷 $L_h(t)$ 与余热回收装置输出热量 $Q_{HR}(t)$ 确定时, 即可得出燃气锅炉的输出 $Q_{SB}(t)$ 。电制冷机的输出 $Q_{EC}(t)$ 取决于冷负荷 $L_c(t)$, 因此电制冷所消耗的电功率 $P_{EC}(t)$ 也仅取决于环境变量中的冷负荷, 无需经过算法决策。在观测到环境状态, 即式(31)中各状态变量已知, 式(18)的电力总线平衡中仅存在 $P_{MT}(t)$ 、 $P_{BC}(t)$ 、 $P_{BD}(t)$ 、 $P_g(t)$ 4 个变量未知, 且 $P_{BC}(t)$ 、 $P_{BD}(t)$ 分别为蓄电池的充放电功率, 本质上可以看成是一个变量, 此外 3 个变量, 在其中 2 个变量确定的情况下第 3 个变量也可通过守恒计算得到。因此 A 可以简化为由联供单元功率 $P_{MT}(t)$ 与电网购电量 $P_g(t)$ 构成, 即

$$\mathcal{A}=[P_{\text{MT}}(t),P_{\text{g}}(t)]$$

(34)

状态空间与动作空间确定之后,即可确定 DQN 中神经网络的输入输出变量的维数。该神经网络的输入为微能源网的环境信息 \mathcal{S} , 输出各动作对应的动作价值函数 $Q(s,a,\omega)$, 具有 h 层全连接隐层, 每层隐层由 u 个神经元构成, 神经元的激活函数选择线性修正函数(rectified linear unit, ReLU)

$$f(x)=\max(0,x)$$

(35)

4.3 仿真结果及分析

4.3.1 能量调度结果

本文使用 6 种不同超参数设置的 DQN 算法、2 种不同粒度设置的 Q 学习算法及遗传算法, 一共 9 种不同参数的算法进行学习和优化。对于强化学习算法, 神经网络结构以外的其他超参数都取相同值, 学习率 $\alpha=0.9$, 衰减率 $\gamma=0.9$, 迭代步数 50 000 步。表 2 展示了各算法的超参数设置及相应的结果, $h=5$ 、 $u=50$ 的 DQN(表 2 中序号 4)取得了最低的日运行成本, 同时序号 2、3 和 7 的算法也有较好的表现。强化学习算法展现良好的后效性, 算法会考虑当前操作对未来产生的影响, 不会贪婪地追求静态时间断面上的最优操作, 而是把全时间段内的最佳结果作为目标, 通过牺牲目前的部分收益以换取未来的更大收益。而作为启发式算法的代表, 遗传算法不具有后效性, 无法考虑当前操作对未来产生的影响, 优化结果较差。

表 2 不同算法参数设置及学习结果比较(1)				
Tab. 2 Learning results of different algorithms(1)				
序号	算法	超参数设置	最低日运行成本/元	策略集占用空间/MB
1	DQN	$h=1,u=1891$	1574	—
2		$h=2,u=50$	1519	—
3		$h=2,u=100$	1518	—
4		$h=5,u=50$	1494	—
5		$h=5,u=200$	1556	—
6		$h=5,u=500$	1605	—
7	Q 学习	$g=10$	1516	0.46
8		$g=3$	1611	5.63
9	GA	—	1622	—

图 7 展示了 5 个日前能量调度结果, 其中, C_{ost} 为该调度计划的成本, LD 为电负荷, GP 为电网输入电功率, 其他缩写含义见表 1。由图 7 可知, 遗传算法对联供系统和蓄电池的利用率都较低。联供系统开启时有一半的时间处于较低的运转负荷, 而蓄电池则较长时间处于荷电状态的最低限值。由强化学习算法得出的各运行方案的调度结果则有以下共同特征:

1) 电储能能在低负荷的谷电时段($t=0\sim6\text{h}$)充分利用可再生能源和谷价电能充电, 且在高负荷的峰电

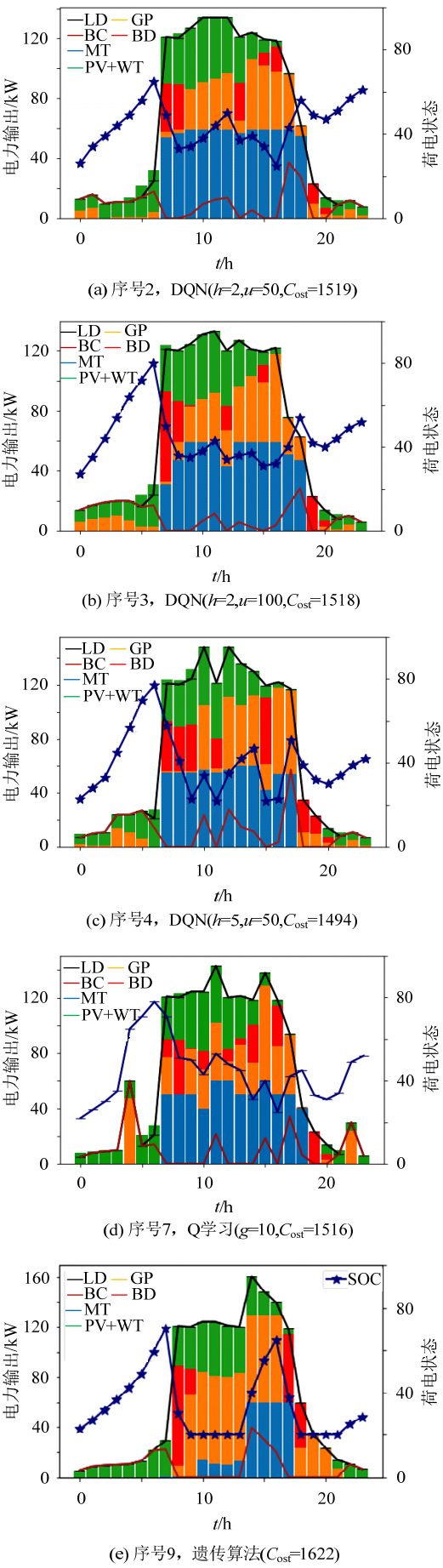


图 7 电力总线能量调度结果

Fig. 7 Dispatch results of the electric power bus

价时段放电,发挥了削峰填谷的作用。

2) 联供单元仅在高负荷时段($t=7\sim 18\text{h}$)开启,同时供应电负荷与热负荷,且负载率均在50%以上,保持较高的运行效率。

3) 电储能选择在高负荷的平电价时段($t=17\sim 18\text{h}$)进行充电,有效满足之后时间段电负荷。

由表2和图7可知, Q 学习算法与具有不同神经网络结构的DQN都可以对微能源网进行有效调度,满足其负荷需求。但在相同的迭代步数下,各算法最优结果的奖励值、收敛速度及占用内存空间不同。

进一步,从表2序号7和8的结果,应用 Q 学习算法易导致维数灾难,同时造成收敛速度慢等后果。对于DQN算法,由于策略集以深度神经网络的形式储存,策略集占用的内存空间可以忽略不计,可见DQN算法在空间复杂度上较 Q 学习算法有绝对的优势。

4.3.2 收敛过程比较与分析

神经网络的结构(即超参数的设置)对策略集的收敛速度及寻优影响较大。图8和图9展示了传统 Q 学习算法和不同结构的DQN算法在学习过程中最佳动作序列奖励和策略集平均奖励的收敛过程。最佳动作序列指当前策略集中日运行成本最低的

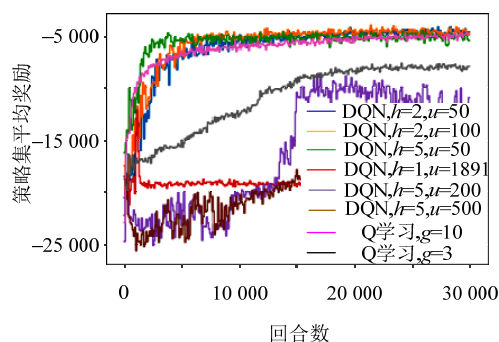


图8 不同超参数DQN及Q学习最佳动作序列学习过程
Fig. 8 Learning process of optimal episode with different DQN and Q-learning hyper-parameters

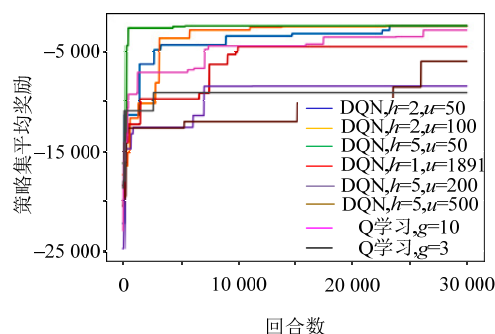


图9 不同超参数DQN及Q学习平均表现学习过程
Fig. 9 Learning process of strategy set with different DQN and Q-learning hyper-parameters

调度结果,代表当前策略集中的最优解;策略集平均奖励指策略集中所有动作序列的平均运行成本,代表整体策略集的平均表现。

由图8和图9可知,超参数设置较为合理(如 $h=5$, $u=50$)的DQN算法较传统 Q 学习算法在最优动作序列的收敛速度、最优动作序列的表现及策略集的收敛速度上都更好的表现。对过于复杂的深度神经网络(如 $h=5$, $u=200$),在迭代初期,记忆库中没有足够的训练样本学习,寻优慢,策略集难收敛;迭代后期($\text{Episode} > 15\,000$),记忆库积累了足够表现优良的训练样本,策略集表现出现较大的提升,但随后又陷入不稳定的振荡过程。粒度较小的 Q 学习算法($g=3$)的 Q 表规模较大,收敛极慢。

综合以上分析,对于DQN算法在微能源网能量管理中的应用,需要选择合适的神经网络结构以得到良好的收敛特性。结果显示,超参数设置为 $h=5$, $u=50$ 的DQN有最好的收敛特性和寻优结果,该结构的DQN算法适用于当前微能源网的能量管理。

4.3.3 策略集继承与智能微能源网

在微能源网中应用深度强化学习算法的意义不仅限于进行能量管理的寻优,更重要的是利用深度强化学习的泛化感知能力和智能决策能力对微能源网进行能量管理。借鉴迁移学习的思想^[23],通过已知场景训练得到一个策略集,再利用深度强化学习的泛化能力,同一微能源网的不同场景具有很高的相似性,利用继承已有策略集的DQN算法对新场景的进行优化。定义图6中的微网场景为S1,新场景为S2, S2的负荷、可再生能源出力和分时电价数据如图10所示。

选取超参数设置为 $h=5$, $u=50$ 的DQN算法分别在策略集完全初始化的情况下对场景S1和S2进行迭代,得到结果R1和R2。用相同的算法,在R1策略集的基础上对S2进行学习得到R3,学习过程如图11所示。由图11可知,继承了R1策略集

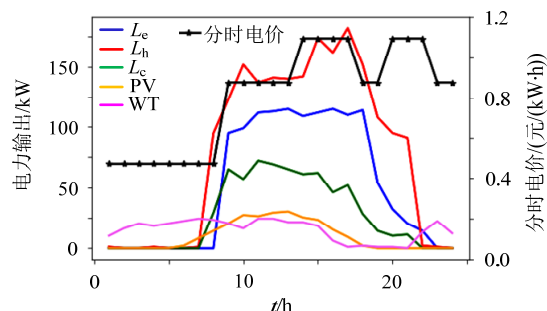


图10 微能源网负荷和可再生能源预测曲线(场景2)
Fig. 10 Predicted curve of loads, renewable energy output(scene 2)

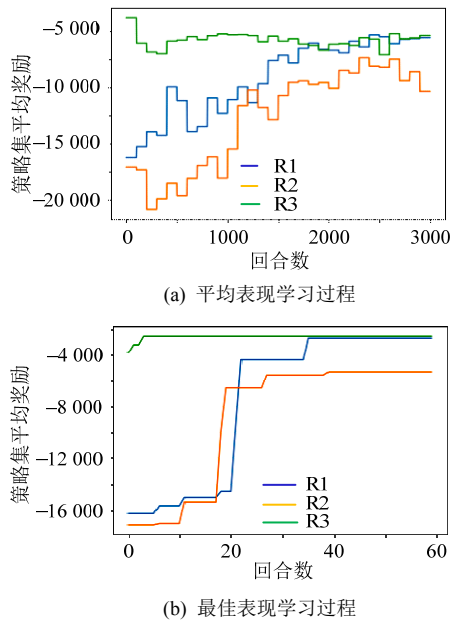


图 11 继承策略集和初始化策略集的 DQN 算法学习过程
Fig. 11 Learning process of DQN with inherited strategy set and initialized strategy set

的 R3 迭代过程明显优于基于初始化策略集的 R2，具有更快的收敛性。

综合以上分析，强化学习算法能有效应用在微能源网能量管理领域。在应用过程中，深度强化学习在空间复杂度和收敛速度上较传统强化学习算法有较大的优势。借助迁移学习思想，从继承策略集的深度学习算法的表现可以看出，相较于基于初始化策略集的深度强化学习，继承策略集的深度学习算法在不同场景的微能源网能量管理中具有更快的收敛速度和更好的收敛结果。

4.3.4 DQN 的在线优化

通过前述的结论，可推论出继承已训练策略集的 DQN 能在收敛速度及优化结果上具有较好的效果。为了验证此结论，本节对遗传算法和经过训练的 DQN 算法在长时间尺度的多能微网环境中的在线优化表现进行对比。算法的优化结果如图 12 和表 3 所示，EH 代表电制热设备。由优化可知，相比于遗传算法，DQN 的优化结果具有以下特点：

- 1) 电储能能在低负荷的谷电时段利用可再生能源和谷价电能充电。而遗传算法较少使用谷价电能进行充电，削峰填谷效果较差。这体现了深度强化学习算法的后效性，这点在长时间尺度场景的在线优化中更加凸显。
- 2) DQN 的结果中，联供单元开启时负载率较高(高于 50%)，可以使燃气轮机保持较高的运行效率。
- 3) DQN 在总运行成本和计算耗时上的表现均优于遗传算法，计算耗时方面的优势尤为明显。
- 4) 对储能的利用程度两者相近，但在 DQN 的

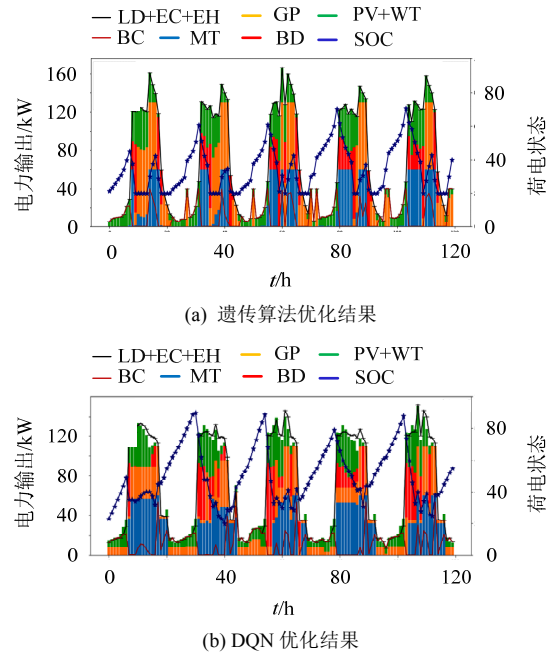


图 12 遗传算法和 DQN 的微能源网优化结果
Fig. 12 Optimization for multi-energy grid of DQN and GA

表 3 不同算法参数设置及学习结果比较(2) Tab. 3 Optimizaiton results of DQN and GA(2)			
算法	5 日运行 总成本/元	耗时/ min	储能等效 充放电循环/次
GA	7738	36	3.26
DQN($h=5, u=50$)	7436	8	3.20

优化结果中，储能的等效充放电循环次数略小于遗传算法，并且在 SOC 始终保持在中高水平，总体上优于遗传算法。

5 结论

微能源网是最有前景的能源利用形式之一。微能源网的能量管理与运行优化是实现能源高效灵活利用的前提。微能源网的多能流耦合等特点使其能量管理面临诸多挑战。本文在针对微能源网的能量管理问题，以经济性为目标，完成了微能源网的能量管理与优化。基于能量总线的微能源网系统模型，利用 DQN 算法，提出基于深度强化学习的微能源网能量管理方法。通过场景仿真验证了算法的可行性和有效性。在此基础上，本文使用不同参数的深度强化学习算法进行场景优化，完成了优化结果和迭代过程的比较。从比较可以看出，对特定结构的微能源网，需要找到合适的神经网络参数才能使算法达到较快的迭代过程和较好的寻优结果。此外，本文通过对比分别基于继承策略集和初始化策略集的 DQN 算法在不同微能源网能量管理场景中的表现，证明通过继承策略集使微能源网具备“记忆”和决策能力能有效提升算法的收敛速度和优化结果。此外，本文还对比了遗传算法与 DQN

算法在长时间尺度多能微网环境的在线优化任务中的表现,证明 DQN 在优化结果和优化速度上的优势,对复杂的微能源能量管理与优化方法提供了有益参考。

参考文献

- [1] 马腾飞, 吴俊勇, 郝亮亮. 含冷热电三联供的微能源网能量流计算及综合仿真[J]. 电力系统自动化, 2016, 40(23): 22-27.
Ma Tengfei, Wu Junyong, Hao Liangliang, et al. Energy flow calculation and integrated simulation of micro-energy grid with combined cooling, heating and power[J]. Automation of Electric Power Systems, 2016, 40(23): 22-27(in Chinese).
- [2] 张涛, 张福兴, 张彦. 面向能源互联网的能量管理系统研究[J]. 电网技术, 2016, 40(1): 146-155.
Zhang Tao, Zhang Fuxing, Zhang Yan. Study on energy management system of energy internet[J]. Power System Technology, 2016, 40(1): 146-155(in Chinese).
- [3] Olivares D E, Canizares C A, Kazerani M. A centralized energy management system for isolated microgrids[J]. IEEE Transactions on Smart Grid, 2017, 5(4): 1864-1875.
- [4] Li Y, Yang Z, Li G, et al. Optimal scheduling of an isolated microgrid with battery storage considering load and renewable generation uncertainties[J]. IEEE Transactions on Industrial Electronics, 2019, 66(2): 1565-1575.
- [5] Li Y, Wang C, Li G, et al. Improving operational flexibility of integrated energy system with uncertain renewable generations considering thermal inertia of buildings[J]. Energy Conversion and Management, 2020(207): 112526.
- [6] Askarzadeh A. A memory-based genetic algorithm for optimization of power generation in a microgrid[J]. IEEE Transactions on Sustainable Energy, 2017, PP(99): 1-1.
- [7] 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.
Wu Dinghui, Gao Cong, Ji Zhicheng. Economic optimization operation of the microgrid using the hybrid particle swarm optimization algorithm[J]. Control Theory & Applications, 2018, 35(4): 457-467(in Chinese).
- [8] 曾君, 徐冬冬, 刘俊峰, 等. 考虑负荷满意度的微电网运行多目标优化方法研究[J]. 中国电机工程学报, 2016, 36(12): 3325-3333.
Zeng Jun, Xu Dongdong, Liu Junfeng, et al. Multi-objective optimal operation of microgrid considering dynamic loads[J]. Proceedings of the CSEE, 2016, 36(12): 3325-3333(in Chinese).
- [9] 王鲁浩, 李歧强, 王桂荣, 等. 多可再生能源冷热电联供微网系统环境经济优化调度[J]. 控制与决策, 2016, 31(5): 913-918.
Wang Luhao, Li Qiqiang, Wang Guirong, et al. Optimal environmental economic scheduling of combined cooling heating and power microgrid system with multi-renewable energy[J]. Control and Decision, 2016, 31(5): 913-918(in Chinese).
- [10] 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(in Chinese).
- [11] Tang Y, He H, Wen J, et al. Power system stability control for a wind farm based on adaptive dynamic programming[J]. IEEE Transactions on Smart Grid, 2015, 6(1): 166-177.
- [12] Ruelens F, Claessens B, Vandaal S, et al. Residential demand response of thermostatically controlled loads using batch Reinforcement Learning[J]. IEEE Transactions on Smart Grid, 2017, 8(5):

2149-2159.

- [13] Foruzan E, Soh L K, Asgarpour S. Reinforcement learning approach for optimal distributed energy management in a microgrid[J]. IEEE Transactions on Power Systems, 2018, 33(5): 5749-5758.
- [14] Mnih V, Kavukcuoglu K, Silver D, et al. Human-level control through deep reinforcement learning[J]. Nature, 2015, 518(7540): 529.
- [15] Silver D, Huang A, Maddison C J, et al. Mastering the game of Go with deep neural networks and tree search[J]. Nature, 2016, 529(7587): 484-489.
- [16] 唐振韬, 邵坤, 赵冬斌, 等. 深度强化学习进展: 从 AlphaGo 到 AlphaGo Zero[J]. 控制理论与应用, 2017, 34(12): 1529-1546.
Tang Zhenhao, Shao Kun, Zhao Dongbin, et al. Recent progress of deep reinforcement learning: from Alphago to AlphaGo Zero [J]. Control Theory & Applications, 2017, 34(12): 1529-1546(in Chinese).
- [17] 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).
- [18] 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).
- [19] François-lavet V, Taralla D, Ernst D, et al. Deep reinforcement learning solutions for energy microgrids management[C]//European Workshop on Reinforcement Learning(EWRL 2016). Barcelona, Spain: Pompeu Fabra University, 2016: 1-7.
- [20] Haochen H, Yuchao Q, et al. Optimal energy management strategies for energy internet via deep reinforcement learning approach [J]. Applied Energy, 2019(239): 598-609.
- [21] Watkins C J C H. Learning from delayed rewards[J]. Robotics & Autonomous Systems, 1989, 15(4): 233-235.
- [22] Jaakkola T, Jordan M I, Singh S P. On the convergence of stochastic iterative dynamic programming algorithms[J]. Neural Computation, 1993, 6(6): 1185-1201.
- [23] Pan S J, Yang Q. A survey on transfer learning[J]. IEEE Transactions on Knowledge & Data Engineering, 2010, 22(10): 1345-1359.



刘俊峰

收稿日期: 2020-03-18。

作者简介:

刘俊峰(1978), 男, 博士, 副教授, 研究方向为高频微电网关键技术、电力电子及其控制技术, E-mail: aujfliu@scut.edu.cn;

陈剑龙(1994), 男, 硕士研究生, 研究方向为深度强化学习在微电网中的应用, E-mail: chenjl94@foxmail.com;

王晓生(1997), 男, 硕士研究生, 研究方向为机器学习在微电网中的应用, E-mail: xiaosheng_1997@163.com;

曾君(1979), 女, 博士, 通信作者, 教授, 研究方向为微电网能量管理及优化、可再生能源发电系统中的电力电子及控制技术, E-mail: junzeng@scut.edu.cn;

黄倩颖(1995), 女, 硕士研究生, 研究方向为微电网优化调度及能量管理, E-mail: 494579905@qq.com。

(责任编辑 徐梅)