基于强化学习的综合能源系统管理综述

熊珞琳[1] 毛 帅[1] 唐 漾[1] 孟 科[2] 董朝阳[2] 钱 锋[1]

摘 要 为了满足日益增长的能源需求并减少对环境的破坏, 节能成为全球经济和社会发展的一项长远战略方针, 加强能源管理能够提高能源利用效率、促进节能减排. 然而, 可再生能源和柔性负载的接入使得综合能源系统 (Integrated energy system, IES) 发展成为具有高度不确定性的复杂动态系统, 给现代化能源管理带来巨大的挑战. 强化学习 (Reinforcement learning, RL) 作为一种典型的交互试错型学习方法, 适用于求解具有不确定性的复杂动态系统优化问题, 因此在综合能源系统管理问题中得到广泛关注. 本文从模型和算法的层面系统地回顾了利用强化学习求解综合能源系统管理问题的现有研究成果, 并从多时间尺度特性、可解释性、迁移性和信息安全性 4 个方面提出展望.

关键词 强化学习, 能源管理, 电力系统, 综合能源系统

引用格式 熊珞琳, 毛帅, 唐漾, 孟科, 董朝阳, 钱锋. 基于强化学习的综合能源系统管理综述. 自动化学报, 2021, **47**(10): 2321−2340

DOI 10.16383/j.aas.c210166

Reinforcement Learning Based Integrated Energy System Management: A Survey

XIONG Luo-Lin[1] MAO Shuai[1] TANG Yang[1] MENG Ke[2] DONG Zhao-Yang[2] QIAN Feng[1]

Abstract In order to meet the growing energy demand and reduce the damage to the environment, energy conservation has become a long-term strategic policy for global economic and social development. The enhancement of energy management can improve energy efficiency, as well as promote energy conservation and emission reduction. However, the integration of renewable energy and flexible load makes the integrated energy system (IES) become a complex dynamic system with high uncertainty, which brings great challenges to modern energy management. Reinforcement learning (RL), as a typical interactive trial-and-error learning method, is suitable for solving optimization problems of complex dynamic systems with uncertainty, and therefore it has been widely considered in integrated energy system management. This paper systematically reviews the existing works of using reinforcement learning to solve integrated energy system management problems from the perspective of models and algorithms, and puts forward prospects from four aspects: Multi-time scale, interpretability, transferability, and information security.

Key words Reinforcement learning (RL), energy management, power system, integrated energy system (IES)

Citation Xiong Luo-Lin, Mao Shuai, Tang Yang, Meng Ke, Dong Zhao-Yang, Qian Feng. Reinforcement learning based integrated energy system management: A survey. *Acta Automatica Sinica*, 2021, **47**(10): 2321−2340

能源是人类社会生存和发展的重要物质基础, 社会的发展伴随着能源需求日益增长, 化石能源的大量使用带来环境污染、生态破坏和全球气候变暖等一系列问题[1−2]. 为了解决能源可持续供应以及环境污染等问题, 以电能为核心, 在源端整合了太阳能、风能、生物质能、海洋能、地热能等清洁可再生能源, 在终端实现热、电、冷联供的综合能源系统 (Integrated energy system, IES) 成为当今世界能源领域研究的热点[3]. 随着全球能源供应多元化和社会对各类能源需求的不断增加, 加强对综合能源的管理不仅能够提高能源利用率、减少对环境的破坏, 也能提升经济发展质量和效益[4]. 电能作为综合能源的核心, 是把握国家经济命脉的关键因素[3], 因此本文从系统层面将综合能源管理问题分为仅考虑单一电能的电力系统管理问题和考虑多种能源的综合能源系统管理问题.

综合能源系统的大规模区域互联使其逐渐发展成为大型高维系统, 间歇性可再生能源和包含电动汽车 (Electric vehicle, EV)、分布式储能设备在内

收稿日期 2021-03-02 录用日期 2021-05-20
Manuscript received March 2, 2021; accepted May 20, 2021
国家自然科学基金基础科学中心项目 (61988101), 国家杰出青年科学基金 (61725301), 中央高校基本科研业务费专项资金 (222202117006), 上海市优秀学术带头人计划 (20XD1401300) 资助
Supported by Project of Basic Science Center of National Natural Science Foundation of China (61988101), National Science Fund for Distinguished Young Scholars (61725301), the Fundamental Research Funds for the Central Universities (222202117006), and Program of Shanghai Academic Research Leader (20XD1401300)
本文责任编委 孙秋野
Recommended by Associate Editor SUN Qiu-Ye
1. 华东理工大学信息科学与工程学院 上海 200237 中国 2. 新南威尔士大学电气工程与电子通信学院 新南威尔士州 2052 澳大利亚
1. School of Information Science and Engineering, East China University of Science and Technology, Shanghai 200237, China 2. School of Electrical Engineering and Telecommunications, University of New South Wales, NSW 2052, Australia

的柔性负载的接入增加了综合能源系统的复杂动态特性[5−6], 另外用户能源消耗行为的随机性、能源多样性和不同形式能源之间的耦合关系也给现代化能源管理带来了巨大的挑战[7−8]. 混合整数规划[9]、线性规划[10]、非线性规划[11] 等传统优化算法往往依赖于精确的数学模型和参数, 考虑到综合能源系统是具有高度不确定性的复杂动态系统, 精确的模型构造十分困难, 因此传统优化算法在求解综合能源系统管理问题中的应用受到限制[12].

作为人工智能的一个重要分支, 强化学习 (Reinforcement learning, RL) 因其强大的自主学习能力, 获得了许多专家学者的关注[13−19]. 具体来讲, 强化学习不需要监督信号来直接指导学习, 只依赖于一个反馈回报信号, 对其 "试错"过程进行评估, 间接指导智能体向反馈回报值最大的方向进行学习, 从而减少对精确的系统模型的依赖. 目前, 强化学习算法已广泛应用于机器人导航[13]、计算机游戏[14]、计算机视觉[15] 和化学合成[16] 等领域.

针对综合能源系统的高度不确定性, 传统优化方法需要对不确定因素提前预测[20] 并利用动态场景生成方法对环境进行估计, 进一步建立能源系统动态模型. 这类方法不仅计算量大, 而且优化结果极大程度上取决于不确定因素预测和动态场景生成的准确度, 当预测结果偏差较大时, 即使性能优良的求解算法也无法得到最优解[21]. 然而在强化学习方法中, 智能体可以在不同的系统状态下尝试不同的动作, 并从奖励回报中学习知识以获得最优策略, 智能体与环境交互的整个过程可以不依赖于详细精确的模型信息, 因此所得策略的性能也不受制于预测结果的精度[22].

针对综合能源系统的变量高维度特性, 强化学习可以采用多层马尔科夫决策过程 (Markov decision process, MDP) 模型进行分层优化. 在面对一些具有连续动作和状态空间的问题时, 强化学习还可以与具有出色数据处理能力的深度学习相结合构成深度强化学习算法 (Deep reinforcement learning, DRL), 进而求解得到具有高维变量的综合能源系统的最优管理策略[23], 并且该方法相较于传统优化方法在实际生活场景下更容易实现[21].

基于强化学习的无模型依赖性、变量复杂性的优点, 许多专家学者致力于利用强化学习算法来处理综合能源系统管理问题, 并取得了一系列研究成果[17−19]. 同时一些学者基于这些研究作了相关综述, 例如文献 [24] 从拓扑结构、优化目标、时间尺度、调度优化结构等方面综述了互联微电网的能源管理方案; 文献 [25] 基于大功耗家庭供暖通风空调控制系统 (Heating, ventilation, and air conditioning, HVAC)、智能家庭、智能商业和住宅建筑这三个系统的能源管理问题, 综述了利用深度强化学习算法求解的能源管理方案; 文献 [26] 系统地总结了强化学习、深度强化学习和多智能体强化学习分别在电力和能源系统中的应用.

本文在现有研究成果和相关综述的基础上, 从模型和算法两个方面系统回顾了基于强化学习的综合能源系统管理问题. 在模型方面, 将单一电能从综合能源中提出来单独讨论, 把综合能源管理问题分为电力系统和综合能源系统管理问题, 在电力系统管理中依次讨论了微电网、智能家庭以及公共电动汽车这三个关注度较高的电能优化管理问题, 即互联微电网电能调度、智能家庭用电管理和电动汽车充放电规划. 在算法方面, 主要分析各类问题中用到的不同强化学习算法并对比其性能. 图 1 是本文的结构框架及主要内容. 第 1 节主要介绍强化学习算法的定义、分类及面临的挑战和解决方法; 第 2 节主要总结了强化学习算法在电力系统优化管理中的应用; 第 3 节聚焦于多种异质能源协调优化、互补互济的综合能源系统中, 分别介绍了综合能源系统优化管理模型和利用强化学习算法求解得到的综合能源系统管理方案; 第 4 节对综合能源系统管理问题面临的挑战进行展望, 并结合强化学习方法提出相应的潜在解决方案; 第 5 节对本文工作进行简单总结.

1 强化学习简单介绍及分类

随着人工智能技术的发展进入新的历史阶段, 强化学习作为人工智能领域中一种快速、高效的学习算法, 是当前的研究热门, 受到许多学者的广泛关注[13−19]. 强化学习与依赖直接监督信息的监督学习不同, 它让智能体通过与环境的持续交互获取环境知识, 并通过采取最优动作获得最大回报以实现其目标. 在解决具有延时回报的序列决策问题中, 智能体与环境的交互过程通常被建模为马尔科夫决策过程模型[27].

强化学习在马尔科夫决策过程中主要使用的方法包括自适应动态规划 (Adaptive dynamic programming, ADP)[28]、时间差分 (Temporal difference, TD) 学习[29]、蒙特卡洛法 (Monte carlo, MC)[27] 等.

根据学习方式的不同, 强化学习可以分为在线策略和离线策略[27]. 其中, 在线策略是指生成样本的策略与网络更新参数时使用的策略不同, 即与环境互动和网络更新同时进行, 一边采样一边更新. 离线策略则是指生成样本的策略与网络更新参数时

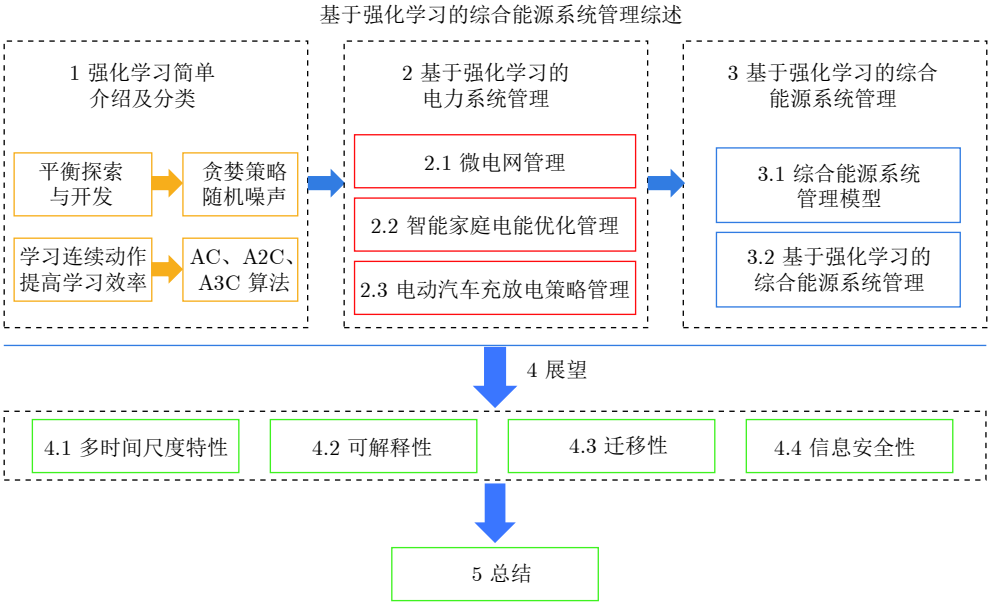

图 1　结构及主要内容

Fig. 1　The structure and main contents

使用的策略相同, 采用先采样后集中更新的方式进行学习[27]. 两者的本质区别在于, 更新 Q 值的方法是沿用既定策略还是新策略. 以此为依据, 时间差分学习又分为状态–动作–回报–状态–动作 (State-action-reward-state-action, SARSA) 算法和 Q 学习算法 (Q Learning)[27].

根据动作的选择依据, 强化学习又可以分为基于价值的强化学习和基于策略的强化学习[27]. 其中, 基于价值的强化学习是在知晓所有动作价值的基础上, 根据最高价值来选择动作, 因此并不适用于选取连续动作. 基于策略的强化学习则是通过对环境的分析, 直接输出下一步可能采取的各种动作的概率, 然后根据概率采样选取行动[27].

在强化学习中, 系统的模型包括环境的状态空间、动作空间以及状态转移概率等. 根据模型是否完全给定, 强化学习还可以分为基于模型的强化学习和无模型的强化学习[27]. 其中, 基于模型的强化学习依赖于环境在各个动作下的状态转移概率, 而无模型的方法不需要完整的环境信息, 当给予适当的奖励时智能体可以自主学习最优策略[27].

强化学习在应用过程中会面临许多挑战, 例如如何平衡探索与开发、如何处理高维决策问题、如何减小状态动作价值的估计误差、如何提升学习效率等. 在选择策略的过程中如何平衡探索与开发是一个常见的问题, 其中探索是指尝试之前没有执行过的动作以期望获得超过当前最优动作的奖励回报, 开发是指执行已经学习到的能获得最大奖励回报的动作, 即贪婪动作. 因此以现有的动作价值为参考, 开发是相对正确的, 但是由于一些具有更高价值的动作可能还未被发现, 从长期来看探索可能会比开发带来更大的收益. 所以需要在开发和探索之间找到一个平衡, 避免陷入局部最优, 并收敛到全局最优. 一种平衡探索与开发的方法是采取贪婪策略, 智能体在每个状态有 $1-\epsilon$ 的概率选择进行开发, 有 ϵ 的概率选择进行探索. 当动作空间为 A 时, $|A|$ 是该空间中的动作总数, 除贪婪动作外各个动作被采取的概率为 $\epsilon/(|A|-1)$[30]. 另一种方法是在每次得到贪婪动作的基础上添加随机噪声, 使得采取的动作是在贪婪动作邻域内随机探索的结果[31]. 但是由于没有考虑每次探索动作的价值, 添加随机噪声的方法存在数据利用率低、充分探索需要无限长时间等不足.

为了处理高维决策问题, 具有感知能力的深度学习和具有决策能力的强化学习相结合产生了深度强化学习算法[23]. 深度学习中深度神经网络从高维数据中提取低维特征, 能够有效解决维度灾害的问题, 再与强化学习相结合解决具有高维状态和动作空间的序列决策问题. 深度 Q 网络 (Deep Q network, DQN)[14]、演员–评论家算法 (Actor-critic, AC)[32] 都是常见的深度强化学习算法. 此外, 针对变量耦合的问题, 传统优化算法中耦合变量和耦合的约束条件使得建立机理模型存在困难, 也为后续的求解增加了难度. 然而强化学习算法具有无模型依赖性, 智能体从与环境交互过程获得的奖励回报中学习知识, 可以克服复杂耦合变量和约束条件带来的困难[33].

在所有目标的状态动作价值都是通过执行贪婪动作直接得到的情况下, DQN 中目标 Q 值的计算更新公式如式 (1) 所示[14]

$$Q(s_t,\ a_t) = r_t + \gamma \max_{\hat{a}} Q(s_{t+1},\ a_{t+1}) \tag{1}$$

其中, r_t 是 t 时刻在状态 s_t 下采取动作 a_t 得到的奖励回报, γ 是折扣因子, $Q(s_{t+1},\ a_{t+1})$ 指下一时刻的状态动作价值. 这里的 max 操作会使现有状态动作价值 $Q(s_t,\ a_t)$ 被高估, 对不同动作不同程度的高估可能会导致某些次优策略价值超过实际最优策略的价值, 从而永远无法找到最优策略. 针对 Q 值被高估的问题, 有学者提出了深度双 Q 网络 (Double deep Q network, Double DQN), 通过解耦动作的选择和目标 Q 值的计算, 来解决过度估计问题, 提升算法性能[34]. 类似地, 深度竞争 Q 网络 (Dueling deep Q network, Dueling DQN) 也能提高估计值的精确度, 提升算法稳定性[35].

Q 学习无法在连续动作空间中选择合适的动作, 策略梯度方法能有效解决这个问题, 但是传统的策略梯度方法采用回合更新的形式, 降低了学习效率. 因此有学者提出了演员−评论家算法[32], 该算法融合了以状态动作价值为基础 (比如 Q 学习) 和以动作概率为基础 (比如策略梯度) 的两类强化学习算法. 优势演员−评论家算法 (Advantage actor-critic, A2C)、异步优势演员−评论家算法 (Asynchronous advantage actor-critic, A3C)[36]、置信域策略梯度算法 (Trust region policy optimization, TRPO)[37]、近端策略优化算法 (Proximal policy optimization, PPO)[38]、深度确定性策略梯度算法 (Deep deterministic policy gradient, DDPG)[39−40] 都是基于演员−评论家算法改进得到的算法, 并被众多专家学者用于高效求解具有连续动作空间的能源管理问题[41−45]. 按照是否基于模型、选择动作的依据和学习方式, 本文对强化学习算法进行了如表 1 所示的分类.

2 基于强化学习的电力系统管理

优化电能分配方式、提高电能利用效率在促进可持续发展进程中起到重要作用, 因此本文首先聚焦电力系统管理问题. 本节将依次介绍面向微电网、智能家庭、电动汽车管理问题的基于强化学习的方法. 这些问题具有相似的经济性和社会性优化目标, 例如降低购电成本、系统运营成本或操作成本以提升系统经济性, 降低负荷曲线峰均比以提升电力系统安全性、稳定性; 它们也面临相似的挑战, 例如系统的高度不确定性、变量的高维耦合特性以及难以建立精确的系统模型等. 由于智能体在与环境交互的过程中可以自主学习环境知识, 不依赖于精确的环境模型, 因此相较于依赖不确定因素预测精度的传统优化方法, 强化学习能够更好地处理无模型的综合能源系统管理优化问题. 但是由于不同场景中电能管理的时间尺度是不同的, 例如对电价的优化可以是日前调度, 而对涡轮发电机、电动汽车或家庭用电设备的调度则需要更小时间尺度下的日内滚动优化或实时调整[46], 因此用到的强化学习算法也有一定差异. 下面将对上述问题进行详细的分析和总结.

2.1 微电网管理

微电网是集成了分布式电源、储电系统、电能转换设备和用电负载的小型配电系统[47]. 在微电网电能优化管理中, 优化变量主要包括电力交易价格、功率分配方案等, 优化目标包括最大化运营商收益、最小化购电成本、提高用户用电满意度、减少能量传输损失、提高新能源利用率、提高系统稳定性等. 其常见模型如式 (2) ~ 式 (4) 所示[48]

$$\begin{aligned}\min \sum_{t=1}^{N_T}\Bigg(&\sum_{k\in m} C_{DG}^{P}\left(P_k^{DG}(t)\right) + \eta_m \lambda(t) P_m^{\text{grid}}(t) + \\ &\sum_{z=1}^{Z} ec_m^z q_m^z(t) u_m^z(t) + \\ &\rho_{es}\left|SOC_{es}(t) - SOC_{es}(t-1)\right|\Bigg)\end{aligned} \tag{2}$$

表 1 强化学习算法分类

Table 1 The classification of reinforcement learning algorithm

强化学习算法类型	模型		选择动作的依据		学习方式	
	有模型	无模型	基于价值	基于策略	在线策略	离线策略
MC[27]		√	√		√	
SARSA[27]		√	√		√	
Q Learning[27]		√	√			√
ADP[28]	√		√			
DQN[14] (Dueling DQN[35]/Double DQN[34])		√	√			√
AC[32] (A2C/A3C[36]/TRPO[37]/PPO[38]/DDPG[39-40])		√	√	√		

s. t.

$$SOC_{es}(t) = SOC_{es}(t-1) + \eta_{es} P_{es}^{ch}(t)\Delta - \frac{P_{es}^{\text{dis}}(t)}{\eta_{es}\Delta} \tag{3}$$

$$P_m^{\text{Load}}(t) - P_m^{\text{grid}}(t) - \sum_{k\in m} P_k^{DG}(t) - \sum_{es\in m}\left(P_{es}^{\text{dis}}(t) - P_{es}^{ch}(t)\right) - \sum_{z=1}^{Z} q_m^z(t) u_m^z(t) = 0 \tag{4}$$

其中, $P_m^{\text{Load}}(t)$ 表示第 m 个微电网的负载功率, $P_m^{\text{grid}}(t)$ 表示微电网从主网购买的电量, $P_k^{DG}(t)$ 表示微电网中第 k 个分布式电源的发电量, $P_{es}^{\text{dis}}(t)$ 和 $P_{es}^{ch}(t)$ 分别表示微电网中储电设备发电量和充电量, $q_m^z(t)$ 表示微电网中第 z 个需求响应块的响应电量, $u_m^z(t)$ 为表示该响应块是否响应的二值变量. 式 (2) 是微电网管理问题的一种常见的目标函数, 表示最小化调度周期 N_T 内第 m 个微电网的运行成本. 第 1 项是分布式电源的发电成本, $C_{DG}^P\left(P_k^{DG}(t)\right)$ 是与发电量 $P_k^{DG}(t)$ 有关的二次多项式; 第 2 项是微电网与主网之间电力交换的成本, η_m 表示电能传输过程中的网络损耗, $\lambda(t)$ 表示公共耦合点处的时变电力零售价格; 第 3 项是微电网内部调度成本, ec_m^z 表示调度单价; 第 4 项是储电设备的损耗成本, ρ_{es} 表示电量变化引起储电设备性能退化、寿命变短的系数, $SOC_{es}(t)$ 表示储电设备的电量状态. 式 (3) 是储电设备电量状态变化的约束, 其中 η_{es} 表示储电设备充放电效率, Δ 表示充放电时间. 式 (4) 是微电网功率平衡的约束. 此外, 常见的约束条件还包括发电机容量约束、储电设备充放电速率约束、电能存储容量约束等[48].

如图 2 所示, 对微电网实施能源优化管理主要从供电侧、储电系统和需求侧三个方面进行考虑.

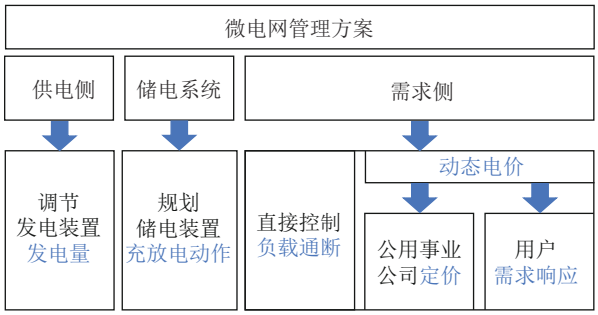

图 2　微电网管理方案

Fig. 2　Microgrid management approach

供电侧管理与调节发电装置的发电量有关[49]. 储电系统管理通过规划充放电动作来协调系统电能供求关系[41, 50−55]. 需求侧管理主要分为两类, 一类是直接控制负载通断[56], 另一类通过动态电价间接管理功率分配. 动态电价对电能的间接管理又可以细分为两个方向[57]: 其一, 站在公用事业公司的立场, 通过设计有效的定价策略最大程度地提高社会福利, 包括智能电网中所有消费者的总收益和公用事业公司售电获得的收益[42, 47, 58−60]; 其二, 站在消费者的立场, 针对公用事业公司的定价策略, 设计有效的价格感知型需求调度策略以最大化消费者个人收益. 在本节将着重讨论动态定价方案设计问题, 用户需求响应和用电设备的规划将在第 2.2 节中阐述.

表 2 罗列了基于强化学习的微电网管理的相关文献, 从时间尺度、管理方案和求解算法这三个角度进行总结, 并从收敛稳定性、计算速度、隐私保护和适应性 4 个方面分析这些算法的性能, 其中 √ 表示文献提出的算法在这方面具有较好的性能, 空白的单元格表示文中没有提到这方面的性能. 根据时间尺度的不同, 微电网的电能优化管理问题中具体包括了日前调度、日内滚动优化和实时调整三种时

表 2　基于强化学习的微电网管理

Table 2　Microgrid management based on reinforcement learning

文献	时间尺度	管理方案	求解算法	算法性能			
				收敛稳定	计算速度	隐私保护	适应性
[22]	日内滚动	公用事业公司定价	自适应强化学习		√	√	√
[41]	实时调整	储电装置调节	深度确定性策略梯度				√
[43]	日前调度	消费者价格感知	有限时域深度确定性策略梯度		√		
	日内滚动		有限时域递归确定性策略梯度		√	√	
[47]	日内滚动	公用事业公司定价	博弈论 + 强化学习	√			√
[48]	日前调度	公用事业公司定价	蒙特卡洛法			√	
[50]	实时调整	储电装置调节	深度双 Q 网络	√			
[56]	日前调度	直接负载控制	深度竞争 Q 网络	√			
[58]	日内滚动	公用事业公司定价	Q 学习				√

间尺度的优化类型. 日前调度是阶段性的, 考虑到电力系统的高度不确定性, 预测可能存在偏差, 所以需要更小时间尺度的优化方案, 例如日内滚动优化和实时调整, 其中, 日内滚动优化是指在某个时间窗口内以日前计划作为参考, 利用时间窗口动态更新的模型数据滚动求得最优策略[61]. 实时调整则是以小时或者更短的时间为单位进行实时优化. 预测精度随时间尺度的减小而逐渐提高, 更小的时间尺度优化往往具有更好的性能, 但也需要进行更复杂的计算[62].

2.1.1 常规算法求解微电网管理问题

微电网系统包含的可再生能源 (例如太阳能光伏发电、风力发电) 生成的不确定性以及用户需求的随机性, 使微电网管理问题模型难以建立, 同时高维优化变量以及非线性约束的存在也为求解此类优化问题带来困难.

传统优化求解算法包括遗传算法 (Genetic algorithm, GA)[63]、粒子群算法 (Particle swarm optimization, PSO)[64−65]、混合整数线性规划 (Mixed integer linear programming, MILP)[63] 以及动态规划算法[53, 66] 等. 例如 Shu 等[63] 面向公用事业公司提出一种融合遗传算法和混合整数线性规划的混合优化算法来确定最优动态零售电价, 在提高了公用事业公司利润的同时改善了大型工业用户的用电方式, 降低其平均用电成本. Mirzaei 等[64] 通过自适应粒子群算法求解由多个微电网和电动汽车组成的双层能源系统管理问题, 以减少高峰时段的负载需求使得负载曲线平整化. Jin 等[53] 在存在分布式可再生能源和时变电力价格的场景中, 利用动态规划算法求解得到储电设备的最佳运行策略, 最大程度地降低消费者的电力消费成本.

针对可再生能源生成的不确定性问题, Li 等[67] 使用区间预测方法预测微电网中风力涡轮机和光伏电池的不确定功率输出, 并通过混沌群体搜索优化方法求解满足微电网运行经济性、电能质量和安全性要求的多目标优化问题. 针对用户用电需求的随机性问题, Bao 等[68] 提出了一种面向工业客户的多时标需求侧最优调度框架, 用动态场景生成方法模拟调度时段内客户用电时间序列的不确定性.

然而, 上述区间预测和动态场景生成方法不仅计算量大, 而且策略优化性能极大程度地取决于不确定因素预测和动态场景生成的准确度, 当预测结果偏差较大时, 即使性能优良的求解算法也无法得到最优策略. 考虑到优化变量的高维性和模型的不可知, 一些文章采用强化学习算法[22, 43], 在节约计算成本的同时提升了算法在面向不同场景的优化问题时的适应性.

2.1.2 强化学习求解储电系统管理问题

一些学者利用强化学习方法对储电系统进行管理, 进而解决电能分配调度优化问题. 例如通过在用电低峰期充电、用电高峰期放电来降低用电成本, 平整负荷曲线; 在光照强或者风力大的时候利用光伏发电或风力发电为储电设备充电, 在电价高或用户用电需求增加时放电, 以满足用户用电需求并降低电力成本. 文献 [41] 提供了一种利用 DDPG 算法进行训练的控制器用于管理储电系统的充放电状态, 同时为电网提供频率响应服务. Qazi 等[69] 提出了基于 DQN 的孤立微电网集群能源和储备调度的概念, 通过共享能源和储备来提高微电网的经济效益, 最大程度降低其运营成本. Jayaraj 等[70] 面向包含光伏单元和电池的微电网利用 Q 学习算法实现经济调度, 减少了电网的净交易成本, 并给出以 24 小时为周期的电池运行调度方案. 文献 [50] 提出了一种基于 Double-Q 学习的方法, 在实时电价和煤炭价格不确定的情况下, 求解得到并网微电网中的储电套利策略. 其中, Double-Q 学习的主要思想是使用两个神经网络将选择策略和评估策略进行分离, 因此该算法可以在迭代更新后更准确地收敛到最优解.

2.1.3 强化学习求解需求侧电能管理问题

一些学者从需求侧管理的角度通过直接控制负载通断对电能进行管理. 文献 [56] 使用 Dueling-DQN 算法学习控制可中断负载的状态, 实现电压调节并减少分布式系统的总操作成本. Dueling-DQN 算法用两个深度网络分别表示状态价值网络 $V(s_t)$ 和动作优势函数网络 $A(s_t, a_t)$, 其输出将两者结合以产生状态动作值 $Q(s_t, a_t)$, 克服传统 DQN 中的噪声和不稳定性, 提高模型收敛稳定性.

制定动态电价是一种更为常见的需求侧能源管理方法, 从公用事业公司的角度来看, 一般将提升利润作为首要优化目标. 例如, Liang 等[42] 采用 DDPG 算法求解公用事业公司的电价竞标策略, 最终实现社会收益最大化. 文献 [58] 提出了一种用于分级电力市场能源管理的动态定价需求响应算法, 将动态定价问题建模为离散有限马尔科夫决策过程, 服务供应商通过 Q 学习算法在线自适应地制定零售电价, 同时实现提高服务供应商利润、降低客户成本、平衡电力市场的能源供需、提高电力系统的可靠性等优化目标. 除了经济性目标之外, 合理分配功率以增加用户用电满意度和降低峰均比也是微电网能源管理中的重要优化目标. 例如文献 [48] 提出了一种基于深度神经网络和无模型强化学习算

法的多微电网能源管理方法, 配电系统运营商 (Distribution system operator, DSO) 利用深度神经网络来预测各微电网的功率交换而无需直接访问用户信息. DSO 通过蒙特卡洛方法求解得到零售定价策略, 既能使 DSO 的利润最大化又能降低需求侧的峰均比, 提高用电可靠性. 与之相似, Zhang 等[22] 在无法直接访问用户信息的条件下, 让智能体基于自适应强化学习框架通过函数逼近来预测微电网功率分配行为, 并优化价格信号, 最终最大化微电网总收益.

从消费者的角度来看, 对实时电价进行感知并调整配电策略可以节省用电费用. Lei 等[43] 针对深度强化学习算法的不稳定性和有限时域模型的独特性, 提出了两种新的 DRL 算法, 即有限时域深度确定性策略梯度算法和有限时域递归确定性策略梯度算法, 分别在有、无完全可观测状态信息这两种情况下学习到包含柴油发电机、光伏电池板和蓄电池的孤立微电网的能源调度策略, 在满足用户电力需求的基础上降低了分布式电源的发电成本, 并最大限度地利用了可再生能源.

2.1.4 电力系统管理中的博弈和隐私保护问题

由于我国电网具有多主体、强不确定性、多目标的特征, 电力系统管理决策问题已逐步由单人优化决策向具有不同目标的多决策者博弈转换. 例如在消纳发电侧大规模风电和光电的问题中, 存在经济性与环保性的权衡; 输电网中, 为保障不确定环境下电网安全, 大自然和电网存在着博弈; 在电动汽车等储电设备灵活接入微电网进行充电和放电的问题中, 存在售电商与用户之间的博弈, 所以在电能管理问题中考虑博弈论和强化学习结合的方法是十分必要的[42]. 例如文献 [47] 提出了一种用于微/纳米电网的内部能源管理和外部能源交易的三层优化方案. 第一层提出一个在线随机需求侧能源管理模型, 并用强化学习算法求解各个网络内部的用电调度方案; 第二层制定了双重拍卖机制, 使各个网络之间可以直接进行电力交易; 第三层由中央控制器制定最佳功率分配策略, 以减少功率传输损耗和局部能源交易可能会带来的破坏性影响.

随着社会和科技的发展, 用户的隐私保护问题得到越来越多的关注[71], 例如文献 [22]、[48] 在无法观测用户对价格作出响应的情况下, 只能选择通过神经网络或函数逼近来预测特定价格信号下的功率交换信息. 文献 [43] 在状态信息部分可观的情况下利用历史信息进行优化. 此外, 文献 [64] 中的双层能源管理模型也在一定程度上保护了用户的隐私.

2.2 智能家庭电能优化管理

随着太阳能光伏电池板、智能电表、电动汽车、家用电池和其他 “智能”设备的普及[72], 智能家庭的概念进入人们视野, 由此家庭耗能优化管理问题得到广泛关注, 一些学者把研究目光聚焦到家庭用电设备的调度管理上来. 由于家用设备数量较多, 而且不同设备具有不同的控制策略, 例如对照明设备的控制可能是连续的功率控制、对洗碗机的控制是离散的开关控制, 传统优化方法对家用设备的管理需要针对不同设备建立不同的模型, 而强化学习算法可以只用一个网络输出不同的参数, 对不同的设备同时优化提高效率.

实时定价和能源调度是家庭能源管理的两个重要组成部分[73]. 在实时定价方面, 主要考虑包含可再生能源发电设备、储电设备[74] 和可充放电电动汽车的家庭在开放市场中进行交易的场景. 文献 [75] 提出了一种基于深度演员–评论家的多智能体扩展算法, 在环境部分可观测并且感知非平稳的条件下学习实时定价方案, 以降低所有家庭总能耗峰均比和用电成本.

在能源调度方面, 不同文献采用不同的方式对负载进行分类, 可以分为不可调负载、运行时间可调负载、运行功率可调负载[21], 或者再进一步将时间可调负载分为连续时间工作负载和可中断负载[44, 76]. 然后根据各类负载运行特性分别实施调度策略, 在节约用电成本的同时, 提升用户用电满意度和舒适度. 文献 [77] 将负载分为常开负载、可开关负载和可灵活调节负载, 并用双向长短期记忆 (Long short term memory, LSTM) 网络预测电力和能源价格, 在此基础上用 Q 学习算法进行优化, 实现了能耗减少和成本降低. 文献 [44] 提出了一种基于置信域策略梯度的家用电器高效需求响应算法, 该方法不依赖模型, 并且通过同一个策略网络输出不同概率分布的参数, 基于不同的概率分布进一步采样得到不同类型设备优化后的离散动作或连续动作. 文献 [76] 对比了 DQN 和确定性策略梯度法 (Deterministic policy gradient, DPG) 的优化性能, 根据电价实时在线优化用电设备的动作, 实现用电总花费最小, 同时考虑了开关频率对用电设备和用户舒适度的影响, 仿真结果证明 DPG 算法在降低用电成本和降低峰均比方面有更好的效果.

此外, 由于多重不确定因素在不同时间尺度上表现出不同的分布特性, 许多文章选择在不同时间尺度上进行优化, 包括日前调度、日内滚动优化和实时调整. 例如, Xu 等[21] 在滚动时间窗口下利用神经网络对不确定性因素预测并进行优化, 而 Lu 等[78]

则提出了一种提前一小时的家庭能源管理实时需求响应算法. 常见的目标函数如式 (5) 所示[21]

$$\begin{aligned}\min R = -\sum_{t\in T}\Bigg\{\lambda_t^G\Bigg(&\left[P_{it}^{d,\,NS}-E_{it}^{PVs}\right]^+ - \\ &\left[P_{jt}^{d,\,PS}-E_{jt}^{PVs}\right]^+ - \\ &\left[u_{mt}P_{mt}^{d,\,TS}-E_{mt}^{PVs}\right]^+ - P_{nt}^{d,\,EV}\Bigg) - \\ &\Bigg(\alpha_j^{PS}\left(P_{j,\,\max}^{d,\,PS}-P_{jt}^{d,\,PS}\right)^2-\alpha_m^{TS}\left(t_m^s-t_m^{\text{ini}}\right)^2 - \\ &\alpha_n^{EV}\left(P_{n,\,\max}^{d,\,EV}-P_{nt}^{d,\,EV}\right)^2\Bigg)\Bigg\}\end{aligned}\tag{5}$$

其中, $[\bullet]^+$ 表示到非负数上的投影运算, 即 $[x]^+=\max(x,\,0)$. 第 1 项为用电成本, 主要由时间周期 T 中 i 个不可调负载消耗的功率 $P_{it}^{d,\,NS}$、j 个功率可调负载消耗的功率 $P_{jt}^{d,\,PS}$、m 个时间可调负载消耗的功率 $P_{mt}^{d,\,TS}$、n 辆电动汽车消耗的功率 $P_{nt}^{d,\,EV}$ 与光伏发电产生的功率 E_t^{PVs} 之差和电价 λ_t^G 决定, u_{mt} 是一个表示时间可调负载开关状态的二值变量. 第 2 项为用户需求没有被满足时的不满意成本, 主要由降低功率可调负载的工作功率 $(P_{j,\,\max}^{d,\,PS}-P_{jt}^{d,\,PS})$ 引起的不满意成本、增加时间可调负载运行等待时间 $(t_m^s-t_m^{\text{ini}})$ 引起的不满意成本和电动汽车未充满电 $(P_{n,\,\max}^{d,\,EV}-P_{nt}^{d,\,EV})$ 情况下担心不能到达目的地的焦虑成本组成, 其中 $P_{j,\,\max}^{d,\,PS}$ 表示功率可调负载工作功率的上限, t_m^s 和 t_m^{ini} 分别表示时间可调设备实际开始工作的时间和正常情况下应该开始工作的时间, $P_{n,\,\max}^{d,\,EV}$ 表示电动汽车能耗上限, α_j^{PS}、α_m^{TS}、α_n^{EV} 为不满意系数, 较高的 α 值意味着调度设备时更容易引起用户不满意. 当负载分类不同时, 目标函数具有类似的形式[44].

对家庭供暖通风空调控制系统进行管理也是家庭能源管理的一个热点, 许多专家在这方面进行了深入的研究[40, 79−80]. 由于模型和参数的不确定性 (如可再生能源发电、电力需求、室外温度和电价) 以及时间耦合约束的存在, 文献 [40] 提出了具有注意力机制的多智能体深度强化学习方法, 在不需要任何关于不确定参数的先验知识和建筑物热动力学模型的情况下进行学习, 并获得优化控制策略. 类似地, 文献 [45] 设计的基于 DDPG 的能源管理算法, 也不需要参数和模型的先验知识, 仿真结果验证了该算法的有效性和鲁棒性. 文献 [25] 按照模型规模从小到大的顺序对基于深度强化学习的智能建筑能源管理作了相关综述, 从大功耗 HVAC、智能家庭、智能商业和住宅建筑三个方面进行了详细而全面的总结.

值得一提的是, 在家庭能源管理问题中, 包括用电时间、用电量等能体现用户偏好习惯的私人信息也可以得到有效的保护. 例如, 文献 [81−82] 通过增加各个设备耗电量和使用时间的相似度或者加入储电设备充放电操作来掩盖用户用电偏好信息, 文献 [83] 通过添加噪声来隐藏有效的用户用电信息, 文献 [84−85] 通过平整负载曲线来加强隐私保护.

2.3 电动汽车充放电策略管理

得益于国家政策的扶持以及电池技术和电动马达技术的发展, 电动汽车市场逐年扩张[86], 如何通过调度电动汽车充放电行为达到降低充电成本的目标一直是人们关注的焦点. 鉴于电动汽车充放电的灵活性, 许多研究场景考虑利用随机的太阳能或风能为其充电. 私人电动汽车在第 2.2 节中作为一种特殊的家庭负载或移动储电设备已经被讨论, 因此本节主要讨论公用电动汽车的充放电规划调度问题.

庞大的电动汽车数量使调度优化变量具有高维特性, 并且可再生能源发电和用户需求的不确定性使得模型难以建立. 文献 [87] 设计了基于参数自适应差分进化的多目标优化算法, 但是该方法需要在计算风电功率的概率基础上建立电动汽车−风能集成电力系统协调调度模型. 文献 [88] 采用的基于场景树的动态规划方法必须具备对不确定性模型完全准确可知的能力, 并生成场景树来描述系统动态变化.

针对电动汽车充放电规划问题中的不确定性主要有两种处理方法. 一种是在决策优化之前对其进行预测得到估计值[89−90]. 其中, 通过物理模型或者概率分布来预测不确定性因素是较为简单且常见的[91−92], 适用于精确度要求较低的场景 (如日前小时级预测需求); 通过利用历史数据训练得到的神经网络进行预测[93] 的方法, 对数据要求比较高、计算复杂, 更适合精确度要求较高的场景 (如日内分钟级预测需求)[94], 因此它往往出现在单独的预测问题中. 另一种方法得益于深度强化学习算法的兴起, 它将历史数据作为系统状态直接输入到智能体中, 智能体通过神经网络自行提取其中的特征, 而后输入策略网络进行学习得到最优策略. 该过程无需输出预测结果值, 属于数据驱动的方法, 因此得到的策略优劣也不依赖于预测结果的精度. 本文将电动汽车充放电策略管理问题中处理不确定因素的方法分为三种, 即机理模型驱动 (简单模型预测)、数据驱动和

模型已知 (包含通过精确预测得到的模型, 在本文中不作详细讨论). 例如, 文献 [95] 提出的基于深度强化学习的方法包含两个网络: 一个代表网络, 用于从电价中提取特征; 一个 Q 网络, 用于近似最佳动作价值函数. 类似地, 文献 [31] 利用 LSTM 网络从历史能源价格中提取相关特征, 用充电控制深度确定性策略梯度方法进行优化.

针对优化变量的高维特性, 一些文献从模型上通过定义电动汽车聚合器、事件和子状态将具有高维变量的电动汽车充放电管理问题进行分层优化, 以降低每一层的变量维数, 同时能一定程度上保护下层用户的隐私信息, 适用于有隐私保护需求的高维系统优化问题. 例如将具有相同剩余电量或相同剩余停车时间或停在同一位置的电动汽车定义为一个电动汽车聚合器, 构建双层或者三层[96] 优化模型, 上层对电动汽车聚合器群体进行电量分配, 下层对各个聚合器内部的电动汽车进行充放电管理[97]. 文献 [97] 基于双层马尔科夫模型开发了一种双层近端策略优化算法来实现充电成本最小化. 文献 [98] 提出了一种基于事件的策略迭代方法, 在假设风能服从正态分布和充电量服从基于停车时间的正态分布条件下, 在上层定义了一系列事件以确定每个聚合器要充电的电动汽车数量, 下层具体决定每辆电动汽车的充电计划, 有效降低了电动汽车的充电成本. 文献 [99] 提出一种基于分布式模拟的策略改进方法对基于经验的策略进行改进, 并且通过将建筑集合群内的电动汽车定义为一个子状态来避免维度灾难. 另外, 文献 [100] 采用一种新颖的二维表格从模型上简化电动汽车充电调度问题, 其中一维表示需要充电的时间, 另一维表示剩余停车时间, 每个单元格的值表示该状态电动汽车数量占总数的比例, 因此模型大小仅与充电时间和剩余停车时间相关, 不会随着电动汽车数量增加而呈指数上升, 从而有效避免维度灾难. 而且二维表格的建模方式还具有可扩展性, 例如当电动汽车具有异质性时, 可以将表格扩展到三维, 第三维表示不同电动汽车的充电效率.

强化学习算法作为解决具有多重不确定性的复杂动态系统管理问题的另一种思路, 可以与深度学习结合从而解决高维状态空间和动作空间的难题. 利用强化学习求解电动汽车充放电策略问题, 首先需要建立马尔科夫决策过程模型, 其中系统状态主要包括风力发电量 W_t^j、剩余所需充电量 E_t^i、剩余停车时间或剩余行驶时间 L_t^i, 和电动汽车位置 D_t^i, 动作可以用简单的二值变量 1 和 0 表示是否充电, 更复杂的情况可以考虑多个离散动作 (例如充电、放电、既不充电也不放电) 或者连续动作 (连续数值表示充放电具体电量). 在给定系统状态和动作的情况下, 电动汽车状态动力学如式 (6) ~ 式 (8) 所示[99]

$$L_{t+1}^i=\begin{cases}L_t^i-\Delta t, & \text{if}\quad L_t^i>0\\ \tau^{t+1}, & \text{if}\quad L_t^i=0,\ D_t^i=0\\ \eta^{t+1}, & \text{if}\quad L_t^i=0,\ D_t^i>0\end{cases}\tag{6}$$

$$D_{t+1}^i=\begin{cases}D_t^i, & \text{if}\quad L_t^i>0\\ R^{t+1}, & \text{if}\quad L_t^i=0,\ D_t^i=0\\ M+1, & \text{if}\quad L_t^i=0,\ D_t^i>0\end{cases}\tag{7}$$

$$E_{t+1}^i=\begin{cases}E_t^i-z_t^i\times P\times\Delta t, & \text{if}\ L_t^i\geq 0,\ D_t^i>0\\ \max(B_{t+1}^i-f_i(\eta^{t+1}),\ 0), & \text{if}\ L_t^i=0,\ D_t^i=0\\ 0, & \text{otherwise}\end{cases}\tag{8}$$

其中, 式 (6) 表示电动汽车剩余停车时间或剩余行驶时间的动态特性, Δt 表示时间间隔, τ^{t+1} 和 η^{t+1} 为两个随机变量, 分别表示电动汽车 i 在 $t+1$ 时刻到达时的剩余停车时间和离开后的剩余行驶时间; 式 (7) 表示电动汽车位置变化情况, R^{t+1} 表示电动汽车 i 在 $t+1$ 时刻所到达的位置; 式 (8) 表示剩余所需充电量的变化, z_t^i 是一个二维动作变量, 1 和 0 分别表示第 i 辆电动汽车是否充电, P 为恒定的充电功率, B_{t+1}^i 表示电动汽车到达时的电量状态, $f_i(\eta^{t+1})$ 表示电动汽车能耗与其行驶时间 η^{t+1} 的关系.

此外, 电动汽车充电决策问题中的约束条件可以分为可行性约束和安全性约束. 其中, 可行性约束主要针对策略的可行性, 例如电池电量状态受电池容量限制、充电状态受电动汽车位置限制等; 安全性约束主要考虑供电量与充电需求间的平衡、单位时间充电功率大小限制等. 对于有复杂约束的优化问题, 很难实现直接求解, 常见的方法是将各个约束考虑为优化目标进行加权求和, 从而将具有复杂约束的优化问题转化为多目标无约束优化问题. 通过设计惩罚函数对违反约束的动作进行惩罚也是一种常见的处理约束的方法. 基于类似的思想, 强化学习可以对不同动作设计不同的奖励回报值来惩罚违反约束的动作、奖励满足约束的动作. 但是由于过高的惩罚会使智能体学习效率降低, 过低的惩罚不利于系统的安全性, 因此设计适当的奖励函数存在一定的困难. 在文献 [101] 中, 作者将电动汽车充放电调度问题建模为约束马尔科夫决策过程, 并提出了一种基于安全深度强化学习的无模型方法, 在不需要关于不确定因素的任何知识、不需要设计惩罚项或调整惩罚系数的情况下, 直接使用深度神

经网络学习满足约束的最佳充放电策略. 此外, 将约束嵌入环境模型也是一种处理约束的方法. 例如文献 [102] 利用约束深度双 Q 网络, 在包含随机风能的场景中, 将动作约束模型嵌入到深度双 Q 学习网络中, 以解决状态空间过大且决策受限的 MDP 问题, 减少了 Q 值估计的误差, 并通过生成更有效的训练数据提高充电策略的准确性.

在电动汽车充放电规划调度问题中, 优化目标除了包含最小化充电成本外, 还包含提升用户用车需求满足率、降低弃风率[87]、降低充电时间成本[103]、避免电动汽车充放电导致变压器过载等实际需求[104]. 文献 [105] 提出了一种针对配电网的最优电动汽车充电策略, 在满足所有物理约束的同时最大化配电系统运营商的利润, 并利用 DDPG 算法来分析不确定的用户用车行为对充电策略的影响, 满足电动汽车电能需求的同时最大程度地减少用户的充电费用. 表 3 总结了处理可再生能源发电和用户用车需求不确定性以及高维变量问题的一些常规算法和强化学习算法, 并从计算速度和算法适应性角度分析了算法的性能. 其中 √ 表示文献提出的算法在这方面具有较好的性能, ×表示文献中提到的算法在这方面具有较差的性能, 空白的单元格表示文中没有提到这方面的性能. 从表 3 可以看出, 利用强化学习可以处理具有不确定性的无模型问题, 而深度网络既可以解决高维变量带来的困难, 也能对不确定因素进行预测[106]. 因此, 深度强化学习算法能更好地解决此类具有多重不确定性的高维无模型问题.

3 基于强化学习的综合能源系统管理

受理论方法和各种能源技术的限制, 以前天然气、电能和热能等能源系统往往是独立计划和运行的, 相互之间缺乏协调, 由此产生的诸如能量利用率低、能源系统的灵活性和可靠性低等问题亟待解决[107]. 可再生能源技术、分布式发电技术、综合能源利用技术和能源管理技术的迅速发展为综合能源系统的形成和发展提供了技术支持. 以电力为核心, 耦合了燃气、热力及其他能源的综合能源系统已经成为国际能源领域的重要战略方向[108], 其目标是通过拓宽能源来源和减少能源消耗建立可持续的能源系统, 从而缓解能源危机并减少环境污染. 在此背景下, 除了综合能源系统外[109], 类似的多能协调、互补共济的能源利用形式还包括能源互联网 (Energy Internet, EI)[110] 和自能源 (We-energy)[111]. 其中, 能源互联网是以电力网络、热力网络、天然气网络及交通网络等复杂网络为物理实体的一种新型开放式能源生态系统, 自能源是能够实现能量间双向传输及灵活转换的能源互联网子单元. 本节主要讨论综合能源系统管理问题.

3.1 综合能源系统管理模型

协同管理多种能源可以提高能源利用率, 保证用能可靠性, 提升用户满意度, 解决能源可持续供应以及环境污染等问题[108, 112]. 然而, 综合能源系统具有多元大数据、源荷双端不确定、时空多维耦合等特征, 亟需理论方法和关键技术的突破. 对于综合能源系统, 许多文章运用智能能源枢纽 (Smart energy hub, SEH)[113]、多能载波 (Multi-energy carrier, MEC)[114]、热电联产 (Cogeneration, combined heat and power, CHP)、冷热电三联产 (Combined cooling, heating and power, CCHP)[115] 的概念协调优化多种能源以实现经济性和社会性目标. 例如文献 [116] 在优化能源枢纽 (Energy hub, EH) 调度时考虑了三种目标函数, 分别是最小化当前净成本、

表 3 电动汽车充放电管理算法

Table 3 The algorithm of charge and discharge management of electric vehicle

文献	不确定性处理	高维变量处理	求解算法	算法性能		
				计算速度	适应性	备注
[31]	数据驱动	深度网络	充电控制深度确定性策略梯度	√		
[87]	机理模型驱动	—	参数自适应差分进化			计算时间仅相对于传统差分进化法可以接受
[88]	模型已知	分层优化	基于场景树的动态规划	×		
[95]	数据驱动	深度网络	深度 Q 网络		√	
[97]	模型已知	分层优化	双层近端策略优化	√		仅总体性能优于其他策略、能较好地跟踪风能发电
[99]	机理模型驱动	分层优化	基于分布式模拟的策略改进			分布式方法具有可扩展性
[100]	数据驱动	二维表格	拟合 Q 迭代			性能受给定训练集时间跨度的影响
[101]	数据驱动	—	安全深度强化学习		√	
[103]	数据驱动	深度网络	深度 Q 网络		√	
[105]	数据驱动	深度网络	深度确定性策略梯度	×		闭环控制框架严格保证电压安全性

最小化二氧化碳总排放量以及同时最小化当前净成本和二氧化碳总排放量. 在文献 [117] 中, 通过调度电力和天然气的交换以及能源枢纽的能源分配, 不仅减小运营成本实现经济性目标, 而且顺应可持续发展规律减少碳排放实现社会性目标. 常见的如式 (9a)、(9b)、(9c) 和式 (10) 所示, 优化目标为最小化能源成本 $J(t)$[118]

$$\min_{P_{gC}(t),\, P_{gB}(t)} \quad \sum_t J(t) = \sum_t J_e(t) + J_g(t) \tag{9a}$$

$$J_e(t) = \left(L_e(t) - P_{gC}(t) \times \eta_e^C\right) \times Pr_e(t) \tag{9b}$$

$$J_g(t) = \left(P_{gC}(t) + P_{gB}(t)\right) \times Pr_g(t) \tag{9c}$$

s. t.

$$P_{gC}(t) \times \eta_h^C + P_{gB}(t) \times \eta_h^B \geq L_h(t) + \frac{L_c(t)}{\eta_C} \tag{10}$$

其中, $J_e(t)$ 指电力成本, $J_g(t)$ 指天然气成本. 电力成本取决于电力负荷 $L_e(t)$、天然气输入热电联产系统后的发电量 $P_{gC}(t) \times \eta_e^C$ 和电力价格 $Pr_e(t)$, 其中 η_e^C 表示热电联产系统的电能转化效率; 天然气成本主要由热电联产和锅炉的天然气输入 $P_{gC}(t)$、$P_{gB}(t)$ 以及天然气价格 $Pr_g(t)$ 决定. 此外, 求解得到的优化策略还需要满足用户热能需求的约束条件 (10) 以及热电联产系统和锅炉等的输入容量约束, 式 (10) 中 η_h^C 和 η_h^B 表示热电联产系统和锅炉的热能转化效率, η_C 表示制冷机的效率, $L_h(t)$ 和 $L_c(t)$ 表示热负载和冷负载需求. 不同文献考虑的约束有所不同, 例如文献 [119] 还考虑了电力负载平衡约束、热能供需平衡约束、各个设备能量输出上下限的约束等. 文献 [120] 已经对优化目标和传统求解算法进行了总结, 本文主要聚焦于综合能源系统模型的规模级别和时间尺度分析, 并对常规算法进行简单对比.

从规模级别来看, 综合能源系统包括城市能源系统[121]、社区能源系统[122]、工厂能源系统[123] 和家庭能源系统[124]. 文献 [125] 将由多个能源枢纽构成的合作社区作为研究对象, 研究了共享能量的合作经济调度问题, 将能源交换和定价问题建模为合作博弈过程, 在考虑不同 EH 目标的条件下实现 Pareto 最优的平衡. 对多个决策者应用分布式优化算法寻找合作系统的议价解决方案, 保证了 EH 的自主调度和信息保密性.

由于综合能源系统是一个多时空尺度的耦合系统, 不同优化对象具有不同时空特性, 例如热能具有热惯性, 因此对热能进行管理调度的频率可以比电能低. 用户对价格变化的响应较快, 因此以价格为导向的需求响应往往比较快. 准确预测不确定性因素并进行提前计划能够提升方案性能, 日前调度[126] 是一种常见的方式, 但是在实际运行中计划情况可能会与实际情况发生偏差, 导致计划方案的可行性降低. 因此, 除日前调度外, 往往需要在更小的时间尺度内进行更为精确的优化, 例如文献 [122] 在社区级能源系统的合作交易模式下提出一种实时滚动能源管理模型. 在日前调度的基础上, 还可以与日内滚动优化、实时调整等不同时间尺度的调度相配合, 形成多时间尺度优化, 进一步提高优化策略的性能[127]. 表 4 从综合能源系统规模级别和不同时间尺度的角度对部分文献进行总结.

从算法的角度, 表 4 主要总结几种传统算法, 在运用这些方法的过程中, 不同文献利用不同方式处理双端不确定性、多种能源耦合、非线性目标等问题. 例如 Ma 等[122] 考虑了光伏发电的随机特征和可变负荷, 用风险条件值综合考虑当期成本和未来成本. 文献 [128] 通过混合整数非线性规划优化方法解决了不确定环境下的能源枢纽非线性调度问题. 文献 [129] 将多能载波系统 (Multiple energy carrier systems, MECS) 的分布式多周期多能量运行模型调度问题描述为混合整数二阶锥规划问题,

表 4　综合能源系统管理的常规算法

Table 4　Conventional algorithm for integrated energy system management

文献	规模级别	时间尺度	算法	附加考虑
[123]	社区	实时调整	混合整数线性规划	考虑光伏生产者的随机特征和风险条件值
[125]	社区	实时调整	合作博弈	考虑各个能源枢纽自主调度和信息保密性
[126]	—	日前计划	交替方向乘子法	提升算法收敛性、实现信息保护
[127]	—	日内滚动、实时调整	博弈论	以较低的社会福利为代价显著缩短了运行时间
[128]	—	日内滚动	混合整数非线性规划	减轻计算负担
[129]	城市	日内滚动	混合整数二阶锥规划 交替方向乘子法	解决多能量网络中的强耦合和固有的非凸性 解决异构能源枢纽局部的能源自主性
[130]	城市	多时间尺度	基于多目标粒子群优化的双层元启发式算法	KPI 的应用与国家范围内的战略目标密切相关
[131]	社区	日内滚动	计算机化算法	将复杂的 EH 模型分为几个简单的 EH 模型

随后通过顺序二阶锥规划方法解决多能量网络中的强耦合和固有非凸性问题, 以确保令人满意的收敛性能. 同时考虑到相邻的异构能源枢纽的自主性, 利用一种完全分布式的基于一致性的交替方向乘子法, 仅需要相邻信息交换便可优化多能量流. 文献 [130] 提出了一种基于 EH 的双层模型: 上层领导者从大的时间范围基于输入信息和限制的功率单元数量处理能源枢纽的规划和设计问题, 在此基础上下层运营部门对各类负载进行操作分配, 然后利用基于多目标粒子群优化的双层元启发式算法使关键绩效指标 (Key performance indicators, KPI) 最小化.

本文将综合能源系统管理问题的优化目标从经济和社会两个角度进行分类. 经济角度主要包括系统建设运行维护成本、能源消费成本和能源利用率, 社会角度包括降低能耗峰均比、平整负荷曲线提升能源网络稳定性、提升用户满意度以及环境友好性. 鉴于文献 [120] 已进行这方面的总结, 在此不再赘述. 此外, 隐私保护[125−126] 和减轻计算负担[127−128, 131] 也被纳入考虑范围.

3.2 基于强化学习的综合能源系统管理

基于前文提到的强化学习具有无模型依赖性、环境适应性等优点, 本节聚焦于利用强化学习算法求解综合能源系统管理问题. 首先简要介绍综合能源系统中的马尔科夫决策过程模型, 包括系统状态 $s(t)$、动作 $a(t)$ 和奖励函数 $r(t)$ [118]

$$r(t) = -\left(P_{gC}(t) + P_{gB}(t)\right) \times Pr_g + P_{gC}(t) \times \eta_e^C \times Pr_e(t) \tag{11a}$$

$$s(t) = Pr_e(t) \tag{11b}$$

$$a(t) = P_{gC}(t) \tag{11c}$$

其中, 式 (11a) 是由耗能成本的相反数构成的奖励函数, 第 1 项为天然气成本, 由输入热电联产系统和输入锅炉的天然气总量 $P_{gC}(t)$、$P_{gB}(t)$ 与天然气价格 Pr_g 决定, 第 2 项是由天然气转化为电能进而节约的电力成本, η_e^C 是天然气经热电联产系统转化为电能的效率, $Pr_e(t)$ 是时变电价; 式 (11b) 是由时变电价 $Pr_e(t)$ 构成的系统状态; 式 (11c) 是由输入热电联产系统的天然气量 $P_{gC}(t)$ 构成的动作[118]. 此外, 当降低碳排放、提升用户用能满意度也作为优化目标时, 相应的奖励函数也应该考虑这些因素, 例如加上碳排放成本和用户不满意成本的相反数作为新的奖励函数[132].

在本节中, 综合能源系统管理的优化目标和优化变量仍然没有大的改变, 即问题的背景、难点与第 3.1 节一致. 但是深度强化学习算法的引进增强了面对无精确模型[118, 133]、可变场景[134−135]、多重不确定性[117] 等情况的求解能力. 表 5 从综合能源系统管理的优化目标、强化学习算法及性能这三个方面进行了总结. 其中由于经济性目标是普遍存在的, 因此表 5 主要对社会性目标进行总结, 具体包括用户满意度、环境友好性以及负荷平滑程度. Ye 等[133] 提出了一种不依赖于模型的优先深度确定性策略梯度方法来求解住宅综合能源系统实时自主能源管理策略, 该方法用 TD 误差的大小来衡量 Q 值估计的准确度并指导学习. TD 误差表明了一个智能体可以从一次试错中学到知识的效果, 较大的正 TD 误差表明这是一次非常成功的尝试, 而较大的负 TD 误差表明智能体的此次尝试是失败的. 在训练期间对这些经验的重演进行优先级排序可以使智能体基于

表 5 基于强化学习的综合能源系统管理

Table 5 Integrated energy system management based on reinforcement learning

文献	社会性目标	求解算法	算法性能		
			计算速度	适应性	备注
[118]	—	蒙特卡洛法	√		收敛速度快
[119]	环境友好	Q 学习	√		
[133]	—	优先深度确定性策略梯度	√		
[134]	—	分布式近端策略优化		√	保证收敛性
[135]	负荷平滑	深度确定性策略梯度		√	
[136]	—	人工神经网络 + 强化学习			同时优化能源枢纽系统设计和运行策略
[137]	用户满意	置信域策略梯度算法 + 深度确定性策略梯度			DDPG 得到的策略更优、两者都无法一步获得最优配置和控制策略
[138]	—	多智能体议价学习 + 强化学习	√		较强的全局搜索能力, 能处理大型复杂的能源系统分布式优化问题
[139]	用户满意	深度双神经拟合 Q 迭代	√		提高鲁棒性, 无模型算法性能不及基于模型的算法
[140]	环境友好	演员-评论家算法	√		有较好的稳定性

成功的尝试更快地优化策略, 防止其选择某些状态下的不利动作, 从而提高策略学习的质量与效率. 文献 [118] 在住宅智能能源枢纽中采用蒙特卡洛方法来寻找近似最佳的解决方案以降低运营成本.

在场景适应性方面, Zhou 等[134] 利用分布式近端策略优化算法训练智能体以探索热电联产系统的最佳经济调度, 并且能够自适应地学习不同场景下的优化管理策略. 文献 [135] 采用 DDPG 方法解决动态能量转换和管理决策问题, 系统运营商基于在线过程自适应地协调电气装置和发电机的运行, 进而平滑电力和天然气的净负荷曲线, 同时兼顾了经济性目标. 在能源价格不确定的条件下, Hua 等[117] 提出了条件随机场方法来分析能源的动态价格弹性, 基于这些内在特征设计了强化学习算法来调度电力和天然气的交换以及能源枢纽的能源分配, 以减小运营成本和降低碳排放. 此外, 文献 [119] 为了满足 We-energy 的功率和热能需求, 同时实现运营成本最小化和降低污染物排放, 在智能能源管理系统中将 Q 学习算法与资格迹理论结合以获得最优策略并加快计算速度.

优化问题除了常见的能源设备运行策略外, 不少文献还考虑了能源枢纽系统的设计和配置问题. 文献 [136] 提出了一种基于强化学习的双层调度策略, 用于同时优化 EH 系统的设计和运行策略. 文献 [137] 中智能体通过强化学习方法找到 EH 的最佳配置, 即燃气轮机、熔炉、变压器和存储设备的组合以及这些设备的最佳控制策略, 最大程度地降低设备总成本和单位成本, 同时满足用户的电热负载需求.

在能源管理优化决策问题中, 当面临的问题模型是单智能体时, 智能体所在的环境是相对稳定不变、可预测的. 但是在多智能体强化学习中, 例如多微网优化管理、多个家庭能源交易或多种能源调度的问题[141], 环境是复杂的、动态的, 给学习训练带来很大的挑战. 而且多智能体之间可能包含合作与竞争等多重关系, 例如在选择能源种类时, 存在不同种类能源供给之间的博弈; 在制定能源价格时, 存在多个能源供应商之间的博弈; 在优化购买能源策略时, 存在能源供应商和能源消费者之间的博弈[142]. 因此引入博弈的概念, 将博弈论与强化学习相结合可以很好地处理这些问题. Zhang 等[138] 针对多能载波系统的分布式能源枢纽经济调度问题 (Energy hub economic dispatch, EHED) 提出了一种多智能体议价学习方法. 每个智能体利用带联想记忆的经典 Q 学习获取知识, 买方与卖方利用讨价还价博弈的方法进行有效协调, 从而实现所有能源枢纽的总收益最大化. 对于分布式的 EHED, 每个能源枢纽都可以看作是讨价还价博弈过程中的一个参与者, 在该模型中具有最多种输出能量类型的枢纽可以被选择作为卖方, 卖方智能体只负责对不同买方报价, 相比之下, 每个买方智能体不仅需要与卖方进行谈判, 还需要搜索潜在的更优解决方案.

在住宅级别, 智能电表的推出和智能设备的快速部署是综合能源系统自治的基础, 该系统可以利用智能电表提供的实时信息来优化调度不同智能设备的运行, 从而最大程度地减少终端用户能源成本. 但是在耦合了多种能源的综合能源系统中, 隐私保护问题仍然值得关注. 与单一电能的管理类似, 从模型角度考虑, 可以建立分层马尔科夫决策过程模型或加入噪声以掩盖用户隐私信息[64, 83]; 从优化目标角度考虑, 可以平整负荷曲线来隐藏耗能信息[84–85]; 从算法角度考虑, 强化学习算法在不需要用户用能数据的情况下, 从与环境交互获得的奖励回报中学习到最优能源管理策略, 可以一定程度上保护用户隐私[139]. 在综合能源系统管理中, 分布式优化算法是一种较为常见的保障信息私密性的方法[125–126].

除了住宅级别综合能源系统, 基于强化学习的综合能源系统管理方案还能用于建筑物供暖系统和更复杂的工业场景中. 文献 [139] 提出了双深度神经拟合 Q 迭代方法控制建筑物室内温度, 在降低能耗和成本的同时确保居住者舒适, 该算法不仅有更短的计算时间, 而且能提高对建筑物动态非平稳过程的鲁棒性. Wang 等[140] 针对钢铁行业综合能源系统中各类能源输入量的优化问题, 提出了基于演员–评论家的分层优化模型及循环求解方法. 该方法既能解决非线性约束, 又可以有效获得最优能源分配方案, 降低生产钢的能耗并确保气体排放达标.

4 展望

本文所提到的综合能源系统管理优化问题的求解难度体现在系统的高度不确定性、难以建立精确的系统模型、维度灾难以及变量耦合等方面. 分层马尔科夫决策过程是一种求解具有高维变量问题的思路, 而且能一定程度上保护用户隐私信息, 适用于有隐私保护需求的高维综合能源系统管理优化问题. 强化学习由于不具有模型依赖性, 可以在没有先验知识的情况下通过与环境交互进行学习, 解决新能源发电和用户用能需求的不确定性带来的问题, 同时深度神经网络的引入还可以解决维度灾难和复杂优化变量耦合的问题, 因此深度强化学习在求解具有复杂动态特性的综合能源系统管理问题中具有极大潜力. 然而, 强化学习方法也具有一定的

局限性, 例如学习性能很大程度上依赖于人为设计的奖励函数, 降低了可解释性, 而且奖励函数还需要适用于不同种类能源和具有不同特性用能设备的学习, 设计存在一定的困难. 在电能的优化管理方面, 尤其是与家庭能源系统和电动汽车相关的研究中, 强化学习算法已经是一种常见的求解方法, 具有卓越的性能. 然而在综合能源系统中, 传统算法仍然是主流, 未来可以更多地尝试将具有强大自主学习能力的强化学习方法用于解决具有复杂动态特性的综合能源系统优化调度问题.

结合现有的强化学习和深度强化学习在综合能源系统管理中的研究进展和研究趋势, 下面将从多时间尺度特性、可解释性、迁移性和信息安全性 4 个方面对综合能源系统管理问题进行展望.

4.1 多时间尺度特性

日前调度虽然计算相对比较简单, 但由于时间尺度较大, 而且综合能源系统存在较大的不确定性, 在对综合能源系统的实际管理中计划情况可能会与实际情况发生较大偏差, 导致优化效果不佳. 因此考虑更为复杂的日内滚动优化、实时调整或者三者相互结合的多时间尺度优化, 这样能更加准确地对实际情况进行预估. 但是这同时会导致计算量增加, 计算时间成本上升, 难以满足综合能源系统管理实时性的要求. 强化学习方法能在特定场景下对智能体进行针对性训练, 当该场景下的参数随着时间推移发生变化时, 训练好的智能体也能快速求得最优管理策略, 从而提高算法效率以达到实时性的要求, 因此与强化学习算法相结合的多时间尺度优化可以得到更好的应用.

4.2 可解释性

可解释性是近年来专家学者讨论比较多的一个话题, 在综合能源系统中, 能源管理的策略最终是面向用户的, 可解释性的提高能够增加社会的接受度[143]. 其中解释性是指人们能够理解人工智能算法所作的决策, 也就是基于对模型特征、结构和相关参数的整体认知来理解算法如何作出决策. 从这个层面上讲, 由于基于强化学习的各种衍生算法都是基于策略迭代和策略提升的原理逐步演变而来, 不同的网络结构和目标函数分别解决什么样的问题都已阐明, 具有强逻辑性和强可解释性. 但由于在面对一些具有连续动作和状态空间的综合能源系统管理问题时引入了深度学习, 用数据驱动的神经网络来拟合策略函数、值函数; 在面对新能源发电和用户耗能需求不确定性时, 一些基于强化学习的方法也用到深度网络对不确定因素进行预测, 这都使强化学习在能源管理问题中的可解释性受到一定程度的影响. 因此, 如何提升深度强化学习的可解释性是未来深度强化学习方法应用于实际综合能源系统管理中要面临的一个重要问题.

4.3 迁移性

不论在电力系统还是综合能源系统中, 能源管理优化问题都可能遇到仅有的少量数据不足以支持完成网络训练的情况. 数据量不足的可能原因主要有两种: 1) 在综合能源系统中, 由于系统规模较大, 所涉及的设备较多, 数据收集复杂且昂贵, 出于技术和成本的原因, 综合能源系统本身无法提供大量的数据; 2) 随着时间推移, 综合能源系统迅速发展, 当系统中的某些设备或用户用能偏好发生变化时, 原有的数据不再包含充足的实时有效信息[144]. 基于综合能源系统中的旧场景和历史数据花费大量时间训练得到的网络无法在新场景中作出最优决策, 需要再次利用大量时间和实时数据进行重新学习. 因此, 在综合能源系统的管理问题中, 如何利用先验知识和少量数据进行学习是当下研究热门.

深度学习具有严重的数据依赖性, 加速学习过程是强化学习方法面临的一个重要问题. 在机器学习中, 迁移学习作为一种运用相似任务已经训练好的网络中包含的知识来求解目标任务的方法, 主要思想为: 解决类似任务的知识会加速目标任务的学习过程, 并且在类似任务数据充足的前提下有效降低对目标任务的数据依赖[145]. 由此可以看出, 迁移学习可以解决综合能源系统中的跨任务学习问题, 对于出现的新的能源管理任务体现了时效性优势, 而且降低了对目标任务的数据依赖性.

迁移学习过程中, 利用目标任务数据对迁移过来的相似任务网络进行训练或者微调, 源任务与目标任务之间越相似, 迁移就越容易, 迁移效果也越好[146]. 由此可见, 这种方法局限于相似任务间的迁移, 而不能用于学习全新的任务, 因此针对经常发生变化的综合能源系统管理问题, 进一步可以考虑使用元学习. 通俗地讲, 元学习通过研究如何让神经网络充分利用旧的综合能源系统中获得的知识经验来指导新系统中的学习任务, 使得神经网络能针对新系统中的能源管理任务进行适当调整, 从而具有学会学习的能力[147]. 一个好的元学习模型能够很好地推广到从未遇到过的新的综合能源系统管理场景中, 最终经过模型的自我调整可以完成新的综合能源系统管理任务. 其中小样本学习是元学习的一种典型方法[148], 可以克服综合能源系统中数据样本少的困难, 并降低数据采集成本. 此外, 元学习还可

以与强化学习结合构成元强化学习, 减少强化学习方法对超参数、策略网络参数、奖励函数等的依赖[147]. 基于此, 未来在综合能源系统管理优化问题中, 可以通过迁移学习、小样本学习甚至元学习与深度强化学习相结合来解决迁移性的问题, 同时克服数据依赖并加快学习过程.

4.4 信息安全性

信息技术的发展使得人们对信息安全问题越来越重视. 随着智能电表和智能设备的发展, 人们的用电偏好和习惯包含在用户数据信息中, 并可以随时被获取, 如何掩盖这些信息成为新的研究热点. 由于在处理具有不完全信息的优化问题中的突出表现, 强化学习方法在不需要新能源发电和用户用能数据的情况下, 通过与环境交互获得的奖励回报中学习到最优能源管理策略, 一定程度上保护了用户隐私信息, 提升信息安全性[22, 43, 47, 65].

5 总结

本文综述了基于强化学习的综合能源系统管理优化研究. 首先从模型角度将综合能源系统管理问题分为对单一电能的管理和对综合能源的管理. 在电能管理问题中, 分别从微电网、智能家庭和电动汽车三个方面进行阐述, 总结发现相较于传统优化求解方法, 强化学习在解决没有先验知识且具有多重不确定性的优化问题中具有突出表现. 当多种能源通过耦合技术相互转换、相互连接形成综合能源系统之后, 由于变量之间相互耦合, 不同种类的能源具有不同的特性使得场景变得更加复杂. 此时在对比传统求解算法的基础上, 对已有的基于强化学习的相关文献进行分析, 结果表明强化学习在求解综合能源系统管理问题时具有卓越性能. 最后本文对综合能源系统管理问题进行展望, 得益于人工智能的发展, 利用深度强化学习算法能够处理具有高维变量的复杂动态系统优化问题. 未来能源管理中多时间尺度特性、可解释性、迁移性和信息安全性的问题将得到人们越来越多的重视, 相应的多时间尺度优化、机理知识与数据驱动相融合的方法以及迁移学习、元学习等算法也将与强化学习算法相结合, 用于综合能源系统管理优化问题.

References

1 Sun Qiu-Ye, Teng Fei, Zhang Hua-Guang. Energy internet and its key control issues. *Acta Automatica Sinica*, 2017, **43**(2): 176−194
(孙秋野, 滕菲, 张化光. 能源互联网及其关键控制问题. 自动化学报, 2017, **43**(2): 176−194)

2 Yang T, Zhao L Y, Li W, Zomaya A Y. Reinforcement learning in sustainable energy and electric systems: A survey. *Annual Reviews in Control*, 2020, **49**: 145−163

3 National Development and Reform Commission and National Energy Administration. Revolutionary strategy of energy production and consumption (2016−2030). *China Electrical Equipment Industry*, 2017(5): 39−47
(国家发改委和国家能源局. 能源生产和消费革命战略 (2016−2030). 电器工业, 2017(5): 39−47)

4 National Development and Reform Commission. Medium and long term special plan for energy conservation. *Energy Conservation and Environmental Protection*, 2004(11): 3−10
(国家发展和改革委员会. 节能中长期专项规划. 节能与环保, 2004(11): 3−10)

5 Ping Zuo-Wei, He Wei, Li Jun-Lin, Yang Tao. Sparse learning for load modeling in microgrids. *Acta Automatica Sinica*, 2020, **46**(9): 1798−1808
(平作为, 何维, 李俊林, 杨涛. 基于稀疏学习的微电网负载建模. 自动化学报, 2020, **46**(9): 1798−1808)

6 Yao W F, Zhao J H, Wen F S, Dong Z Y, Xue Y S, Xu Y, et al. A multi-objective collaborative planning strategy for integrated power distribution and electric vehicle charging systems. *IEEE Transactions on Power Systems*, 2014, **29**(4): 1811−1821

7 Moghaddam M P, Damavandi M Y, Bahramara S, Haghifam M R. Modeling the impact of multi-energy players on electricity market in smart grid environment. In: Proceedings of the 2016 IEEE Innovative Smart Grid Technologies-Asia (ISGT-Asia). Melbourne, Australia: IEEE, 2016. 454−459

8 Carli R, Dotoli M. Decentralized control for residential energy management of a smart users′ microgrid with renewable energy exchange. *IEEE/CAA Journal of Automatica Sinica*, 2019, **6**(3): 641−656

9 Farrokhifar M, Aghdam F H, Alahyari A, Monavari A, Safari A. Optimal energy management and sizing of renewable energy and battery systems in residential sectors via a stochastic MILP model. *Electric Power Systems Research*, 2020, **187**: Article No. 106483

10 Moser A, Muschick D, Golles M, Nageler P, Schranzhofer H, Mach T, et al. A MILP-based modular energy management system for urban multi-energy systems: Performance and sensitivity analysis. *Applied Energy*, 2020, **261**: Article No. 114342

11 Alipour M, Zare K, Abapour M. MINLP probabilistic scheduling model for demand response programs integrated energy hubs. *IEEE Transactions on Industrial Informatics*, 2018, **14**(1): 79−88

12 Sadeghianpourhamami N, Demeester T, Benoit D F, Strobbe M, Develder C. Modeling and analysis of residential flexibility: Timing of white good usage. *Applied Energy*, 2016, **179**: 790−805

13 Bruce J, Sunderhauf N, Mirowski P, Hadsell R, Milford M. One-shot reinforcement learning for robot navigation with interactive replay. arXiv: 1711.10137, 2017.

14 Mnih V, Kavukcuoglu K, Silver D, Graves A, Antonoglou I, Wierstra D, et al. Playing Atari with deep reinforcement learning. arXiv: 1312.5602, 2013.

15 Wang X, Huang Q Y, Celikyilmaz A, Gao J F, Shen D H, Wang Y F, et al. Reinforced cross-modal matching and self-supervised imitation learning for vision-language navigation. In: Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, USA: IEEE, 2019. 6622−6631

16 Segler M H S, Preuss M, Waller M P. Planning chemical syntheses with deep neural networks and symbolic AI. *Nature*, 2018, **555**(7698): 604−610

17 Hua H C, Qin Y C, Hao C T, Cao J W. Optimal energy management strategies for energy internet via deep reinforcement learning approach. *Applied Energy*, 2019, **239**: 598−609

18 Kim S, Lim H. Reinforcement learning based energy management algorithm for smart energy buildings. *Energies*, 2018,

11(8): Article No. 2010

19 Liu T, Hu X S, Li S E, Cao D P. Reinforcement learning optimized look-ahead energy management of a parallel hybrid electric vehicle. *IEEE/ASME Transactions on Mechatronics*, 2017, **22**(4): 1497−1507

20 Kong W C, Dong Z Y, Jia Y W, Hill D J, Xu Y, Zhang Y. Short-term residential load forecasting based on LSTM recurrent neural network. *IEEE Transactions on Smart Grid*, 2019, **10**(1): 841−851

21 Xu X, Jia Y W, Xu Y, Xu Z, Chai S J, Lai C S. A multi-agent reinforcement learning-based data-driven method for home energy management. *IEEE Transactions on Smart Grid*, 2020, **11**(4): 3201−3211

22 Zhang Q Z, Dehghanpour K, Wang Z Y, Huang Q H. A learning-based power management method for networked microgrids under incomplete information. *IEEE Transactions on Smart Grid*, 2020, **11**(2): 1193−1204

23 Sun Chang-Yin, Mu Chao-Xu. Important scientific problems of multi-agent deep reinforcement learning. *Acta Automatica Sinica*, 2020, **46**(7): 1301−1312
(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. 自动化学报, 2020, **46**(7): 1301−1312)

24 Zou H L, Mao S W, Wang Y, Zhang F H, Chen X, Cheng L. A survey of energy management in interconnected multi-microgrids. *IEEE Access*, 2019, **7**: 72158−72169

25 Yu L, Qin S Q, Zhang M, Shen C, Jiang T, Guan X H. Deep reinforcement learning for smart building energy management: A survey. arXiv: 2008.05074, 2020.

26 Cao D, Hu W H, Zhao J B, Zhang G Z, Zhang B, Liu Z, et al. Reinforcement learning and its applications in modern power and energy systems: A review. *Journal of Modern Power Systems and Clean Energy*, 2020, **8**(6): 1029−1042

27 Sutton R S, Barto A G. *Reinforcement Learning: An Introduction (Second Edition)*. Cambridge: MIT Press, 2018.

28 Wang F Y, Zhang H G, Liu D R. Adaptive dynamic programming: An introduction. *IEEE Computational Intelligence Magazine*, 2009, **4**(2): 39−47

29 Tesauro G. Temporal difference learning and TD-Gammon. *Communications of the ACM*, 1995, **38**(3): 58−68

30 Tokic M. Adaptive ε-greedy exploration in reinforcement learning based on value differences. In: Proceedings of the 33rd Annual German Conference on Artificial Intelligence. Karlsruhe, Germany: Springer Press, 2010. 203−210

31 Zhang F Y, Yang Q Y, An D. CDDPG: A deep-reinforcement-learning-based approach for electric vehicle charging control. *IEEE Internet of Things Journal*, 2021, **8**(5): 3075−3087

32 Konda V R, Tsitsiklis J N. On actor-critic algorithms. *SIAM Journal on Control and Optimization*, 2003, **42**(4): 1143−1166

33 Yu L, Sun Y, Xu Z B, Shen C, Yue D, Jiang T, et al. Multi-agent deep reinforcement learning for HVAC control in commercial buildings. *IEEE Transactions on Smart Grid*, 2021, **12**(1): 407−419

34 Van Hasselt H, Guez A, Silver D. Deep reinforcement learning with double Q-learning. In: Proceedings of the 30th AAAI Conference on Artificial Intelligence. Phoenix, USA: AAAI, 2016. 2094−2100

35 Wang Z Y, Schaul T, Hessel M, Van Hasselt H, Lanctot M, De Freitas N. Dueling network architectures for deep reinforcement learning. In: Proceedings of the 33rd International Conference on Machine Learning. New York, USA: PMLR Press, 2016. 1995−2003

36 Babaeizadeh M, Frosio I, Tyree S, Clemons J, Kautz J. Reinforcement learning through asynchronous advantage actor-critic on a GPU. In: Proceedings of the 5th International Conference on Learning Representations. Toulon, France: OpenReview.net, 2017.

37 Schulman J, Levine S, Abbeel P, Jordan M I, Moritz P. Trust region policy optimization. In: Proceedings of the 32nd International Conference on Machine Learning. Lille, France: PMLR Press, 2015. 1889−1897

38 Schulman J, Wolski F, Dhariwal P, Radford A, Klimov O. Proximal policy optimization algorithms. arXiv: 1707.06347, 2017.

39 Mnih V, Kavukcuoglu K, Silver D, Rusu A A, Veness J, Bellemare M G, et al. Human-level control through deep reinforcement learning. *Nature*, 2015, **518**(7540): 529−533

40 Lillicrap T P, Hunt J J, Pritzel A, Heess N, Erez T, Tassa Y, et al. Continuous control with deep reinforcement learning. arXiv: 1509.02971, 2015.

41 Gorostiza F S, Gonzalez-Longatt F M. Deep reinforcement learning-based controller for SOC management of multi-electrical energy storage system. *IEEE Transactions on Smart Grid*, 2020, **11**(6): 5039−5050

42 Liang Y C, Guo C L, Ding Z H, Hua H C. Agent-based modeling in electricity market using deep deterministic policy gradient algorithm. *IEEE Transactions on Power Systems*, 2020, **35**(6): 4180−4192

43 Lei L, Tan Y, Dahlenburg G, Xiang W, Zheng K. Dynamic energy dispatch based on deep reinforcement learning in IoT-driven smart isolated microgrids. *IEEE Internet of Things Journal*, 2021, **8**(10): 7938−7953

44 Li H P, Wan Z Q, He H B. Real-time residential demand response. *IEEE Transactions on Smart Grid*, 2020, **11**(5): 4144−4154

45 Yu L, Xie W W, Xie D, Zou Y L, Zhang D Y, Sun Z X, et al. Deep reinforcement learning for smart home energy management. *IEEE Internet of Things Journal*, 2020, **7**(4): 2751−2762

46 Zhang C, Xu Y, Dong Z Y, Wong K P. Robust coordination of distributed generation and price-based demand response in Microgrids. *IEEE Transactions on Smart Grid*, 2018, **9**(5): 4236−4247

47 Latifi M, Rastegarnia A, Khalili A, Bazzi W M, Sanei S. A self-governed online energy management and trading for smart Micro/Nano-grids. *IEEE Transactions on Industrial Electronics*, 2020, **67**(9): 7484−7498

48 Du Y, Li F X. Intelligent multi-microgrid energy management based on deep neural network and model-free reinforcement learning. *IEEE Transactions on Smart Grid*, 2020, **11**(2): 1066−1076

49 Hafeez G, Alimgeer K S, Wadud Z, Khan I, Usman M, Qazi A B, et al. An innovative optimization strategy for efficient energy management with day-ahead demand response signal and energy consumption forecasting in smart grid using artificial neural network. *IEEE Access*, 2020, **8**: 84415−84433

50 Yu Y J, Cai Z F, Huang Y S. Energy storage arbitrage in grid-connected micro-grids under real-time market price uncertainty: A double-Q learning approach. *IEEE Access*, 2020, **8**: 54456−54464

51 Thirugnanam K, Kerk S K, Yuen C, Liu N, Zhang M. Energy management for renewable microgrid in reducing diesel generators usage with multiple types of battery. *IEEE Transactions on Industrial Electronics*, 2018, **65**(8): 6772−6786

52 Morstyn T, Hredzak B, Agelidis V G. Control strategies for microgrids with distributed energy storage systems: An overview. *IEEE Transactions on Smart Grid*, 2018, **9**(4): 3652−3666

53 Jin J L, Xu Y J, Khalid Y, Hassan N U. Optimal operation of energy storage with random renewable generation and AC/DC loads. *IEEE Transactions on Smart Grid*, 2018, **9**(3): 2314−2326

54 Bouakkaz A, Mena A J G, Haddad S, Ferrari M L. Efficient energy scheduling considering cost reduction and energy saving in hybrid energy system with energy storage. *Journal of Energy Storage*, 2021, **33**: Article No. 101887

55 Luo F J, Meng K, Dong Z Y, Zheng Y, Chen Y Y, Wong K P. Coordinated operational planning for wind farm with battery energy storage system. *IEEE Transactions on Sustainable Energy*, 2015, **6**(1): 253−262

56 Wang B, Li Y, Ming W Y, Wang S R. Deep reinforcement learning method for demand response management of interruptible load. *IEEE Transactions on Smart Grid*, 2020, **11**(4): 3146−3155

57 Zhang Y, Van Der Schaar M. Structure-aware stochastic storage management in smart grids. *IEEE Journal of Selected Topics in Signal Processing*, 2014, **8**(6): 1098−1110

58 Lu R Z, Hong S H, Zhang X F. A dynamic pricing demand response algorithm for smart grid: Reinforcement learning approach. *Applied Energy*, 2018, **220**: 220−230

59 Rasheed M B, Qureshi M A, Javaid N, Alquthami T. Dynamic pricing mechanism with the integration of renewable energy source in smart grid. *IEEE Access*, 2020, **8**: 16876−16892

60 Yang H M, Zhang J, Qiu J, Zhang S H, Lai M Y, Dong Z Y. A practical pricing approach to smart grid demand response based on load classification. *IEEE Transactions on Smart Grid*, 2018, **9**(1): 179−190

61 Marquant J F, Evins R, Bollinger L A, Carmeliet J. A holarchic approach for multi-scale distributed energy system optimisation. *Applied Energy*, 2017, **208**: 935−953

62 Wang Q, Wu H Y, Florita A R, Martinez-Anido C B, Hodge B M. The value of improved wind power forecasting: Grid flexibility quantification, ramp capability analysis, and impacts of electricity market operation timescales. *Applied Energy*, 2016, **184**: 696−713

63 Shu J, Guan R, Wu L, Han B. A Bi-level approach for determining optimal dynamic retail electricity pricing of large industrial customers. *IEEE Transactions on Smart Grid*, 2019, **10**(2): 2267−2277

64 Mirzaei M, Keypour R, Savaghebi M, Golalipour K. Probabilistic optimal bi-level scheduling of a multi-microgrid system with electric vehicles. *Journal of Electrical Engineering & Technology*, 2020, **15**(6): 2421−2436

65 Dadashi-Rad M H, Ghasemi-Marzbali A, Ahangar R A. Modeling and planning of smart buildings energy in power system considering demand response. *Energy*, 2020, **213**: Article No. 118770

66 Liu D R, Xu Y C, Wei Q L, Liu X L. Residential energy scheduling for variable weather solar energy based on adaptive dynamic programming. *IEEE/CAA Journal of Automatica Sinica*, 2018, **5**(1): 36−46

67 Li Y Z, Wang P, Gooi H B, Ye J, Wu L. Multi-objective optimal dispatch of microgrid under uncertainties via interval optimization. *IEEE Transactions on Smart Grid*, 2019, **10**(2): 2046−2058

68 Bao Z J, Qiu W R, Wu L, Zhai F, Xu W J, Li B F, et al. Optimal multi-timescale demand side scheduling considering dynamic scenarios of electricity demand. *IEEE Transactions on Smart Grid*, 2019, **10**(3): 2428−2439

69 Qazi H S, Liu N, Wang T. Coordinated energy and reserve sharing of isolated microgrid cluster using deep reinforcement learning. In: Proceedings of the 5th Asia Conference on Power and Electrical Engineering (ACPEE). Chengdu, China: IEEE, 2020. 81−86

70 Jayaraj S, Ahamed T P I, Kunju K B. Application of reinforcement learning algorithm for scheduling of microgrid. In: Proceedings of the 2019 Global Conference for Advancement in Technology (GCAT). Bangalore, India: IEEE, 2019. 1−5

71 Mao S, Tang Y, Dong Z W, Meng K, Dong Z Y, Qian F. A privacy preserving distributed optimization algorithm for economic dispatch over time-varying directed networks. *IEEE Transactions on Industrial Informatics*, 2021, **17**(3): 1689−1701

72 Parag Y, Sovacool B K. Electricity market design for the prosumer era. *Nature Energy*, 2016, **1**(4): Article No. 16032

73 Ruan L N, Yan Y, Guo S Y, Wen F S, Qiu X S. Priority-based residential energy management with collaborative edge and cloud computing. *IEEE Transactions on Industrial Informatics*, 2020, **16**(3): 1848−1857

74 Wei Q L, Liu D R, Liu Y, Song R Z. Optimal constrained self-learning battery sequential management in microgrid via adaptive dynamic programming. *IEEE/CAA Journal of Automatica Sinica*, 2017, **4**(2): 168−176

75 Lee J, Wang W B, Niyato D. Demand-side scheduling based on deep actor-critic learning for smart grids. arXiv: 2005.01979, 2020.

76 Mocanu E, Mocanu D C, Nguyen P H, Liotta A, Webber M E, Gibescu M, et al. On-line building energy optimization using deep reinforcement learning. *IEEE Transactions on Smart Grid*, 2019, **10**(4): 3698−3708

77 Diyan M, Khan M, Cao Z B, Silva B N, Han J H, Han K J. Intelligent home energy management system based on Bi-directional long-short term memory and reinforcement learning. In: Proceedings of the 2021 International Conference on Information Networking (ICOIN). Jeju Island, South Korea: IEEE, 2021. 782−787

78 Lu R Z, Hong S H, Yu M M. Demand response for home energy management using reinforcement learning and artificial neural network. *IEEE Transactions on Smart Grid*, 2019, **10**(6): 6629−6639

79 Huang Z T, Chen J P, Fu Q M, Wu H J, Lu Y, Gao Z. HVAC optimal control with the multistep-actor critic algorithm in large action spaces. *Mathematical Problems in Engineering*, 2020, **2020**: Article No. 1386418

80 Yoon A Y, Kim Y J, Moon S I. Optimal retail pricing for demand response of HVAC systems in commercial buildings considering distribution network voltages. *IEEE Transactions on Smart Grid*, 2019, **10**(5): 5492−5505

81 Chen Z, Wu L. Residential appliance DR energy management with electric privacy protection by online stochastic optimization. *IEEE Transactions on Smart Grid*, 2013, **4**(4): 1861−1869

82 Li B Y, Wu J, Shi Y Y. Privacy-aware cost-effective scheduling considering non-schedulable appliances in smart home. In: Proceedings of the 2019 IEEE International Conference on Embedded Software and Systems (ICESS). Las Vegas, USA: IEEE, 2019. 1−8

83 Rottondi C, Barbato A, Chen L, Verticale G. Enabling privacy in a distributed game-theoretical scheduling system for domestic appliances. *IEEE Transactions on Smart Grid*, 2017, **8**(3): 1220−1230

84 Chang H H, Chiu W Y, Sun H J, Chen C M. User-centric multiobjective approach to privacy preservation and energy cost minimization in smart home. *IEEE Systems Journal*, 2019, **13**(1): 1030−1041

85 Kement C E, Gultekin H, Tavli B, Girici T, Uludag S. Comparative analysis of load-shaping-based privacy preservation strategies in a smart grid. *IEEE Transactions on Industrial Informatics*, 2017, **13**(6): 3226−3235

86 Wu Y, Zhang J, Ravey A, Chrenko D, Miraoui A. Real-time energy management of photovoltaic-assisted electric vehicle charging station by Markov decision process. *Journal of Power Sources*, 2020, **476**: Article No. 228504

87 Li Y Z, Ni Z X, Zhao T Y, Yu M H, Liu Y, Wu L, et al. Co-

ordinated scheduling for improving uncertain wind power adsorption in electric vehicles — Wind integrated power systems by multiobjective optimization approach. *IEEE Transactions on Industry Applications*, 2020, **56**(3): 2238−2250

88 Yang Y, Jia Q S, Guan X H. Stochastic coordination of aggregated electric vehicle charging with on-site wind power at multiple buildings. In: Proceedings of the 56th IEEE Annual Conference on Decision and Control (CDC). Melbourne, Australia: IEEE, 2017. 4434−4439

89 Wan C, Xu Z, Pinson P, Dong Z Y, Wong K P. Optimal prediction intervals of wind power generation. *IEEE Transactions on Power Systems*, 2014, **29**(3): 1166−1174

90 Wan C, Xu Z, Pinson P, Dong Z Y, Wong K P. Probabilistic forecasting of wind power generation using extreme learning machine. *IEEE Transactions on Power Systems*, 2014, **29**(3): 1033−1044

91 Kabir M E, Assi C, Tushar M H K, Yan J. Optimal scheduling of EV charging at a solar power-based charging station. *IEEE Systems Journal*, 2020, **14**(3): 4221−4231

92 Zhao J H, Wen F S, Dong Z Y, Xue Y S, Wong K P. Optimal dispatch of electric vehicles and wind power using enhanced particle swarm optimization. *IEEE Transactions on Industrial Informatics*, 2012, **8**(4): 889−899

93 Almalaq A, Hao J, Zhang J J, Wang F Y. Parallel building: A complex system approach for smart building energy management. *IEEE/CAA Journal of Automatica Sinica*, 2019, **6**(6): 1452−1461

94 Wang Yan-Wu, Cui Shi-Chang, Xiao Jiang-Wen, Shi Yang. A review on energy sharing for community energy prosumers. *Control and Decision*, 2020, **35**(10): 2305−2318
(王燕舞, 崔世常, 肖江文, 施阳. 社区产消者能量分享研究综述. 控制与决策, 2020, **35**(10): 2305−2318)

95 Wan Z Q, Li H P, He H B, Prokhorov D. Model-free real-time EV charging scheduling based on deep reinforcement learning. *IEEE Transactions on Smart Grid*, 2019, **10**(5): 5246−5257

96 Jin J L, Xu Y J, Yang Z Y. Optimal deadline scheduling for electric vehicle charging with energy storage and random supply. *Automatica*, 2020, **119**: Article No. 109096

97 Long T, Ma X T, Jia Q S. Bi-level proximal policy optimization for stochastic coordination of EV charging load with uncertain wind power. In: Proceedings of the 2019 IEEE Conference on Control Technology and Applications (CCTA). Hong Kong, China: IEEE, 2019. 302−307

98 Long T, Tang J X, Jia Q S. Multi-scale event-based optimization for matching uncertain wind supply with EV charging demand. In: Proceedings of the 13th IEEE Conference on Automation Science and Engineering (CASE). Xi′an, China: IEEE, 2017. 847−852

99 Yang Y, Jia Q S, Deconinck G, Guan X H, Qiu Z F, Hu Z C. Distributed coordination of EV charging with renewable energy in a microgrid of buildings. *IEEE Transactions on Smart Grid*, 2018, **9**(6): 6253−6264

100 Sadeghianpourhamami N, Deleu J, Develder C. Definition and evaluation of model-free coordination of electrical vehicle charging with reinforcement learning. *IEEE Transactions on Smart Grid*, 2020, **11**(1): 203−214

101 Li H P, Wan Z Q, He H B. Constrained EV charging scheduling based on safe deep reinforcement learning. *IEEE Transactions on Smart Grid*, 2020, **11**(3): 2427−2439

102 Ming F Z, Gao F, Liu K, Wu J, Xu Z B, Li W M. Constrained double deep Q-learning network for EVs charging scheduling with renewable energy. In: Proceedings of the 16th IEEE International Conference on Automation Science and Engineering (CASE). Hong Kong, China: IEEE, 2020. 636−641

103 Qian T, Shao C C, Wang X L, Shahidehpour M. Deep reinforcement learning for EV charging navigation by coordinating smart grid and intelligent transportation system. *IEEE Transactions on Smart Grid*, 2020, **11**(2): 1714−1723

104 Da Silva F L, Nishida C E H, Roijers D M, Costa A H R. Coordination of electric vehicle charging through multiagent reinforcement learning. *IEEE Transactions on Smart Grid*, 2020, **11**(3): 2347−2356

105 Ding T, Zeng Z Y, Bai J W, Qin B Y, Yang Y H, Shahidehpour M. Optimal electric vehicle charging strategy with Markov decision process and reinforcement learning technique. *IEEE Transactions on Industry Applications*, 2020, **56**(5): 5811−5823

106 Xu Ren-Chao, Yan Wei-Wu, Wang Guo-Liang, Yang Jian-Cheng, Zhang Xi. Time series forecasting based on seasonality modeling and its application to electricity price forecasting. *Acta Automatica Sinica*, 2020, **46**(6): 1136−1144
(徐任超, 阎威武, 王国良, 杨健程, 张曦. 基于周期性建模的时间序列预测方法及电价预测研究. 自动化学报, 2020, **46**(6): 1136−1144)

107 Wu J Z, Yan J Y, Jia H J, Hatziargyriou N, Djilali N, Sun H B. Integrated energy systems. *Applied Energy*, 2016, **167**: 155−157

108 Yu Xiao-Dan, Xu Xian-Dong, Chen Shuo-Yi, Wu Jian-Zhong, Jia Hong-Jie. A brief review to integrated energy system and energy internet. *Transactions of China Electrotechnical Society*, 2016, **31**(1): 1−13
(余晓丹, 徐宪东, 陈硕翼, 吴建中, 贾宏杰. 综合能源系统与能源互联网简述. 电工技术学报, 2016, **31**(1): 1−13)

109 Hu Xu-Guang, Ma Da-Zhong, Zheng Jun, Zhang Hua-Guang, Wang Rui. An operation state analysis method for integrated energy system based on correlation information adversarial learning. *Acta Automatica Sinica*, 2020, **46**(9): 1783−1797
(胡旭光, 马大中, 郑君, 张化光, 王睿. 基于关联信息对抗学习的综合能源系统运行状态分析方法. 自动化学报, 2020, **46**(9): 1783−1797)

110 Huang B N, Li Y S, Zhang H G, Sun Q Y. Distributed optimal co-multi-microgrids energy management for energy internet. *IEEE/CAA Journal of Automatica Sinica*, 2016, **3**(4): 357−364

111 Sun Qiu-Ye, Hu Jing-Wei, Yang Ling-Xiao, Zhang Hua-Guang. We-energy hybrid modeling and parameter Identification with GAN technology. *Acta Automatica Sinica*, 2018, **44**(5): 901−914
(孙秋野, 胡旌伟, 杨凌霄, 张化光. 基于 GAN 技术的自能源混合建模与参数辨识方法. 自动化学报, 2018, **44**(5): 901−914)

112 Qiu J, Dong Z Y, Zhao J H, Xu Y, Zheng Y, Li C X, et al. Multi-stage flexible expansion co-planning under uncertainties in a combined electricity and gas market. *IEEE Transactions on Power Systems*, 2015, **30**(4): 2119−2129

113 Sheikhi A, Rayati M, Bahrami S, Ranjbar A M. Integrated demand side management game in smart energy hubs. *IEEE Transactions on Smart Grid*, 2015, **6**(2): 675−683

114 O′Malley M J, Anwar M B, Heinen S, Kober T, McCalley J, McPherson M, et al. Multicarrier energy systems: Shaping our energy future. *Proceedings of the IEEE*, 2020, **108**(9): 1437−1456

115 He J, Yuan Z J, Yang X L, Huang W T, Tu Y C, Li Y. Reliability modeling and evaluation of urban multi-energy systems: A review of the state of the art and future challenges. *IEEE Access*, 2020, **8**: 98887−98909

116 Farah A, Hassan H, Kawabe K, Nanahara T. Optimal planning of multi-carrier energy hub system using particle swarm optimization. In: Proceedings of the 2019 IEEE Innovative Smart Grid Technologies-Asia (ISGT Asia). Chengdu, China: IEEE, 2019. 3820−3825

117 Hua W Q, You M L, Sun H J. Real-time price elasticity reinforcement learning for low carbon energy hub scheduling based on conditional random field. In: Proceedings of the 2019

IEEE/CIC International Conference on Communications Workshops in China (ICCC Workshops). Changchun, China: IEEE, 2019. 204−209

118 Rayati M, Sheikhi A, Ranjbar A M. Applying reinforcement learning method to optimize an energy hub operation in the smart grid. In: Proceedings of the 2015 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT). Washington, USA: IEEE, 2015. 1−5

119 Sun Q Y, Wang D L, Ma D Z, Huang B N. Multi-objective energy management for we-energy in energy internet using reinforcement learning. In: Proceedings of the 2017 IEEE Symposium Series on Computational Intelligence (SSCI). Honolulu, USA: IEEE, 2017. 1−6

120 Guo Xu-Tao, Han Gao-Yan, Lyu Hong-Kun. Review of modeling methods of combined cooling, heating and power system. *Zhejiang Electric Power*, 2020, **39**(4): 83−93
(国旭涛, 韩高岩, 吕洪坤. 冷热电三联供系统建模方法综述. 浙江电力, 2020, **39**(4): 83−93)

121 Van Beuzekom I, Gibescu M, Slootweg J G. A review of multi-energy system planning and optimization tools for sustainable urban development. In: Proceedings of the 2015 IEEE Eindhoven PowerTec. Eindhoven, Netherlands: IEEE, 2015. 1−7

122 Ma L, Liu N, Zhang J H, Wang L F. Real-time rolling horizon energy management for the energy-hub-coordinated prosumer community from a cooperative perspective. *IEEE Transactions on Power Systems*, 2019, **34**(2): 1227−1242

123 Paudyal S, Canizares C A, Bhattacharya K. Optimal operation of industrial energy hubs in smart grids. *IEEE Transactions on Smart Grid*, 2015, **6**(2): 684−694

124 Rastegar M, Fotuhi-Firuzabad M, Zareipour H, Moeini-Aghtaieh M. A probabilistic energy management scheme for renewable-based residential energy hubs. *IEEE Transactions on Smart Grid*, 2017, **8**(5): 2217−2227

125 Fan S L, Li Z S, Wang J H, Piao L J, Ai Q. Cooperative economic scheduling for multiple energy hubs: A bargaining game theoretic perspective. *IEEE Access*, 2018, **6**: 27777−27789

126 Yang Z, Hu J J, Ai X, Wu J C, Yang G Y. Transactive energy supported economic operation for multi-energy complementary microgrids. *IEEE Transactions on Smart Grid*, 2021, **12**(1): 4−17

127 Bahrami S, Toulabi M, Ranjbar S, Moeini-Aghtaie M, Ranjbar A M. A decentralized energy management framework for energy hubs in dynamic pricing markets. *IEEE Transactions on Smart Grid*, 2018, **9**(6): 6780−6792

128 Dolatabadi A, Jadidbonab M, Mohammadi-Ivatloo B. Short-term scheduling strategy for wind-based energy hub: A hybrid stochastic/IGDT approach. *IEEE Transactions on Sustainable Energy*, 2019, **10**(1): 438−448

129 Xu D, Wu Q W, Zhou B, Li C B, Bai L, Huang S. Distributed multi-energy operation of coupled electricity, heating, and natural gas networks. *IEEE Transactions on Sustainable Energy*, 2020, **11**(4): 2457−2469

130 Martinez R E, Perez B E Z, Reza A E, Rodriguez-Martinez A, Bravo E C, Morales W A. Optimal planning, design and operation of a regional energy mix using renewable generation. Study case: Yucatan peninsula. *International Journal of Sustainable Energy*, 2021, **40**(3): 283−309

131 Liu T H, Zhang D D, Dai H, Wu T. Intelligent modeling and optimization for smart energy hub. *IEEE Transactions on Industrial Electronics*, 2019, **66**(12): 9898−9908

132 Rayati M, Sheikhi A, Ranjbar A M. Optimising operational cost of a smart energy hub, the reinforcement learning approach. *International Journal of Parallel, Emergent and Distributed Systems*, 2015, **30**(4): 325−341

133 Ye Y J, Qiu D W, Wu X D, Strbac G, Ward J. Model-free real-time autonomous control for a residential multi-energy system using deep reinforcement learning. *IEEE Transactions on Smart Grid*, 2020, **11**(4): 3068−3082

134 Zhou S Y, Hu Z J, Gu W, Jiang M, Chen M, Hong Q T, et al. Combined heat and power system intelligent economic dispatch: A deep reinforcement learning approach. *International Journal of Electrical Power & Energy Systems*, 2020, **120**: Article No. 106016

135 Zhang B, Hu W H, Li J H, Cao D, Huang R, Huang Q, et al. Dynamic energy conversion and management strategy for an integrated electricity and natural gas system with renewable energy: Deep reinforcement learning approach. *Energy Conversion and Management*, 2020, **220**: Article No. 113063

136 Perera A T D, Nik V M, Mauree D, Scartezzini J L. Design optimization of electrical hubs using hybrid evolutionary algorithm. In: Proceedings of the ASME 10th International Conference on Energy Sustainability Collocated With the ASME 2016 Power Conference and the ASME 2016 14th International Conference on Fuel Cell Science, Engineering and Technology. Charlotte, USA: American Society of Mechanical Engineers (ASME) Press, 2016.

137 Bollenbacher J, Rhein B. Optimal configuration and control strategy in a multi-carrier-energy system using reinforcement learning methods. In: Proceedings of the 2017 International Energy and Sustainability Conference (IESC). Farmingdale, USA: IEEE, 2017. 1−6

138 Zhang X S, Yu T, Zhang Z Y, Tang J L. Multi-agent bargaining learning for distributed energy hub economic dispatch. *IEEE Access*, 2018, **6**: 39564−39573

139 Nagy A, Kazmi H, Cheaib F, Driesen J. Deep reinforcement learning for optimal control of space heating. arXiv: 1805.03777, 2018.

140 Wang Z, Wang L Q, Han Z Y, Zhao J. Multi-index evaluation based reinforcement learning method for cyclic optimization of multiple energy utilization in steel industry. In: Proceedings of the 39th Chinese Control Conference (CCC). Shenyang, China: IEEE, 2020. 5766−5771

141 Ahrarinouri M, Rastegar M, Seifi A R. Multiagent reinforcement learning for energy management in residential buildings. *IEEE Transactions on Industrial Informatics*, 2021, **17**(1): 659−666

142 Wang X C, Chen H K, Wu J, Ding Y R, Lou Q H, Liu S W. Bi-level multi-agents interactive decision-making model in regional integrated energy system. In: Proceedings of the 2019 IEEE 3rd Conference on Energy Internet and Energy System Integration (EI2). Changsha, China: IEEE, 2019. 2103−2108

143 Molnar C. Interpretable machine learning [Online], available: https://christophm.github.io/interpretable-ml-book/, June 14, 2021

144 Zhao Jin-Quan, Xia Xue, Xu Chun-Lei, Hu Wei, Shang Xue-Wei. Review on application of new generation artificial intelligence technology in power system dispatching and operation. *Automation of Electric Power Systems*, 2020, **44**(24): 1−10
(赵晋泉, 夏雪, 徐春雷, 胡伟, 尚学伟. 新一代人工智能技术在电力系统调度运行中的应用评述. 电力系统自动化, 2020, **44**(24): 1−10)

145 Takano T, Takase H, Kawanaka H, Tsuruoka S. Transfer method for reinforcement learning in same transition model-quick approach and preferential exploration. In: Proceedings of the 10th International Conference on Machine Learning and Applications and Workshops. Honolulu, USA: IEEE, 2011. 466−469

146 Shao L, Zhu F, Li X L. Transfer learning for visual categorization: A survey. *IEEE Transactions on Neural Networks and Learning Systems*, 2015, **26**(5): 1019−1034

147 Finn C, Abbeel P, Levine S. Model-agnostic meta-learning for fast adaptation of deep networks. In: Proceedings of the 34th

International Conference on Machine Learning. Sydney, Australia: ACM, 2017. 1126−1135

148 Snell J, Swersky K, Zemel R. Prototypical networks for few-shot learning. In: Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach, USA: ACM, 2017. 4080−4090

熊珞琳 华东理工大学信息科学与工程学院博士研究生. 主要研究方向为强化学习, 智能电网.
E-mail: Y11200038@mail.ecust.edu.cn
(**XIONG Luo-Lin** Ph. D. candidate at School of Information Science and Engineering, East China University of Science and Technology. Her research interest covers reinforcement learning, and smart grid.)

毛 帅 华东理工大学信息科学与工程学院博士研究生. 主要研究方向为多智能体系统, 分布式优化.
E-mail: mshecust@163.com
(**MAO Shuai** Ph. D. candidate at School of Information Science and Engineering, East China University of Science and Technology. His research interest covers multi-agent systems, and distributed optimization.)

唐 漾 博士, 华东理工大学教授. 主要研究方向为分布式估计/控制/优化, 信息物理融合系统, 混杂动力系统, 计算机视觉和强化学习.
E-mail: yangtang@ecust.edu.cn
(**TANG Yang** Ph. D., professor at East China University of Science and Technology. His research interest covers distributed estimation/control/optimization, cyber-physical systems, hybrid dynamical systems, computer vision, and reinforcement learning.)

孟 科 博士, 澳大利亚新南威尔士大学电气工程与电信学院高级讲师. 主要研究方向为电力系统建模, 稳定性分析, 可再生能源系统和电网集成.
E-mail: kemeng@ieee.org
(**MENG Ke** Ph. D., senior lecturer at the School of Electrical Engineering and Telecommunications, University of New South Wales, Australia. His research interest covers electric power system modelling, stability analysis, renewable energy systems, and grid integration.)

董朝阳 博士, 澳大利亚新南威尔士大学电气工程与电信学院能源系统教授. 主要研究方向为智能电网, 电力系统规划, 电力系统安全, 负荷建模, 电力市场和计算智能及其在电力工程中的应用.
E-mail: zydong@ieee.org
(**DONG Zhao-Yang** Ph. D., professor of energy systems at the School of Electrical Engineering and Telecommunications, University of New South Wales, Australia. His research interest covers smart grid, electric power system planning, electric power system security, load modeling, electricity market, and computational intelligence and its application in power engineering.)

钱 锋 博士, 中国工程院院士, 华东理工大学副校长. 主要研究方向为化工过程资源与能源高效利用的流程制造智能控制, 系统集成优化理论方法与关键技术研究. 本文通信作者.
E-mail: fqian@ecust.edu.cn
(**QIAN Feng** Ph. D., Academician of Chinese Academy of Engineering, the Vice President of East China University of Science and Technology. His research interest covers intelligent control of process manufacturing for efficient utilization of chemical process resources and energy, and theory, method and key technology of system integrated optimization. Corresponding author of this paper.)