

基于纳什竞争深度Q网络的综合能源微网电-热-碳联合调度策略

黎海涛¹, 乔禄¹, 谢冬雪¹, 杨艳红²

(1. 北京工业大学 信息科学技术学院, 北京 100124;

2. 中国科学院电工研究所, 北京 100190)

摘要:随着各类可再生能源发电设备以微网形式大规模接入主网,微网间的协调调度运行愈加重要,而现有智能体强化学习在处理调度问题时存在维度灾难、训练效率低的不足。为此,提出基于纳什竞争深度Q网络学习算法的微网联合调度策略。构建计及电、热、碳交易的综合能源微网能量调度模型,在碳排放管理方面采用碳捕集装置和阶梯式碳排放交易机制;分析微网间的博弈关系,并提出基于纳什竞争深度Q网络的电-热-碳联合调度策略。算例仿真结果表明,所提调度策略能够提高微网的经济效益并有效降低碳排放量,实现了微网间能量的时空互补和利益均衡。

关键词:微网;调度;碳捕集;多智能体博弈;纳什均衡;深度强化学习;阶梯碳交易;综合能源系统

中图分类号:TM73

文献标志码:A

DOI:10.16081/j.epae.202601001

0 引言

作为实现碳达峰、碳中和目标的有效途径,分布式可再生能源在近年获得快速发展^[1-4]。各类可再生能源发电设备以微网形式接入主网已成为新型电力系统的发展趋势^[5-7]。本文所研究的微网是包含电、热、氢和碳交易等多种元素的综合能源微网系统(下文简称微网)。随着微网规模的不断扩大,多个微网间的能量互补和协调运行也愈加重要,这不仅有助于提高可再生能源的利用率,还可减小分布式能源波动对主网的影响。针对多微网系统中电、热等能源资源的管理调度:文献[8]构建以运行成本和环境成本为目标的微网群协同优化控制模型,采用基于遗传算法的多目标优化算法进行求解,有效解决了分布式发电微电网运行中的能源管理问题,提高了微电网的经济收益;文献[9]以微网系统运行成本最低为目标函数,提出基于非精确广义不定邻近交替方向乘子法的多微网系统分布式协调优化方案,在保证系统功率平衡的前提下提高了可再生能源的消纳率。文献[10]提出计及分布式水风光时空相关性的多微网协同优化策略,进一步提升了新能源的互补消纳水平。在分布式调度中,各主体间的信息交互带来了数据隐私泄露的风险。文献[11]提出基于一致性交替方向乘子法的分布式鲁棒

优化框架,以增强各运营主体的隐私保护。文献[12]采用差分隐私技术对交互信息进行加密扰动,以保护各园区的隐私。

若多微网系统中各个智能体的目标相互对立,则可能不存在一个明确的最优解,此时能量管理策略通常会寻求均衡。博弈论作为研究不同决策者之间存在利益冲突时合理决策的重要理论,在解决多微网能量管理问题中得到应用。文献[13]提出一种基于纳什均衡理论和 Stackelberg 博弈的社区微网分布式能源协调调度方法,能提升各博弈参与方的收益,更合理地调度系统中的资源。文献[14]基于随机博弈模型和纳什均衡理论,研究微网能量交易问题,通过微网运营商收益函数中的条件风险度量来表征微网收益,采用样本平均近似技术来逼近随机纳什均衡,有效提高了微网经济效益。文献[15]考虑多微网功能区的差异,基于纳什谈判准则,研究计及电、热、碳能源交易的多微网优化调度策略,采用交替方向乘子法求解得到纳什均衡解,提高了系统运行的低碳性与经济性。文献[16]将纳什均衡策略应用于光-储耦合制氢系统,实现了多目标优化控制。文献[17]基于古诺博弈模型,研究综合能源供应商在电-气-碳-绿证多市场中的联合均衡问题。文献[18]考虑氢能系统,构建配电网运营商与多综合能源服务商之间的混合博弈模型。文献[19]考虑主从与合作博弈的混合博弈模型,并引入日前-日内两阶段调度框架。为促进系统低碳运行:文献[20]研究碳捕集、电转气等技术与热电联产的耦合运行模式;文献[21]则构建包含碳捕集电厂的协同优化模型,验证了阶梯式碳交易机制在碳排放上的有效性。为挖掘清洁能源的环境价值,针对电-氢综

收稿日期:2025-03-07;修回日期:2025-12-05

在线出版日期:2026-01-07

基金项目:国家重点研发计划项目(2024YFE0115700);国家自然科学基金资助项目(52277131)

Project supported by the National Key Research and Development Program of China(2024YFE0115700) and the National Natural Science Foundation of China(52277131)

合能源系统,文献[22]提出氢能绿证交易等市场机制来激励绿氢生产。文献[23]提出基于电-氢-热点对点(point to point,P2P)交易的分布式智能电网低碳优化运行方法,实现了氢能多元利用与多能协同共享。

然而,在多主体环境下,采用传统方法求解纳什均衡的计算复杂度较高,甚至难以获得最优解,近年来,有学者研究基于多智能体强化学习的求解策略。文献[24]建立以多智能体间利益均衡为目标的联合博弈决策模型,基于纳什均衡理论和 Q 学习算法(记为Nash- Q 学习算法)求解模型,实现了综合能源微网的协同调度。但随着问题复杂性的增加,Nash- Q 学习算法在高维状态空间中面临维度灾难问题,显著增加了算法的计算复杂度和存储需求,严重制约了Nash- Q 学习算法在复杂环境中的应用。为了应对该挑战,文献[25]提出纳什深度 Q 网络(Nash-deep Q network,Nash-DQN)学习算法来求解多微网协同优化问题,有效解决了求解纳什均衡时的维度灾难问题,使各微网的运行成本达到均衡。尽管Nash-DQN在一定程度上解决了维度灾难问题,但该算法在处理高维状态空间时的训练效率较低且收敛速度较慢。此外,由于DQN的非线性特性,Nash-DQN在复杂环境下的策略评估和改进过程中可能会出现策略不稳定、学习不充分等问题。

针对现有多智能体强化学习算法处理微网间博弈均衡的不足,本文研究新的博弈学习算法来设计多微网能源调度策略。首先,构建包含电、热、碳能源交易的微网能量调度模型。然后,提出适用于多智能体博弈场景的纳什竞争深度 Q 网络(Nash-dueling deep Q network,Nash-Dueling DQN)学习算法。该算法在求解多个微网进行博弈的纳什均衡问题时,通过深度神经网络对价值函数进行参数化表示,并对优势函数进行去中心化处理,不仅有效解决了求解纳什均衡时的维度灾难问题,还提升了学习效率和训练的稳定性,克服了Nash-DQN学习算法的不足。仿真实验结果表明,基于Nash-Dueling DQN学习算法的电-热-碳联合调度策略显著提升了微网的经济收益。

1 综合能源微网能量调度模型

1.1 综合能源微网结构

考虑如图1所示的综合能源微网结构,每个微网由风力发电站、光伏发电站、储能电池、氢气罐、电解槽、燃料电池、燃气锅炉、碳捕集装置等组成。在该微网结构中,风能和太阳能作为可再生能源,优先为微网提供能量,最大化就地消纳;储能电池与氢气罐共同组成一个能满足电力和热能需求的多功能混

合储能系统;电解槽将电能转化成氢能,为燃料电池和燃气锅炉的运行提供能源;燃料电池利用氢气产出电能和热能,增强了系统能源应用的灵活性;燃气锅炉负责将氢气或天然气转化为热能,以满足热量需求;碳捕集装置从燃气锅炉的排放中提取 CO_2 ,以降低对环境的影响。该微网结构以其高效的能源互补和转换机制,为实现微网的能源优化和环境可持续性提供了切实可行的解决方案。各物理组件的详细数学模型如附录A式(A1)~(A6)所示。

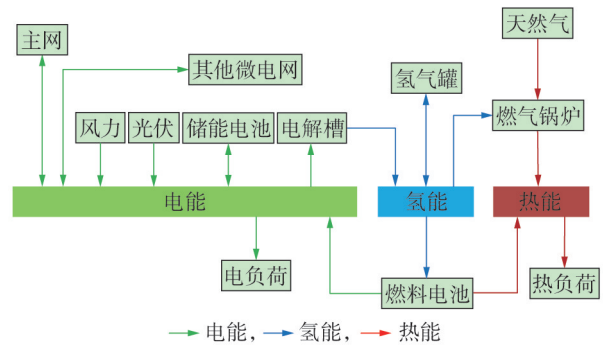


图1 综合能源微网结构

Fig.1 Structure of integrated energy microgrid

1.2 电-热-碳联合调度模型

在微网能量调度过程中,交易机制至关重要。本文考虑电交易、热交易和碳交易在微网能量调度中的作用与影响。

在电交易方面,微网不仅能与主网进行能量交换,还能与其他微网进行P2P交易,实现能源共享和互补。考虑到传输损耗,微网间交易价格通常设定在主网售电价格与购电价格之间。

在热交易方面,微网热能主要由燃气锅炉通过燃烧氢气或天然气提供。其中,天然气燃烧会产生碳排放。

在碳交易方面,为有效促进碳减排,本文采用阶梯式碳交易机制。微网的总碳排放源于向主网购电以及天然气燃烧,而碳捕集装置可以捕获并减少部分碳排放。微网从主网购电时可获得一定数量的无偿碳排放配额,若实际碳排放量超出该配额,则必须在碳市场中购买碳排放配额,且购买价格随着超出量的增加而呈阶梯式上涨。相较于传统碳交易,该机制通过惩罚性的价格杠杆,能更强力地引导微网走向低碳运行模式。

微网电-热-碳联合调度是以最大化自身经济收益为目标。该收益定义为微网的总运行收益减去总运行成本。运行成本主要包括发电设备维护成本、与主网及其他微网的电量交互成本、购买天然气的成本和阶梯式碳排放配额购买成本;运行收益主要包括向本地电、热负荷供能的收益以及将富余电量出售给主网或其他微网的收益。在运行过程中,模

型必须时刻满足电能、热能和氢能的动态平衡约束。微网能量调度模型的核心任务在于对微网内部的能源进行优化分配与管理,从而提高整体经济效益。

本文所构建的微网能量调度模型的具体目标函数,各部分成本、收益及约束条件的数学表达式如附录A式(A7)~(A25)所示。在微网能量调度过程中,必须综合考虑电力、热能、碳排放等多种能源形式的协同作用与动态变化,还需要应对负荷需求的波动、可再生能源的间歇性以及储能设备的充放电策略。这些复杂性使得微网能量调度不仅是一个优化问题,还是一个动态决策问题,因此,微网能量调度模型可以视为马尔可夫决策过程(Markov decision process, MDP),微网作为智能体,通过持续感知系统状态,选择相应的动作来最大化累积收益。

由于微网之间可以进行电量交易,多个微网构成多智能体模型,因此将微网调度问题扩展为多智能体博弈模型^[26-27]。在该模型中,各个微网作为独立的智能体,智能体内部通过调度策略进行资源决策,同时微网之间通过电量交互来协调利益。智能体之间的相互作用形成了一个动态博弈过程,各智能体通过独立决策进行博弈并制定自身的运行策略,使其效益达到最优。因此,不同于传统的统一优化方法,微网多智能体博弈模型不是所有智能体共同追求一个单一的优化目标,而是每个智能体独立优化其自身目标。通过这种独立优化,各智能体在相互作用中不断调整自身策略,最终系统达到一个稳定的状态,即纳什均衡。在该均衡状态下,所有智能体的策略都达到一种平衡,每个智能体在现有策略下无法通过单方面改变自身策略来获得更高的收益。

2 联合调度的MDP博弈分析

2.1 MDP博弈

在微网系统的博弈中,每个微网作为独立的智能体,与环境持续交互,并通过调整策略来解决微网能量调度问题。由于智能体的决策过程可能受到外部环境变化和其他智能体行为的影响,且下一时刻的状态仅依赖于当前状态,这一过程可以视为符合MDP的随机博弈。在微网多智能体模型中,MDP博弈模型能够有效描述各智能体之间的交互与优化过程。一个MDP博弈可描述为 $\{S, \mathbf{a}_1, \dots, \mathbf{a}_n, P, r_1, \dots, r_n\}$ 。其中: S 为所有智能体所处环境的状态空间; $\mathbf{a}_1 - \mathbf{a}_n$ 分别为智能体1~ n 的动作空间; P 为状态转移函数; $r_1 - r_n$ 分别为智能体1~ n 的奖励值。

所有智能体在 t 时刻所处环境的状态空间定义为 $S_t = [s_{1,t}, s_{2,t}, \dots, s_{n,t}]$,其中智能体 i 在 t 时刻的状态

定义为:

$$s_{i,t} = [P_{\text{load}}^{i,t}, P_{\text{total,h}}^{i,t}, P_{\text{W}}^{i,t}, P_{\text{PV}}^{i,t}, K_{\text{down}}^{i,t}, K_{\text{up}}^{i,t}, K_{\text{p2p}}^{i,t}] \quad (1)$$

式中: $P_{\text{load}}^{i,t}$ 为电负荷的电需求量; $P_{\text{total,h}}^{i,t}$ 为热负荷的热需求量; $P_{\text{W}}^{i,t}$ 为风力发电组件的发电量; $P_{\text{PV}}^{i,t}$ 为光伏发电组件的发电量; $K_{\text{down}}^{i,t}$ 为向电网销售单位电量的价格; $K_{\text{up}}^{i,t}$ 为向电网购买单位电量的价格; $K_{\text{p2p}}^{i,t}$ 为微网间进行电量交易的价格。

智能体 i 在 t 时刻的动作 $\mathbf{a}_{i,t}$ 包括电量过剩时的动作选择以及电量短缺时的动作选择,即:

$$\mathbf{a}_{i,t} = [\mathbf{a}_0^{i,t}, \mathbf{a}_1^{i,t}] \quad (2)$$

式中: $\mathbf{a}_0^{i,t}$ 为微网电量过剩时的动作选择; $\mathbf{a}_1^{i,t}$ 为微网电量短缺时的动作选择。当电量过剩时,可采取的动作包括出售电量、储能电池存储电量和使用电解槽。当电量短缺时,可采取的动作包括购买电量、储能电池供电和使用燃料电池。

智能体 i 在 t 时刻所获得的奖励值 $r_{i,t}$ 由其奖励函数计算得到,微网基于奖励值并利用强化学习进行决策和优化,以选择最佳的动作策略,从而实现最优经济效益。奖励值为:

$$r_{i,t} = R_{i,\text{load}}^t + R_{i,\text{sale}}^t + R_{i,\text{p2p}}^t - C_{i,\text{W}}^t - C_{i,\text{PV}}^t - C_{i,\text{purchased}}^t - C_{i,\text{p2p}}^t - C_{i,\text{trans}}^t - C_{i,\text{gas}}^t - C_{i,\text{carbon}}^t \quad (3)$$

式中: $R_{i,\text{load}}^t$ 为向本地负荷供电的收益; $R_{i,\text{sale}}^t$ 、 $R_{i,\text{p2p}}^t$ 分别为向主网和其他微网售电的收益; $C_{i,\text{W}}^t$ 、 $C_{i,\text{PV}}^t$ 为设备维护成本; $C_{i,\text{purchased}}^t$ 为购电成本; $C_{i,\text{p2p}}^t$ 为微网间交互成本; $C_{i,\text{trans}}^t$ 为传输成本; $C_{i,\text{gas}}^t$ 为天然气成本; $C_{i,\text{carbon}}^t$ 为碳交易成本。

2.2 博弈均衡分析

由附录A式(A22)可以看出,微网收益受制于自身和其他微网调度策略选择的影响。微网间主要进行电量交易,当某个微网需要购买电量时,可以从其他电力充足的微网购买以降低购电成本。相较于从主网购买电量,这种购电方式更经济高效,然而售电微网的收益也会下降。微网可采用储能电池来增强供电的可靠性和灵活性,但同时也需综合考虑储能设备的成本和效率。在微网内部,购电及售电决策也会影响燃料电池、燃气锅炉等设备的决策。因此,微网间形成了复杂的博弈关系,如图2所示。图中展示了博弈的过程, K 为博弈的轮次。每个微网需在考虑其他微网行为的基础上制定自身调度策略。这种相互关联的决策过程增加了微网间博弈的复杂性,因此微网需要综合考虑各种因素进行权衡以作出最优决策。

在微网博弈结构中,各主体决策相互影响,增加了系统的动态性和不确定性。微网根据实际情况及时调整策略,以达到博弈的最终目标——纳什均衡。纳什均衡作为博弈论分析框架中的关键理论工具,用于分析参与者之间策略相互作用的情况。在

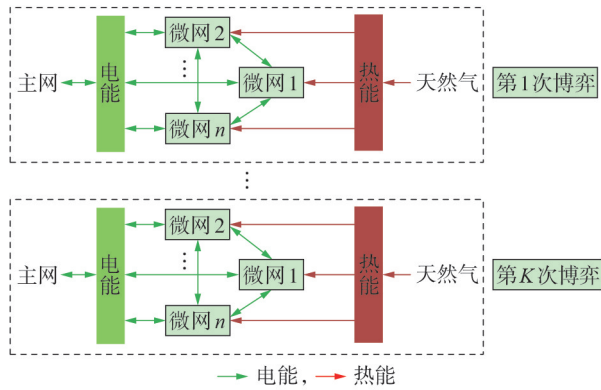


图2 微网间博弈关系

Fig.2 Game relationship among microgrids

持续的时间序列中,微网的相互作用演化为策略博弈,微网通过持续博弈逐渐稳定于纳什均衡。纳什均衡表明了系统的稳定性和策略的相互依赖性,当达到纳什均衡时,每个智能体基于其他智能体的策略作出的决策是最佳的,且智能体之间不存在激励偏差,即任何智能体单独改变策略都不会给自身带来额外的好处^[28-29]。

本文将求解多智能体之间博弈的纳什均衡点作为微网能量调度计划,其目标表达式为 $G = g(N_g, A_g, u_g)$ 。其中: G 为多智能体之间博弈达到的均衡点; $g(\cdot)$ 为多智能体之间的博弈函数; N_g 为参与博弈的智能体数量,本文中表示微网数量; A_g 为参与博弈的智能体行为的集合,本文中表示各微网的动作集合 $\{a_1, a_2, \dots, a_n\}$; u_g 为参与博弈的智能体的效用集合,本文中表示微网的净收益 I 。在微网博弈场景中,如果微网 i 的净收益满足 $I_i(a_i^*, a_{-i}^*) \geq I_i(a_i, a_{-i}^*)$,则称系统达到了纳什均衡。其中 (a_i^*, a_{-i}^*) 为纳什均衡解, a_i^* 为智能体 i 在给定其他智能体动作策略组合时的最佳反应动作, a_{-i}^* 为除微网 i 外其他智能体的动作策略。通过多智能体博弈,微网能够有效管理和分配电-热-碳资源,提高运行效率和可持续性。

3 基于 Nash-Dueling DQN 的综合能源微网联合调度策略

对于马尔可夫决策问题,通常采用 Q 学习算法进行求解,然而,现有方法在处理复杂的微网调度问题时存在一定的局限性。为此,本章提出一种改进的算法,即 Nash-Dueling DQN 算法,以更有效地消除这些局限性,优化微网的调度策略。

3.1 Nash-Q 学习算法

Q 学习算法是被普遍用于求解 MDP 问题的强化学习算法之一,它采用值函数的迭代方式进行在线学习,并依据动态最优原则进行操作。该算法的核心思想是利用之前积累的经验信息来初始化 Q 值

表,可有效缩短算法达到收敛状态所需的时间。 Q 学习算法的值函数及迭代过程分别表示为:

$$Q(s, a) = R(s, s', a) + \gamma \sum_{s' \in S} P(s' | s, a) \max_a Q(s', a) \quad (4)$$

$$Q^{k+1}(s_k, a_k) = Q^k(s_k, a_k) + \tau \left[R(s_k, s_{k+1}, a_k) + \gamma \max_a Q(s_{k+1}, a) - Q^k(s_k, a_k) \right] \quad (5)$$

式中: s, s' 分别为当前状态和下一时刻状态; $Q(s, a)$ 为状态 s 下执行动作 a 的 Q 值; $R(s, s', a)$ 为状态 s 经过动作 a 移到状态 s' 后得到的立即奖励函数值; γ 为折扣因子; $P(s' | s, a)$ 为状态 s 在动作 a 发生后转移到状态 s' 的概率; Q^k 为最优值函数 Q^* 的第 k 次迭代值; τ 为学习因子。

将博弈论与 Q 学习方法相结合,可以实现对多智能体运行策略的学习和优化,进一步提高系统的运行效率和稳定性。Nash- Q 学习算法是一种广泛应用于解决多智能体系统中序贯决策问题的强化学习方法,通过在线学习和动态优化来求解纳什均衡^[30]。Nash- Q 学习算法中状态动作对的价值函数迭代更新公式为:

$$Q_i^{k+1}(s, a_1, \dots, a_n) = (1 - \tau) Q_i^k(s, a_1, \dots, a_n) + \tau [r_i + \gamma Q'_{Nash, i}(s')] \quad (6)$$

$$Q'_{Nash, i}(s') = Q_i(s', a_1^*, \dots, a_n^*) \quad (7)$$

式中: $Q'_{Nash, i}(s')$ 为状态 s' 下智能体 i 的 Nash- Q 值。

3.2 Nash-DQN 学习算法

在多智能体环境中, Nash- Q 学习算法通过 Q 表存储每个状态、动作元组的 Q 值向量。然而,随着环境复杂度的增加, Q 表的大小呈指数级增长,导致时间和空间复杂度急剧上升,造成维度灾难问题。文献[11]借鉴 DQN 的思想,提出 Nash-DQN 算法,即使用神经网络来近似 Q 值函数^[31-32]。通过智能体与环境交互得到表示状态、动作、奖励和下一个状态的四元组,将其存储在经验回放缓冲区中,再从中随机抽取小批量样本进行训练。计算每个样本的目标 Q 值,根据目标 Q 值和当前 Q 值之间的均方误差来定义损失函数,通过梯度下降法最小化损失函数,对 Q 网络的参数进行更新。每隔固定数量的步骤,将 Q 网络的参数复制到目标网络。通过这种方式, Nash-DQN 算法能够有效地利用经验回放和目标网络来稳定训练过程,提高智能体在复杂环境中的学习效果。

虽然 Nash-DQN 算法能够有效解决维度灾难问题,但是在处理高维状态空间时,可能会遇到训练效率低下和收敛速度慢的问题。此外, Nash-DQN 算法在每次更新时只会更新某个动作的 Q 值,而其他动作的 Q 值不变,这可能导致学习不充分。

3.3 Nash-Dueling DQN 学习算法

针对 Nash-DQN 存在的问题,本文提出 Nash-Dueling DQN 算法,以在多微网环境中得到纳什均衡。所提算法将神经网络的输出 Q 值分为独立的状态值和动作值来提高算法鲁棒性^[33],对优势函数进行去中心化处理来提高函数可辨识度与算法稳定性。 Q 值函数为:

$$Q(s, \mathbf{a}; \theta, \alpha, \beta) = V(s; \theta, \beta) + \left[A(s, \mathbf{a}; \theta, \alpha) - \frac{1}{|A|} A(s, \mathbf{a}'; \theta, \alpha) \right] \quad (8)$$

式中: $V(s; \theta, \beta)$ 为仅与状态有关而与选取的动作无关的价值函数, $A(s, \mathbf{a}; \theta, \alpha)$ 为与状态和动作均有关的优势函数, θ 为公共部分的网络参数, β 为价值函数的网络参数, α 为优势函数的网络参数; $\frac{1}{|A|} A(s, \mathbf{a}'; \theta, \alpha)$ 为优势函数的平均值, \mathbf{a}' 为下一时刻的动作。Nash-Dueling DQN 算法的详细学习过程如附录 B 图 B1 所示。

3.4 微网联合调度策略

基于 Nash-Dueling DQN 算法设计的微网联合调度策略流程如下。

1) 初始化神经网络, 构建 2 个具有相同架构的当前网络参数 θ 和目标网络参数 θ' , 分别计算状态的价值和执行特定动作的相对优势。

2) 各微网根据当前状态 s_i 选择动作 \mathbf{a}_i , 并执行该动作, 动作即为微网调度策略, 得到当前奖励 r_i 和下一个状态 s'_i 。

3) 将当前状态 s_i 、动作 \mathbf{a}_i 、当前奖励 r_i 和下一个状态 s'_i 以元组 $(s_i, \mathbf{a}_i, r_i, s'_i)$ 储存在经验池中。

4) 从经验池中随机抽取一批训练样本 $(s_i, \mathbf{a}_i, r_i, s'_i)$, 使用目标网络计算目标 Q 值 y_i , 即:

$$y_i = r_i + \gamma Q_{\text{Nash}}(s'_i, \mathbf{a}_i; \theta', \alpha, \beta)$$

式中: $Q_{\text{Nash}}(\cdot)$ 为基于目标网络参数 θ' 计算的下一时刻纳什均衡状态下的全局 Q 值。

5) 当前网络根据状态 s_i 和动作 \mathbf{a}_i 计算预测 Q 值 $Q(s_i, \mathbf{a}_i; \theta, \alpha, \beta)$ 。

6) 计算损失函数 L 。定义损失函数为目标 Q 值与预测 Q 值的均方差, 即:

$$L = \frac{1}{m} \sum_{i=1}^m [y_i - Q(s_i, \mathbf{a}_i; \theta, \alpha, \beta)]^2$$

式中: m 为样本数量。

7) 利用计算得到的损失函数, 采用随机梯度下降更新当前 Q 网络的参数, 以减少损失并改善策略, 从而得到最优模型。

8) 定期更新目标网络参数 θ' , 使其等于当前网络参数 θ 。

9) 重复步骤 2)~8), 直到达到最大迭代次数。

所提算法通过优化神经网络结构, 将神经网络的输出 Q 值分解为独立的状态价值函数和动作优势函数。这种分离能够更准确地评估每个状态的价值, 而不仅是动作的价值, 从而降低 Q 值估计的不准确性。每次更新时, 状态价值函数均会被更新, 影响所有动作的 Q 值, 使得学习更加准确。这些改进提高了 Q 值的估计精度, 提升了学习效率和训练的稳定性, 有效解决了 Nash-DQN 在复杂环境中的不足。

4 仿真实验与结果

4.1 环境设置

为了在接近真实物理与市场环境下对所提出的基于 Nash-Dueling DQN 的联合调度策略进行量化评估与验证, 下面介绍仿真实验的系统架构、关键参数及运行工况。

为充分检验所提策略在多智能体协调与博弈中的性能, 本文构建的仿真场景包含 2 个通过 P2P 交易互联的并网型综合能源微网。这种 1 个能源富余、1 个能源短缺的异质化结构是微网集群研究中的典型配置, 能有效激发微网间的交易动机, 从而为验证协调调度策略的有效性提供理想的测试平台, 具体设定如下: 微网 1 (少电型), 其可再生能源发电量通常难以满足自身负荷需求, 该微网倾向于作为电能购买方, 用于检验策略降低用能成本的能力; 微网 2 (多电型), 其可再生能源发电量在多数时段超出自身负荷, 该微网倾向于作为电能出售方, 用于检验策略提升运行收益的能力。

为确保仿真数据的真实性, 仿真所用的风力发电、光伏发电、电负荷及热负荷数据均来自 Fingrid^[34]、Pecan Street^[35] 等国际公认的能源数据集, 并处理为 1 h 分辨率。2 个微网内部各关键设备的运行参数如附录 B 表 B1 所示。

4.2 结果分析

4.2.1 算法性能分析

在仿真实验中, 分别采用 Nash-DQN 和本文提出的 Nash-Dueling DQN 算法进行训练, 训练过程中微网 1 和微网 2 的奖励值分别如图 3(a)、(b) 所示, 图中数据经过了小数点左移 3 位处理以优化展示。由图可知, 采用所提算法的微网奖励值更高且收敛更快。在训练的初始阶段, 所提算法执行以探索为主的策略, 导致微网的初始奖励值处于相对较低的水平。随着训练迭代的逐步推进, 经验回放缓冲区积累了大量样本数据, 智能体基于这些数据不断提炼和优化策略。经验的累积效应显著提升了算法的决策能力, 从而使得微网的奖励值得到提高并逐渐趋

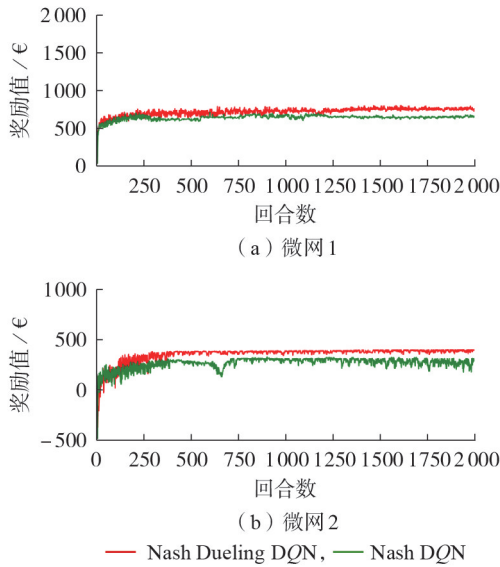


图3 微网奖励值对比

Fig.3 Comparison of microgrid reward

于稳定。

Nash-Dueling DQN算法表现更优的核心原因在于其独特的网络结构。通过将 Q 值分解为与状态相关的价值函数和与动作相关的优势函数,该算法能够更高效地评估在当前微网状态(如负荷大小、电价高低)下的整体优劣,而不是仅评估单个动作的好坏。在复杂的微网调度环境中,许多动作对状态价值的影响甚微,Dueling架构能够识别并忽略这些无关紧要的动作,从而加速学习过程。这种结构上的优化使得智能体能够更快地收敛到稳定的纳什均衡策略,尤其是在本文构建的这种电-热-碳联合调度博弈场景中,其性能提升尤为显著。

4.2.2 电、热、碳交易分析

本文所提策略能够引导各微网根据实时变化的市场价格(如附录B图B2所示)和自身能源需求(如附录B图B3所示)作出经济最优的调度决策。由日内电、热、碳的详细平衡结果(如附录B图B4—B6所示)可知,所提策略实现了高效的能量时空互补。能源短缺的微网1主要通过向微网2和主网购电来满足缺口,而能源富余的微网2则通过向外售电和电解槽制氢来最大化其能源价值。

为验证阶梯式碳交易机制的有效性,对微网1和微网2分别采用阶梯式和传统碳交易机制时连续10 d内的碳排放量进行比较,如图4所示。图4(a)中,微网1在传统碳交易机制下的10 d总碳排放量为6843 kg,而在阶梯式碳交易机制下的总碳排放量降至4848 kg,排放量降低了29.2%;图4(b)中,微网2在传统碳交易机制下的碳排放总量为4655 kg,而在阶梯式碳交易机制下的碳排放总量减少至3389 kg,排放量降低了27.2%。可见,阶梯式碳交易机制在

降低碳排放方面具有显著优势,该机制具有有效性的原因在于其对高碳排放行为的惩罚性成本:通过实施分级的碳定价策略,使排放量越高的行为面临越高的边际成本。在本文的仿真中,由于从主网购电的碳排放强度远高于其他能源形式,当微网1因大量外购电量而面临碳排放超额风险时,这种惩罚机制便被激活,这迫使强化学习智能体在决策时必须对购电的直接成本与潜在的高昂碳罚款进行权衡,从而更倾向于选择向微网2购买低碳电力或利用储能等替代方案。因此,该机制成功地将宏观的减排政策目标转化为每个智能体在博弈过程中的经济优化驱动力,进而引导整个系统走向更节能、低碳的运行模式。

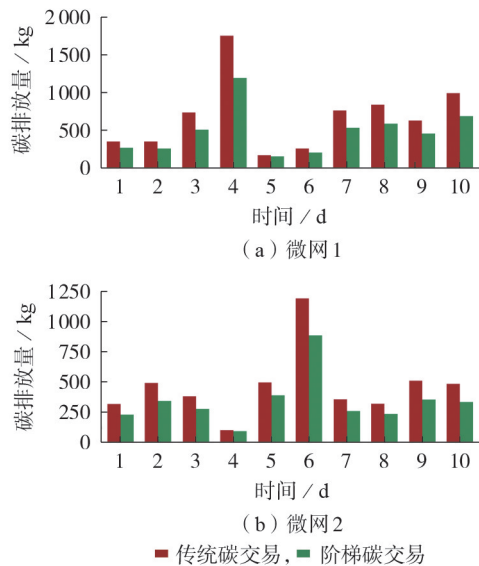


图4 阶梯碳交易与传统碳交易碳排放量对比

Fig.4 Comparison of carbon emission between stepping carbon trading and traditional carbon trading

为定量评估所提 Nash-Dueling DQN算法在低碳调度方面的性能,将其与 Nash DQN算法下的系统日均碳排放进行对比。结果显示,本文算法展现出较强的系统级减排效益,促使多微网系统的总碳排放量从1583.37 kg降低至1149.80 kg,降幅达27.4%。这种整体效益源于算法更优的协同调度能力。能源短缺的微网1的碳排放量从1146.73 kg大幅削减至684.30 kg,这主要归因于算法引导该微网最大化地消纳了来自微网2的低碳电力,替代了高碳排的主网能源。能源富余的微网2的碳排放量从436.64 kg增至465.50 kg。这种现象并非调度性能的不足,而是算法在寻求全局最优解时的策略体现:微网2通过微调自身运行策略来支撑微网1,从而实现了整个系统碳排放的最小化。上述结果表明,所提 Nash-Dueling DQN算法能够发掘并收敛于更优的纳什均衡点,有效突破单一主体追求局部利

益的局限性,最终实现了系统整体的低碳经济运行目标。

4.2.3 运行成本和收益

各微网在所提策略下的成本和收益如图5所示。由图可知,本文算法能引导微网捕捉不同时段主网与微网间的价差波动及自身供需状态,灵活选择最优交易对象,从而在复杂的动态市场环境中实现利益最大化。

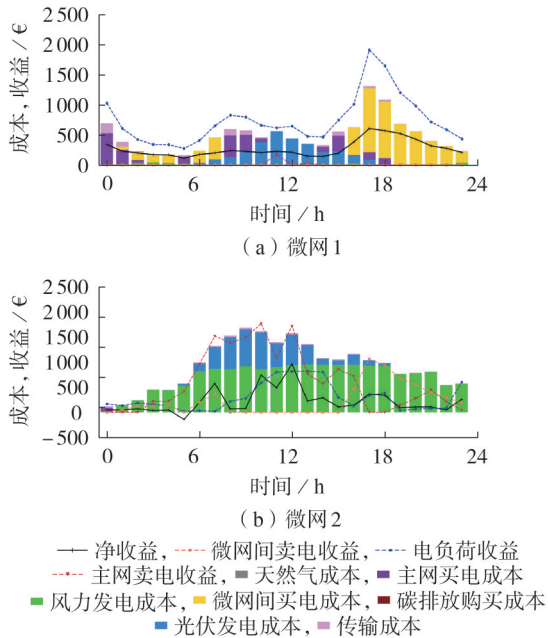


图5 微网成本和收益

Fig.5 Cost and benefit of microgrid

为验证所提算法在提升系统经济效益方面的有效性,对Nash-Dueling DQN与Nash DQN算法调度下的各微网日均净收益进行比较。结果显示,微网1和微网2的日均净收益分别由Nash DQN算法下的€268.77和€-37.83,提升至所提算法下的€269.42和€223.22。本文算法在提升多微网系统整体经济性方面具有显著优势,系统总净收益由€230.94大幅提升至€492.64,增幅高达113.3%。该提升主要源于对“多电型”微网2经济效益的根本性改善。在传统Nash DQN算法下,微网2未能有效利用其富余能源进行交易,最终其日均收益为€-37.83,处于亏损状态。而在本文提出的Nash-Dueling DQN算法的精细化调度下,微网2的日均净收益跃升至€223.22,实现了从亏损到盈利的质变。同时,“少电型”微网1的收益也未因微网2的盈利而受损,其日均净收益为€269.42。该结果表明所提算法能够发掘出更优的博弈均衡策略,通过高效的能量交易与协同调度,不仅避免了无益的内部竞争,还实现了各参与方利益的均衡与系统整体经济效益的最大化。

5 结论

本文构建计及电、热、碳能源交易的微网联合调度模型,并提出基于Nash-Dueling DQN博弈均衡求解算法。所提算法通过优化神经网络结构,加快了收敛速度,增强了学习效率和训练的稳定性,并提升了微网的经济效益。通过算例分析验证了所提算法的可行性和有效性,并得出以下结论:

1)本文提出包含电、热、碳交易的多微网能量调度模型,其中微网智能体之间通过持续的博弈策略互动达到纳什均衡状态,在微网间实现了内部能量的互相补充,各智能体的运行净收益达到了纳什均衡;

2)所提Nash-Dueling DQN算法通过优化神经网络结构,提高了Q值的估计精度,提升了学习效率和训练的稳定性,有效地弥补了Nash-DQN算法在策略评估和动作选择方面的不足,使多智能体在复杂环境下能够根据当前环境状态作出更准确的决策,提升了多智能体强化学习的决策效率;

3)本文充分考虑微网内部多种能源之间的转换效率,在模型中纳入碳捕集技术的应用,并采用阶梯式碳交易机制,减少了碳排放量。

附录见本刊网络版(<http://www.epae.cn>)。

参考文献:

- [1] SAEED M H, WANG F Z, KALWAR B A, et al. A review on microgrids' challenges & perspectives [J]. IEEE Access, 2021, 9: 166502-166517.
- [2] ROURAY S K, MOHAPATRA S, SAHU P C, et al. Microgrid performance study: grid connected and islanding under common disturbance & approach [C] // 2023 International Conference in Advances in Power, Signal, and Information Technology (APSIT). Bhubaneswar, India: IEEE, 2023: 543-548.
- [3] JIA Y W, LYU X, XIE P, et al. A novel retrospect-inspired regime for microgrid real-time energy scheduling with heterogeneous sources [J]. IEEE Transactions on Smart Grid, 2020, 11(6): 4614-4625.
- [4] DOU C X, ZHANG Z Q, YUE D, et al. Improved droop control based on virtual impedance and virtual power source in low-voltage microgrid [J]. IET Generation, Transmission & Distribution, 2017, 11(4): 1046-1054.
- [5] 董雷, 杨子民, 乔骥, 等. 基于分层约束强化学习的综合能源多微网系统优化调度 [J]. 电工技术学报, 2024, 39(5): 1436-1453.
DONG Lei, YANG Zimin, QIAO Ji, et al. Optimal scheduling of integrated energy multi-microgrid system based on hierarchical constraint reinforcement learning [J]. Transactions of China Electrotechnical Society, 2024, 39(5): 1436-1453.
- [6] 蔡瑶, 卢志刚, 潘尧, 等. 计及多重差异的交直流混合多能微网多时间尺度优化调度 [J]. 电工技术学报, 2024, 39(11): 3392-3410.
CAI Yao, LU Zhigang, PAN Yao, et al. Multi-time-scale optimal scheduling of AC-DC hybrid multi-energy microgrid considering multiple differences [J]. Transactions of China Electrotechnical Society, 2024, 39(11): 3392-3410.

- [7] 郭立东,雷鸣宇,杨子龙,等. 光储微网系统多目标协调控制策略[J]. 电工技术学报,2021,36(19):4121-4131.
GUO Lidong, LEI Mingyu, YANG Zilong, et al. Multi-objective coordinated control strategy for photovoltaic and energy-storage microgrid system[J]. Transactions of China Electrotechnical Society, 2021, 36(19): 4121-4131.
- [8] HE G X, LIU K C, ZHANG X H, et al. Cooperative optimal control of microgrid group with distributed generation [C]//2022 China International Conference on Electricity Distribution(CICED). Changsha, China: IEEE, 2022: 852-856.
- [9] 黄堃,付明,翟家祥,等. 基于改进线性化ADMM的多微网经济运行分布式协调优化[J]. 中国电力,2025,58(2):193-202.
HUANG Kun, FU Ming, ZHAI Jiaxiang, et al. Distributed coordination optimization for economic operation of the multi-microgrid system based on improved linearization ADMM[J]. Electric Power, 2025, 58(2): 193-202.
- [10] 刘浩,王丹,刘佳委,等. 计及分布式水风光发电时空相关性的多微网协同优化策略[J]. 电力系统保护与控制,2025,53(13):23-35.
LIU Hao, WANG Dan, LIU Jiawei, et al. Multi-microgrid collaborative optimization strategy considering spatiotemporal correlation of distributed hydro-wind-solar generation[J]. Power System Protection and Control, 2025, 53(13): 23-35.
- [11] 邵振国,林勇棋,陈飞雄,等. 考虑碳-绿证耦合机制的电-气-热综合能源系统分布式鲁棒低碳经济调度[J]. 电力自动化设备,2024,44(6):50-58,76.
SHAO Zhenguo, LIN Yongqi, CHEN Feixiong, et al. Distributionally robust low-carbon economic dispatch of electricity-gas-heat integrated energy system considering carbon-green certificate coupling mechanism[J]. Electric Power Automation Equipment, 2024, 44(6): 50-58, 76.
- [12] 吴严杰,平健,严正,等. 基于差分隐私的多园区综合能源系统分布式调度方法[J]. 电力系统自动化,2025,49(5):80-88.
WU Yanjie, PING Jian, YAN Zheng, et al. Differential privacy based distributed dispatch method for multi-park integrated energy system[J]. Automation of Electric Power Systems, 2025, 49(5): 80-88.
- [13] 程杉,尚冬冬,魏昭彬,等. 基于纳什议价博弈的微电网一体化电站分布式协调优化[J]. 电机与控制学报,2022,26(5):86-95.
CHENG Shan, SHANG Dongdong, WEI Zhaobin, et al. Hierarchical and distributed coordination and optimization of microgrid with CSSIS based on Nash bargaining game[J]. Electric Machines and Control, 2022, 26(5): 86-95.
- [14] LI C J, YU X H, SOKOLOWSKI P, et al. A stochastic game for energy resource trading in the context of Energy Internet [C]//2016 IEEE Power and Energy Society General Meeting (PESGM). Boston, MA, USA: IEEE, 2016: 1-5.
- [15] 初壮,李秋雨,王议坚. 基于非零和博弈的互联异构多微网系统电-热-碳优化调度策略[J]. 电网技术,2024,48(6):2287-2296.
CHU Zhuang, LI Qiuyu, WANG Yijian. Optimal dispatch of interconnected heterogeneous multi-microgrids system considering electricity-heat-carbon trades based on non-zero-sum game [J]. Power System Technology, 2024, 48(6): 2287-2296.
- [16] 梁忠豪,王丽芳,李建林. 基于纳什均衡的光-储耦合制氢系统优化控制方法[J]. 电力系统自动化,2025,49(3):125-134.
LIANG Zhonghao, WANG Lifang, LI Jianlin. Nash equilibrium based optimal control method for hydrogen production system with photovoltaic-storage coupling [J]. Automation of Electric Power Systems, 2025, 49(3): 125-134.
- [17] 王胜彩,王琨,张少华. 电-气-碳-绿证多市场下综合能源供应商联合均衡模型[J]. 电力系统保护与控制,2025,53(16):60-73.
WANG Shengcai, WANG Xian, ZHANG Shaohua. Joint equilibrium model for integrated energy suppliers in coupled multi-markets of electricity-gas-carbon-green certificate[J]. Power System Protection and Control, 2025, 53(16): 60-73.
- [18] 郑杰凯,何山,韩璐,等. 计及混合博弈配电网与氢能系统的多综合能源服务商优化调度[J]. 电力自动化设备,2025,45(1):67-75.
ZHENG Jiekai, HE Shan, HAN Lu, et al. Optimal scheduling of multi-integrated energy service provider considering hybrid game-based distribution network and hydrogen energy system [J]. Electric Power Automation Equipment, 2025, 45(1): 67-75.
- [19] 林顺富,董炯炯,周波,等. 基于混合博弈的多微网日前-日内协同优化调度[J]. 电力自动化设备,2025,45(8):106-113.
LIN Shunfu, DONG Jiongiong, ZHOU Bo, et al. Day-ahead and intra-day coordinated optimization scheduling of multiple microgrids based on mixed-game[J]. Electric Power Automation Equipment, 2025, 45(8): 106-113.
- [20] 骆钊,王菁慧,王华,等. 考虑碳捕集和电转气的综合能源系统优化调度[J]. 电力自动化设备,2023,43(12):127-134.
LUO Zhao, WANG Jinghui, WANG Hua, et al. Optimal scheduling of integrated energy system considering carbon capture and power-to-gas[J]. Electric Power Automation Equipment, 2023, 43(12): 127-134.
- [21] 乐生铜,张亚超,朱蜀,等. 考虑低碳需求响应的碳捕集燃煤电厂配置-运行协同优化[J]. 电力自动化设备,2024,44(7):278-286.
LE Shengtong, ZHANG Yachao, ZHU Shu, et al. Collaborative optimization of configuration-operation for coal-fired power plants with carbon capture considering low-carbon demand response[J]. Electric Power Automation Equipment, 2024, 44(7): 278-286.
- [22] 陈明健,陈胜,王异成,等. 考虑氢能绿证的电-氢综合能源系统机会约束优化调度[J]. 电力自动化设备,2023,43(12):206-213.
CHEN Mingjian, CHEN Sheng, WANG Yicheng, et al. Chance constrained optimal scheduling of electric-hydrogen integrated energy system considering green certificate of hydrogen energy [J]. Electric Power Automation Equipment, 2023, 43(12): 206-213.
- [23] 王杰,贾宏杰,靳小龙,等. 基于电-氢-热P2P交易的分布式智能电网低碳优化运行方法[J]. 电力系统自动化,2025,49(9):40-51.
WANG Jie, JIA Hongjie, JIN Xiaolong, et al. Low-carbon operation optimization method for distributed smart grid based on electricity-hydrogen-heat peer-to-peer trading[J]. Automation of Electric Power Systems, 2025, 49(9): 40-51.
- [24] 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.
- [25] 刘俊峰,王晓生,卢俊波,等. 基于多主体博弈和强化学习的多微网系统协同优化研究[J]. 电网技术,2022,46(7):2722-2732.
LIU Junfeng, WANG Xiaosheng, LU Junbo, et al. Collaborative optimization of multi-microgrid system based on multi-agent game and reinforcement learning [J]. Power System Technology, 2022, 46(7): 2722-2732.

- [26] 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.
- [27] 韩莹, 于三川, 李萃一, 等. 计及阶梯式碳交易的风光氢储微电网低碳经济配置方法[J]. 高电压技术, 2022, 48(7): 2523-2533.
HAN Ying, YU Sanchuan, LI Luoyi, et al. Low-carbon and economic configuration method for solar hydrogen storage microgrid including stepped carbon trading[J]. High Voltage Engineering, 2022, 48(7): 2523-2533.
- [28] ZHAO N, JIA H R, ZHANG Z H. Research on electrical energy allocation strategies based on Nash equilibrium[C]//2023 3rd International Conference on Computer Science, Electronic Information Engineering and Intelligent Control Technology(CEI). Wuhan, China: IEEE, 2024: 719-722.
- [29] PRIYADHARSHINI P, SUPRIYA S. Multi-robot path planning using Nash equilibrium based game theory[C]//2023 International Conference on Recent Advances in Information Technology for Sustainable Development(ICRAIS). Manipal, India: IEEE, 2023: 188-193.
- [30] HUANG H, LI Y G, LIU H Z. Source-network-load-storage cooperative scheduling based on Nash-Q algorithm[C]//2023 5th Asia Energy and Electrical Engineering Symposium(AEES). Chengdu, China: IEEE, 2023: 1208-1214.
- [31] ZHANG Z, XU S L, XIA Y Y, et al. Two improved algorithms based on DQN[C]//2023 IEEE International Conference on Signal Processing, Communications and Computing(ICSPCC). Zhengzhou, China: IEEE, 2024: 1-4.
- [32] XIAO H, PU X W, PEI W, et al. A novel energy management method for networked multi-energy microgrids based on improved DQN[J]. IEEE Transactions on Smart Grid, 2023, 14(6): 4912-4926.
- [33] ZHANG Z, XU S L, XIA Y Y, et al. Two improved algorithms based on DQN[C]//2023 IEEE International Conference on Signal Processing, Communications and Computing(ICSPCC). Zhengzhou, China: IEEE, 2024: 1-4.
- [34] Fingrid. Fingrid open datasets[EB/OL]. (2021-08-09)[2025-01-13]. <https://data.fingrid.fi/en/datasets>.
- [35] Pecan Street Inc. Dataport[EB/OL]. (2018-12-31)[2025-01-13]. <https://www.pecanstreet.org/dataport/>.

作者简介:

黎海涛(1972—),男,副教授,博士,主要研究方向为智能网络与计算(**E-mail**: lihaitao@bjut.edu.cn);

乔禄(2002—),男,硕士研究生,主要研究方向为网络通信与智能计算(**E-mail**: qiaolu@bjut.edu.cn);

杨艳红(1985—),男,创新研究员,博士,通信作者,主要研究方向为电力系统运行控制与优化方法、人工智能在电力系统中的应用(**E-mail**: yangyanhong@mail.iee.ac.cn)。

(编辑 王锦秀)

Nash-dueling deep Q network based electricity-heat-carbon joint dispatch strategy of integrated energy microgrid

LI Haitao¹, QIAO Lu¹, XIE Dongxue¹, YANG Yanhong²

(1. School of Information Science and Technology, Beijing University of Technology, Beijing 100124;

2. Institute of Electrical Engineering, Chinese Academy of Sciences, Beijing 100190)

Abstract: Along with the large-scale integration of various renewable energy generation equipments into the main power grid in the form of microgrid, the coordinated dispatch and operation among microgrids become increasingly important. However, the existing multi-agent reinforcement learning has the shortcomings of dimensionality disaster and low training efficiency when dealing with dispatch problem. Therefore, a joint dispatch strategy of microgrids is proposed based on Nash-dueling deep Q network reinforcement learning. An energy dispatch model considering electricity, heat and carbon trading is constructed for integrated energy microgrid, which adopts carbon capture equipment and stepping carbon emission trading mechanism in the aspect of carbon emission management. The game relationship among microgrids is analyzed, and a Nash-dueling deep Q network based electricity-heat-carbon joint dispatch strategy is proposed. The example simulation results show that the proposed dispatch strategy can improve the economic benefit of microgrid and effectively reduce the carbon emission, which realizes the temporal and spatial energy complementation and benefit equilibrium among microgrids.

Key words: microgrid; dispatch; carbon capture; multi-agent game; Nash equilibrium; deep reinforcement learning; stepping carbon trading; integrated energy system