

DOI: 10.19783/j.cnki.pspc.211454

# 基于深度强化学习的电动汽车充电调度算法研究进展

张延宇<sup>1</sup>, 饶新朋<sup>1</sup>, 周书奎<sup>1</sup>, 周毅<sup>2</sup>

(1. 河南大学人工智能学院, 河南 郑州 450046; 2. 河南省车联网协同技术  
国际联合实验室(河南大学), 河南 郑州 450046)

**摘要:** 对电动汽车的充电过程进行优化调度有利于电网安全稳定运行, 提升道路通行效率, 提高可再生能源利用率, 减少用户充电时间和充电费用。深度强化学习可以有效解决电动汽车充电优化调度面临的随机性和不确定性因素的影响。首先, 概述了深度强化学习的工作原理, 对比分析了不同种类强化学习的特点和应用场合。然后, 从静态充电调度和动态充电调度两方面综述了基于深度强化学习的电动汽车充电调度算法研究成果, 分析了现有研究的不足。最后, 展望了该领域未来的研究方向。

**关键词:** 智能电网; 电动汽车; 深度强化学习; 充电调度

## Research progress of electric vehicle charging scheduling algorithms based on deep reinforcement learning

ZHANG Yanyu<sup>1</sup>, RAO Xinpeng<sup>1</sup>, ZHOU Shukui<sup>1</sup>, ZHOU Yi<sup>2</sup>

(1. College of Artificial Intelligence, Henan University, Zhengzhou 450046, China; 2. International Joint Laboratory of Collaborative Technology for Internet of Vehicles of Henan Province (Henan University), Zhengzhou 450046, China)

**Abstract:** Optimal scheduling of the electric vehicle charging process is beneficial to the safe and stable operation of power grids. It improves road traffic efficiency, facilitates renewable energy utilization, and reduces the charging time and costs for users. Deep reinforcement learning can effectively solve the problems caused by different randomness and uncertainty in the optimal charging scheduling. This paper summarizes the working principle of deep reinforcement learning first, and makes the comparison of the characteristics and applications among different types of reinforcement learning. Then, the research results of deep reinforcement learning for EV charging scheduling are summarized in terms of both static and dynamic charging scheduling, and the shortcomings of existing research are analyzed. Finally, future research directions are discussed.

This work is supported by the National Natural Science Foundation of China (No. 62176088).

**Key words:** smart grid; electric vehicles; deep reinforcement learning; charging scheduling

## 0 引言

近年来, 为应对气候变化、推动绿色能源发展, 且随着技术的成熟, 电动汽车(Electric Vehicle, EV)市场占有率迅速增长。2020年11月, 国务院办公厅正式发布《新能源汽车产业发展规划(2021—2035年)》, 提出以电动汽车为代表的新能源汽车应融合互联网、人工智能等变革性技术, 从单纯交通工具向移动智能终端、储能单元转变; 加强新能源汽车

与电网能量互动以及与可再生能源高效协同发展; 推动新能源汽车与能源、信息产业的深度融合<sup>[1]</sup>。

尽管电动汽车优点显著, 但大量电动汽车无序充电不仅影响电网的安全稳定运行, 还会影响道路通行效率, 并最终损害电动汽车用户的使用体验<sup>[2]</sup>。因此, 设计合理的充电调度策略, 对电动汽车的充电过程进行协同控制十分必要。根据应用场景的不同, 充电调度可分为静态充电调度和动态充电调度。静态充电调度策略通过协调处于泊车状态的电动汽车的充/放电行为, 从而在满足用户出行需求的前提下, 实现形式多样的优化目标, 比如最小化充电费用、最大化可再生能源利用率及最小化充电时间等。

**基金项目:** 国家自然科学基金项目资助(62176088); 河南省科技攻关项目资助(212102210412)

动态充电调度策略(充电导航)对行驶中有充电需求的电动汽车进行充电路径规划,将其诱导至合适的充电站进行充电,从而减少电动汽车总充电时间,缓解充电站电网负荷压力<sup>[3]</sup>。通常情况下,充电过程中不同利益相关方的优化调度目标并不完全一致,需要综合考虑。同时,优化调度算法需要考虑众多不确定性因素的影响,如动态电价、可再生能源出力的波动、用户出行需求的不确定性 & 道路通行状况等。因此,在高度不确定性的环境下实现电动汽车充电的优化调度是一个十分有意义同时又充满挑战性的研究方向。

传统的静态调度策略主要有基于动态规划的方法<sup>[4-5]</sup>、基于日前调度的方法<sup>[6-8]</sup>和基于模型的实时调度方法<sup>[9-11]</sup>。在具有不确定性的环境下,利用动态规划方法难以获得最优充电策略。为此,研究人员提出了基于日前调度的方法,利用鲁棒优化或随机优化来最小化日前调度场景中不确定性因素对电动汽车优化调度结果的影响,但迭代计算消耗资源严重,且难以满足实时性要求。实时充电控制是保证电动汽车高效运行的关键,基于模型的实时调度策略为动态充电需求和分时电价下的电动汽车充电控制提供了实时框架,保证了用户在实时场景下的充电需求。传统的电动汽车动态充电优化算法主要是基于 Dijkstra 等最短路径算法<sup>[12-14]</sup>和仿生学算法<sup>[15-16]</sup>,并融合实时电网状态和交通信息的电动汽车路径规划策略。然而,基于模型的调度方法需要具备建立模型的先验知识,并且系统性能严重依赖于系统模型参数。上述分析表明,不确定性因素的处理是电动汽车充电调度中的难点之一。

深度强化学习(Deep Reinforcement Learning, DRL)在处理不确定性因素上具有独特优势。近年来,基于 DRL 的电动汽车充电优化调度算法得到了越来越多的关注。电动汽车充电优化调度的本质是在多种不确定因素影响下对电动汽车充/放电状态的时序进行优化,DRL 是解决这类序列决策问题的有效方法。智能体通过周期性地观察环境,做出动作,并获得奖励值,然后根据奖励值自动调整策略以寻找最优充电调度策略。

一些文献对强化学习在电力系统中的应用进行了初步的总结和归纳。文献[17]将强化学习在电动汽车充电中的应用研究分为住宅充电、集中式日前调度、分布式电动汽车充电协调和实时充电调度。文献[18]则专注于电力系统能源管理、需求响应、电力市场和运行控制 4 个方面。与这些研究不同,本文专注于基于 DRL 的电动汽车充电优化调度算

法,分别从 DRL 算法和电动汽车充电优化调度场景两个维度对现有研究成果进行分类研究,总结已经取得的成果,分析存在的问题,探讨该领域未来的研究方向。

### 1 深度强化学习

在一个强化学习(Reinforcement Learning, RL)系统中,智能体在每个时间步  $t$  通过观察得到环境状态  $s_t$ ,并根据自身策略  $\pi$  采取动作  $a_t$ ,施加于环境上,环境的状态转移到下一个状态  $s_{t+1}$ ,并向智能体返回一个奖励值  $r_t$ 。智能体通过与环境的交互来学习如何最大化长期累积奖励,如图 1 所示。该过程可建模为一个马尔可夫决策过程( $S, A, P, R, \gamma$ )。其中,  $S$  表示环境中可被感知到的所有状态集合,  $A$  为智能体可采取的所有动作集合,  $P$  是状态转移概率,  $R$  表示特定状态及动作下的即时奖励,  $\gamma \in (0,1)$  表示回报折扣率。

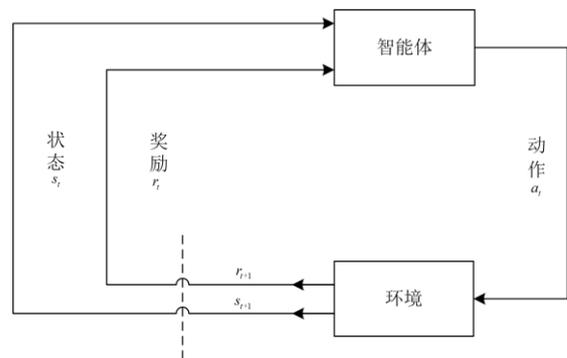


图 1 强化学习原理图

Fig. 1 Schematic diagram of reinforcement learning

深度强化学习采用神经网络来感知目标观测信息并提供当前环境中的状态信息,将当前状态映射为相应的动作,然后基于预期回报评估做出决策。按照算法学习结果的不同,深度强化学习可以分为基于价值的 DRL 算法和基于策略的 DRL 算法两大类。前者通过状态或动作的价值函数表示达到某种状态或执行某个动作后可以得到的回报,智能体倾向于选择价值最大的状态或动作,通过学习获得最优价值函数,再根据最优价值函数来做决策,选出最好的动作;而后者不需要定义价值函数,它为动作定义了概率分布,智能体按照概率分布来选取要执行的动作<sup>[19]</sup>,学习后得到最优策略函数,然后直接用得到的策略函数计算所有动作的概率值,并随机抽样选出一个动作并执行。这两类算法对比如表 1 所示。

表 1 基于价值的强化学习与基于策略的强化学习对比

Table 1 Comparison between value-based reinforcement learning and policy-based reinforcement learning

算法类别	代表算法	算法机制	算法优势	算法不足	适用场景
基于价值	Q-learning <sup>[20]</sup>	计算价值函数, 选取最大价值函数对应的动作, 隐式获得确定性策略。	样本利用率高, 价值函数估值方差小, 避免陷入局部最优。	容易出现过拟合, 处理问题复杂度受限, 收敛性较差。	离散动作空间
	SARSA <sup>[21]</sup>				
	DQN 系列 <sup>[22]</sup>				
基于策略	AC <sup>[23]</sup> 、A3C <sup>[24]</sup>	Actor(Policy-based)根据价值函数更新策略; Critic(Value-based)根据动作计算价值函数, 单步更新。	样本利用率高, 价值函数估计方差小, 整体训练速度快。	算法稳定性不足, 对超参数敏感。	离散/连续动作空间
	DPG <sup>[25]</sup> 、DDPG <sup>[26]</sup>				
	TD3 <sup>[27]</sup> 、SAC <sup>[28]</sup>				

深度 Q 网络(Deep Q Network, DQN)<sup>[22]</sup>是第一个深度强化学习算法, 是一种典型的基于价值的 DRL 算法, 得到了广泛应用。以此为基础, 人们深入研究了 DQN, 形成了一系列改进的 DQN 算法, 如解决过估计问题的 Double DQN<sup>[29]</sup>, 带有优先经验回放、可高效学习的 Prioritized Replay DQN<sup>[30]</sup>, 将环境价值和动作价值解耦的 Dueling DQN<sup>[31]</sup>等。

基于策略的 DRL 算法在处理具有连续动作空间的任務时, 比基于价值的 DRL 算法更加高效。目前最流行的基于策略的 DRL 算法是深度确定性策略梯度算法(Deep Deterministic Policy Gradient, DDPG), 该算法结合了 DQN 算法和演员-评论家(Actor-Critic, AC)算法的优点, 能够高效地解决连续动作空间的任務, 并成功应用于电动汽车充电调度等领域。

## 2 DRL 在电动汽车充电调度中的应用

如前所述, 根据电动汽车的行驶状态, 电动汽车充电调度可分为静态充电调度和动态充电调度(充电导航)。两种场景都可以使用基于价值和基于策略的深度强化学习调度方法。表 2 从算法、状态、奖励和约束条件 4 个方面对比分析了基于 DRL 的电动汽车充电调度算法。

### 2.1 在静态充电调度中的应用

#### 2.1.1 基于值函数的 EV 静态充电调度算法

在静态充电调度场景中, 实现电动汽车用户充电成本最小化是主要优化目标。由表 2 可知, 在静态充电调度中目前主要采用基于值函数的强化学习算法, 如 Q-learning 算法、DQN 算法及其各种改进 DQN 算法等。文献[32]对充电功率和电价进行了离散化, 采用 Q-learning 算法构建一个 Q 表来近似最优动作价值函数, 以找到最优充电计划。文献[33]利用基于集合状态空间的 Q 学习获得了调度策略; 通过在真实电价数据上进行验证, 展示了该调度模式在节约成本方面的优势。文献[34]采用实时在线强化学习算法 Sarsa 优化充电调度过程和定价策略, 使公共电动汽车充电站的系统目标最大化。为了降

低电动汽车电池充电的长期成本, 文献[35]提出了一种批量强化学习算法, 从一批过渡样本中学习降低充电成本的最优充电策略, 结果表明 EV 用户可以节省 10%~50%的充电成本。然而, 上述强化学习方法只适用于离散的动作空间, 存在“维度爆炸”问题, 不适合具有大量动作和状态的应用场景。

深度强化学习算法采用深度神经网络来近似 Q 表格, 有效地解决了上述维度爆炸问题, 在电动汽车充电调度中得到了广泛应用。文献[36]将充电调度问题描述为一个转移概率未知的马尔可夫决策过程, 提出了一种基于深度强化学习的无模型方法来确定该问题的最优策略; 采用一个长短时记忆(Long Short-Term Memory, LSTM)网络提取电价特征, 利用一个 Q 网络逼近最优动作价值函数, 自适应地学习转移概率, 不需要任何系统模型信息。

为确保电动汽车离开充电站时满足所需电量, 上述基于传统 DQN 的方法严重依赖惩罚项的设计和惩罚系数的选择, 该过程相当繁琐, 并且主观性强。为了避免手动选择奖惩系数引起的误差, 文献[37]对 DQN 进行了改进, 提出了安全深度强化学习(Safe Deep Reinforcement Learning, SDRL), 通过深度神经网络直接学习电动汽车的原始状态信息和电价信息, 生成受约束的最优充电/放电时间表, 从而保证电动汽车离开充电站时满足用户电量需求, 并最小化充电成本。与现有的 DRL 方法不同, 文献[37]所提出的方法可以直接处理约束, 并且不需要为约束设计惩罚项和选择惩罚系数。

上述所提强化学习方法基本能够满足实时在线应用的要求, 但控制动作仅考虑了电动汽车是否充电, 有待综合考虑配电网其他控制动作以及深度强化学习的回报激励模型, 从而进一步提高强化学习的实用价值。针对在用电高峰时段给电动汽车集群充电可能会使配电网中变压器过载的问题, 文献[38]提出了一种多智能体自私协作架构(Multi-Agent Selfish Collaborative architecture, MASCO), 它是一种多智能体多目标强化学习体系结构, 能够最小化充电成本和避免变压器过负载。文献[39]将无模型

表 2 基于 DRL 的 EV 充电调度文献

Table 2 Literature about EV charging scheduling based on DRL

静态调度策略	文献	算法	状态	奖励	约束
基于价值	[32]	Q-learning	电池 SOC、电价、剩余充电时隙	单目标: 最小化用户充电成本	电池和充放电功率
	[33]	Sarsa	剩余充电能量、剩余充电时间	单目标: 充电站利润	充电率约束、充电数量约束
	[34]	Batched RL Fitted Q-iteration	电池 SOC、电价	多目标: 最小化充电开销和 电池损耗	驾驶模式约束
	[35]	Fitted Q-iteration	充电能量	单目标: 充电成本	充电功率约束
	[36]	LSTM+DQN	电价、电池 SOC	多目标: 最小化累积充电成本、 电池损耗和里程焦虑	充放电功率
	[37]	Safe DQN	电池 SOC、电价	单目标: 充放电总成本	充电需求约束
	[38]	MASCO	电池 SOC、EV 位置、 变压器负载	多目标: 满足充电需求、最小化 充电成本、减小变压器负载	变压器负载
	[39]	DQN	时间、剩余充电时隙、电量、 电价和光伏、风电、负载功率	多目标: 最小化充电成本和 微电网内部功率波动	充放电边界、充放电功率
静态充电调度	[40]	Customized Actor-Critic	电池 SOC、电价	单目标: 充电成本	电池容量约束、 充电负载约束
	[41]	Soft-Actor-Critic	剩余充电需求、剩余充电时隙	多目标: 充电成本、 完成充电比率	太阳能发电与容量约束、 充电率、剩余可充电量约束
	[42]	Continuous Soft Actor-Critic	电价、驾驶员充电偏好、 电池 SOC 和当前时间	多目标: 充电成本、用户焦虑度	电池容量约束、充电 功率约束
	[43]	Multi-agent AC	充电站运行成本	单目标: 充电站运行成本	电池容量约束
	[44]	DDPG+PER	电价、灵活与非灵活 EV 需求	单目标: 聚合者整体收益	聚合者内能量平衡和 平均价格约束
	[45]	JANET+DDPG	剩余电池容量、分时电价	单目标: 充电成本	电池容量约束
	[46]	DDPG	到达和离开时间、天气和交通 信息、充电负载	单目标: 分布式系统收益	EV 物理约束、网络约束和 充电约束
	[47]	Control DDPG	电池 SOC、电价、到达和 离开时间	多目标: 满足用户充电需求、 最大化收益	EV 到达离开时间、 电池能量消耗
基于策略	[48]	TD3	电池 SOC、电价、离开时间	单目标: 最小化集群 EV 的 充电开销	电池容量约束
	[49]	DQN	所有充电站中未使用的 快慢充电桩数量	多目标: 最小化充电的总时间、最 小化 OD 距离	EV 行驶距离限制、不平衡充 电率约束
	[50]	DQN	交通信息、等待时间、电价	多目标: 最小化充电成本和 等待时间	电池 SOC、功耗、等待和 行驶时间、充电成本
	[51]	DQN	电池 SOC、当前位置方向和已用 停车位、需充电 EV 数量	单目标: 电动汽车到达指定 充电站总开销	停车位、充电可行性
	[52]	GCN+DQN	EV 和充电站位置信息	多目标: 最小化充电时间损耗成 本、提高充电站服务均衡度	EV 剩余电量的安全行驶 里程
	[53]	DQN	EV 和充电站交通网实时数据	多目标: 途中行驶奖励和到站后 行驶奖励	距离能耗成本、 充电站充电费用
	[54]	MASTER	时间、当前可用充电点数、周围 充电请求数、充电功率和 预计到达时间	多目标: 最小化总充电时间、 平均充电功率和充电故障率	—
	[55]	Multi-agent AC	所有 EV 的位置和剩余电量	多目标: 最大限度地提高数据 收集率、降低所有车辆的能耗	—

的深度 Q 学习与基于物理机理的电动汽车充放电特性模型深度融合, 建立了以最小化功率波动和充电费用为目标的实时调度模型; 但设计的模型比较简单, 仅探讨了微电网内功率发生小范围波动场景下的实时调度策略, 且没有顾及电网约束。考虑配电网中电动汽车的空间分布与网络约束以及应对大场景变化的实时调度策略将是静态充电调度问题下一步研究的重点和方向。

### 2.1.2 基于策略梯度的 EV 静态充电调度算法

对于具有连续动作空间的任務, 例如电动汽车充电功率连续控制, DQN 的性能受限, 因其仅估计离散动作空间中的 Q 值。而基于 AC 架构的策略梯度强化学习算法在解决连续动作问题方面表现出了更大的潜力, 在电动汽车实时连续充/放电控制任务方面上表现出了更好的性能。

AC 算法在电动汽车静态充电调度中得到了广泛应用。为了减少车队充电费用和削减充电峰值负荷, 文献[40]提出了一种基于 AC 学习的智能充电算法, 并在此基础上通过降低状态维数提出了一个计算效率更高的 CALC (Customized Actor-critic Learning Charging) 算法。与文献[40]不同, 文献[41]考虑了可再生能源的影响, 提出了一种无模型的 SAC+NMT (Soft-Actor-Critic+ Nodal Multi-Target) 方法, 通过充分利用波动的光伏输出和电价来最小化充电成本。在充电调度过程中, 用户的动态行为也是一个值得关注的因素。为此, 文献[42]引入了集合焦虑的概念, 用以描述驾驶员对电动汽车行驶里程和不确定事件的焦虑程度, 并提出了一种基于连续 SAC 框架的强化学习充电控制方法, 以平衡充电成本和司机的焦虑度。与这些基于单智能体 AC 算法的调度算法不同, 文献[43]提出了一种新的基于多智能体 AC 算法的调度算法, 对配置有光伏系统和储能系统的分布式充电站进行能量管理, 实现多个充电站之间的协作运行; 该方法能够以分布式计算的方式获得多个充电站的调度方案, 同时能够处理与储能系统和电动汽车相关的动态数据, 从而有效减少充电站的运行费用。

AC 算法存在收敛速度较慢的缺点, 为此, 研究人员提出了融合 AC 与 DQN 优势的 DDPG 算法, 该算法在电动汽车静态充电调度中得到广泛应用。文献[44]在 AC 的基础上引入深度神经网络, 提出了一种新的深度强化学习 (DDPG) 方法, 解决了电动汽车充电问题, 将深度确定性策略梯度与优先经验重放 (Prioritized Experience Replay) 策略相结合, 并将问题建立在多维连续状态和动作空间中, 实验结果表明所提算法优于深度 Q 学习方法。文献[45]提

出了一种配电网中的最优电动汽车充电策略, 在满足物理约束的同时, 使配电系统运营商的利润最大化, 并通过马尔可夫决策过程模型来描述不确定性的时间序列, 然后利用基于 DDPG 的强化学习技术来分析不确定性对充电策略的影响, 并通过仿真的方法验证了算法的有效性。文献[46]将 DDPG 与改进的 LSTM 网络相结合, 采用改进的 LSTM 神经网络作为表示层, 从电价信号中提取时间特征, 利用具有连续动作空间的 DDPG 算法求解电动汽车最优充电序列; 该方法可以根据电价自动调整充电策略, 降低电动汽车用户的充电成本。文献[47]提出了一种改进的 CDDPG (Control Deep Deterministic Policy Gradient) 算法来学习电动汽车的最优充电控制策略, 从而在满足用户对电池能量需求的前提下, 最小化电动汽车用户的充电费用。文献[44-47]仿真结果均表明基于 DDPG 的方法在满足用户对电池能量需求和降低充电成本方面优于传统的基于 DQN 的方法。

针对强化学习方法动作空间离散、训练收敛困难、稳定性差, 并且充电调度算法未考虑充电桩效率随充电功率变化的问题, 文献[48]利用双延迟深度确定性策略梯度算法 (Twin Delay Deep Deterministic policy gradient, TD3) 对单辆电动汽车功率连续可调充电过程进行建模, 控制充电功率, 优化电动汽车充电行为。相较于传统的优化控制方法, TD3 在速度和灵活性上具有明显优势。通过对训练得到的智能体进行分布式部署, 该方法实现了对集群电动汽车充电行为的高速实时分布式优化。

考虑到分时电价与电网实际峰值出现时段可能存在偏差, 集群电动汽车的充放电仍然有可能造成电网“峰上加峰”情况。若供应商能根据电网状态, 制定实时的功率调节信号进一步对智能体充电行为进行引导, 将会有效解决这一问题。如何依据各电动汽车与电网的实时状态信息为智能体制定有效的调节信号, 有待进一步研究。

## 2.2 在动态充电调度中的应用

电动汽车动态充电调度算法的优化目标多种多样, 比如最小化行驶路线距离、行驶时间、行驶能耗及充电费用, 或者同时考虑这些因素的综合目标, 深度强化学习在该领域的应用也越来越广泛。

### 2.2.1 基于值函数的 EV 动态充电调度算法

Q-learning、DQN 等基于值函数的深度强化学习算法在电动汽车动态充电调度中依旧得到了广泛应用。文献[49]利用深度 Q 学习算法解决充电路径规划问题, 以最小化电动汽车的总充电时间和最大限度地减少出发地到目的地的距离为目标, 在真实数据集上的实验结果表明, 该算法能显著缩短电动

汽车的充电时间和行驶总里程；文献[50]提出了一种基于深度 Q 学习的电动汽车充电导航方法，旨在使电动汽车到达充电站的总行驶时间和充电成本最小化。文献[51]考虑了充电许可和电价波动，利用基于深度 Q 学习的调度算法最大限度地减少电动汽车包含时间成本和充电费用的总开销，但并未考虑实时交通的复杂性和充电站实时状态的多变性。

智能电网和智能交通系统的协同运行为电动汽车用户提供了丰富的电网和交通网数据，可用于电动汽车充电导航。针对不规律的电动汽车快速充电请求、高维且不规则的环境特征，文献[52]提出了一种基于图强化学习的快速充电站推荐方法，通过构建逐辆推荐式的强化学习框架，结合图卷积网络和强化学习，实现了实时充电站推荐的深度强化学习范式，并通过大量仿真验证了所提方法在城市路网中应用的可行性。然而，现实世界的充电场景有着更加复杂的不确定性信息。为此，文献[53]考虑了电动汽车动态驾驶行为和随机充电行为的多数据融合特性以及多系统建模的复杂性，提出了一种基于数据驱动和深度 Q 学习的电动汽车充电导航策略，为车主推荐最优充电站，并规划最优行驶路径。上述研究表明，综合考虑车-路-网的融合特性，充分利用电网和路网的运行数据，对大规模电动汽车的充电行为进行协同优化，实现电网和路网的协同高效运行值得进一步研究。

### 2.2.2 基于策略梯度的 EV 动态充电调度算法

多智能体强化学习算法在电动汽车动态充电调度领域逐渐受到重视。为了最小化充电等待时间、充电费用和充电失败率，文献[54]提出了一种多智能体时空强化学习 (Multi-Agent Spatio-Temporal Reinforcement Learning) 框架，开发了一个带有集中注意力机制的 Critic 的多智能体 AC 框架，以协调推荐不同地理位置的充电站；与 9 种基线方法相比，所提算法获得了最佳的综合性能。与文献[54]仅考虑充电站优化不同，文献[55]侧重于车辆与充电站之间的合作与竞争，提出一种基于多智能体 AC 算法的分布式充电站车辆调度控制框架 e-Divert，通过车辆与充电站之间的合作与竞争实现最大化能效、数据收集率、地理公平性，同时最小化能耗。

然而，与基于值函数的电动汽车动态充电调度算法相比，利用基于策略梯度的深度强化学习对电动汽车的动态充电过程进行优化的研究还处于起步阶段，相关研究成果还很少，有待进一步深入研究。

## 3 未来研究方向

目前深度强化学习在电动汽车充电调度中的应

用可以分为考虑新能源接入电网的充电桩级和充电站级充电调度、融合路网和配电网的综合充电路径规划以及融合图神经网络和强化学习二者优势的电动汽车充电引导算法。通过对现有成果的研究，未来可从以下 3 个方向深入研究。

1) 考虑新能源接入电网情况下针对充电桩级和充电站级的静态充电调度问题。目前的研究大都围绕分时电价进行调度，以实现电网负荷的“削峰填谷”，并降低运营商或用户的用电成本。但是在新能源接入电网情况下考虑新能源利用率的研究还较少。如何解决出力的不确定性是新能源接入电网的关键问题，传统基于模型的方法难以建立精确的系统模型，而强化学习在解决随机不确定因素时具有一定的优势。然而，强化学习方法也具有一定的局限性，例如在考虑新能源利用率时很大程度上依赖于人为设计奖励函数，如何优化奖励函数，能够使算法收敛的同时并获得预期效果尤为关键。目前深度学习方法(如循环神经网络)已在新能源出力预测领域取得一定成果，如何提高实时预测精度并与强化学习相结合，从而更好地参与配电网的负荷调度，并满足用户需求是潜在的研究热点。

2) 深度融合“车-路-网”信息的电动汽车综合充电路径规划。随着电动汽车的不断普及，未来对深度融合“车-路-网”的多源、异构信息，在高度随机的环境下实现集群电动汽车充电路径的优化调度以降低用户的充电费用，提高路网和电网的运行效率及安全性的研究具有重要意义。多智能体的优势就是合作、博弈，相较于单车调度具有很强的鲁棒性和可靠性，并具有较高的问题求解效率。但目前多智能体强化学习算法在电动汽车充电调度中的研究还很少，随着 MADDPG、Q-MIX 等多智能体深度强化学习算法应用的不断成熟，如何将其应用于解决大规模电动汽车充电调度是未来的重要研究方向之一。

3) 基于图强化学习的电动汽车充电引导算法。电网与交通网数据均呈现出典型的图结构，其拓扑中蕴含大量信息，图神经网络在处理海量图结构数据和复杂关联性问题时具有很大优势，能对不规则环境信息进行特征提取，学习图内包含的知识，对图中各节点间的相关性具有强大的建模能力。电动汽车和充电站之间关系密切，电动汽车快速充电站的引导问题本质就是一个推荐问题，正是图神经网络的强项之一。未来可将图神经网络与强化学习深度结合，图神经网络用于提取交通网特征，强化学习进行充电调度决策，从而设计出一套高效的快速充电引导和控制策略，是提高用户满意度、确保耦

合系统稳定运行的重要前提。基于此, 未来图深度强化学习算法在电动汽车充电调度领域中的应用值得研究人员进一步深入研究。

## 4 结论

本文概述了深度强化学习的基本工作原理, 对比了不同深度强化学习的特点。从电动汽车静态充电调度和动态充电调度两个方面综述了基于深度强化学习的充电调度算法, 分析了研究现状, 探讨了该领域未来值得进一步研究的方向。总体来说, 当前基于深度强化学习的电动汽车充电调度研究处于一个快速发展阶段, 值得研究人员进一步深入研究。

## 参考文献

- [1] 中华人民共和国工业和信息化部. 新能源汽车产业发展规划(2021-2035)年[EB/OL]. [2020-12-20]. [https://www.miit.gov.cn/xwdt/szyw/art/2020/art\\_4390362916324365a260ed97d7558f18.html](https://www.miit.gov.cn/xwdt/szyw/art/2020/art_4390362916324365a260ed97d7558f18.html).  
Ministry of Industry and Information Technology of the People's Republic of China. New energy vehicle industry development plan (2021-2035) [EB/OL]. [2020-12-20]. [https://www.miit.gov.cn/xwdt/szyw/art/2020/art\\_4390362916324365a260ed97d7558f18.html](https://www.miit.gov.cn/xwdt/szyw/art/2020/art_4390362916324365a260ed97d7558f18.html).
- [2] 张聪, 许晓慧, 孙海顺, 等. 基于自适应遗传算法的规模化电动汽车智能充电策略研究[J]. 电力系统保护与控制, 2014, 42(14): 19-24.  
ZHANG Cong, XU Xiaohui, SUN Haishun, et al. Smart charging strategy of large-scale electric vehicles based on adaptive genetic algorithm[J]. Power System Protection and Control, 2014, 42(14): 19-24.
- [3] 邢强, 陈中, 黄学良, 等. 基于数据驱动方式的电动汽车充电需求预测模型[J]. 中国电机工程学报, 2020, 40(12): 3796-3813.  
XING Qiang, CHEN Zhong, HUANG Xueliang, et al. Electric vehicle charging demand forecasting model based on data-driven approach[J]. Proceedings of the CSEE, 2020, 40(12): 3796-3813.
- [4] ZHANG L, LI Y. Optimal management for parking-lot electric vehicle charging by two-stage approximate dynamic programming[J]. IEEE Transactions on Smart Grid, 2017, 8(4): 1722-1730.
- [5] WU J, GAO B, ZHENG Q, et al. Optimal equivalence factor calculation based on dynamic programming for hybrid electric vehicle[C] // 2017 IEEE Chinese Automation Congress (CAC), July 16-20, 2017, Chicago, IL, USA: 6640-6645.
- [6] 阎怀东, 马汝祥, 柳志航, 等. 计及需求响应的电动汽车充电站多时间尺度随机优化调度[J]. 电力系统保护与控制, 2020, 48(11): 71-80.
- [7] YAN Huaidong, MA Ruxiang, LIU Zhihang, et al. Multi-time scale stochastic optimal dispatch of electric vehicle charging station considering demand response[J]. Power System Protection and Control, 2020, 48(11): 71-80.
- [7] 王锡凡, 邵成成, 王秀丽, 等. 电动汽车充电负荷与调度控制策略综述[J]. 中国电机工程学报, 2013, 33(1): 1-10.  
WANG Xifan, SHAO Chengcheng, WANG Xiuli, et al. Survey of electric vehicle charging load and dispatch control strategies[J]. Proceedings of the CSEE, 2013, 33(1): 1-10.
- [8] 陈奎, 马子龙, 周思宇, 等. 电动汽车两阶段多目标有序充电策略研究[J]. 电力系统保护与控制, 2020, 48(1): 65-72.  
CHEN Kui, MA Zilong, ZHOU Siyu, et al. Charging control strategy for electric vehicles based on two-stage multi-target optimization[J]. Power System Protection and Control, 2020, 48(1): 65-72.
- [9] 程杉, 杨堃, 魏昭彬, 等. 计及电价优化和放电节制的电动汽车充电站有序充放电调度[J]. 电力系统保护与控制, 2021, 49(11): 1-8.  
CHENG Shan, YANG Kun, WEI Zhaobin, et al. Orderly charging and discharging scheduling of an electric vehicle charging station considering price optimization and discharge behavior control[J]. Power System Protection and Control, 2021, 49(11): 1-8.
- [10] ZHAO J, WAN C, XU Z, et al. Risk-based day-ahead scheduling of electric vehicle aggregator using information gap decision theory[J]. IEEE Transactions on Smart Grid, 2017, 8(4): 1609-1618.
- [11] BINETTI G, DAVOUDI A, NASO D, et al. Scalable real-time electric vehicles charging with discrete charging rates[J]. IEEE Transactions on Smart Grid, 2015, 6(5): 2211-2220.
- [12] 邢强, 陈中, 冷钊莹, 等. 基于实时交通信息的电动汽车路径规划和充电导航策略[J]. 中国电机工程学报, 2020, 40(2): 534-550.  
XING Qiang, CHEN Zhong, LENG Zhaoying, et al. Route planning and charging navigation strategy for electric vehicle based on real-time traffic information[J]. Proceedings of the CSEE, 2020, 40(2): 534-550.
- [13] 李晓辉, 李磊, 刘伟东, 等. 基于动态交通信息的电动汽车充电负荷时空分布预测[J]. 电力系统保护与控制, 2020, 48(1): 117-125.  
LI Xiaohui, LI Lei, LIU Weidong, et al. Spatial-temporal distribution prediction of charging load for electric vehicles based on dynamic traffic information[J]. Power

- System Protection and Control, 2020, 48(1): 117-125.
- [14] 严奔遥, 罗禹贡, 朱陶, 等. 融合电网和交通网信息的电动车辆最优充电路径推荐策略[J]. 中国电机工程学报, 2015, 35(2): 310-318.  
YAN Yiyao, LUO Yugong, ZHU Tao, et al. Optimal charging route recommendation method based on transportation and distribution information[J]. Proceedings of the CSEE, 2015, 35(2): 310-318.
- [15] 王鑫, 周步祥, 唐浩. 考虑用户因素的电动汽车有序充放电控制策略[J]. 电力系统保护与控制, 2018, 46(4): 129-137.  
WANG Xin, ZHOU Buxiang, TANG Hao. A coordinated charging/discharging strategy for electric vehicles considering customers' factors[J]. Power System Protection and Control, 2018, 46(4): 129-137.
- [16] 牛利勇, 张帝, 王晓峰, 等. 基于自适应变异粒子群算法的电动出租车充电引导[J]. 电网技术, 2015, 39(1): 63-68.  
NIU Liyong, ZHANG Di, WANG Xiaofeng, et al. An adaptive particle mutation swarm optimization based electric taxi charging guidance[J]. Power System Technology, 2015, 39(1): 63-68.
- [17] ABDULLAH H M, GASTLI A, BEN-BRAHIM L. Reinforcement learning based EV charging management systems-a review[J]. IEEE Access, 2021, 9: 41506-41531.
- [18] ZHANG Z, ZHANG D, QIU R C. Deep reinforcement learning for power system applications: an overview[J]. CSEE Journal of Power and Energy Systems, 2020, 6(1): 213-225.
- [19] 肖智清. 强化学习原理与 Python 实现[M]. 北京: 机械工业出版社, 2019.
- [20] WATKINS C J C H, DAYAN P. Q-learning[J]. Machine Learning, 1992, 8(3-4): 279-292.
- [21] SUTTON R S. Generalization in reinforcement learning: successful examples using sparse coarse coding[J]. Advances in Neural Information Processing Systems, 1996, 8: 1038-1044.
- [22] MNIH V, KAVUKCUOGLU K, SILVER D, et al. Playing Atari with deep reinforcement learning[J]. arXiv preprint arXiv:1312.5602, 2013.
- [23] KONDA V R, TSITSIKLIS J N. Actor-critic algorithms[J]. Advances in Neural Information Processing Systems, 2000, 12: 1008-1014.
- [24] BABAEIZADEH M, FROSIO I, TYREE S, et al. GA3C: GPU-based A3C for deep reinforcement learning[J]. CoRR abs/1611.06256, 2016.
- [25] SILVER D, LEVER G, HEES N, et al. Deterministic policy gradient algorithms[C] // International Conference on Machine Learning, June 21-26, 2014, Beijing, China: 387-395.
- [26] LILICRAP T P, HUNT J J, PRITZEL A, et al. Continuous control with deep reinforcement learning[J]. arXiv preprint arXiv: 1509.02971, 2015.
- [27] FUJIMOTO S, HOOF H, MEGER D. Addressing function approximation error in actor-critic methods[C] // International Conference on Machine Learning, June 19-24, 2016, New York City, USA: 1587-1596.
- [28] HAARNOJA T, ZHOU A, HARTIKAINEN K, et al. Soft actor-critic algorithms and applications[J]. arXiv preprint arXiv: 1812.05905, 2018.
- [29] VAN H H, GUEZ A, SILVER D. Deep reinforcement learning with double q-learning[C] // Proceedings of the AAAI Conference on Artificial Intelligence, February 12-17, 2016, Phoenix, USA: 2094-2100.
- [30] SCHAUL T, QUAN J, ANTONOGLU I, et al. Prioritized experience replay[J]. arXiv preprint arXiv:1511.05952, 2015.
- [31] WANG Z, SCHAUL T, HESSEL M, et al. Dueling network architectures for deep reinforcement learning[C] // International conference on machine learning, June 19-24, 2016, New York City, USA: 1995-2003.
- [32] WEN Z, O'NEILL D, MAEI H. Optimal demand response using device-based reinforcement learning[J]. IEEE Transactions on Smart Grid, 2015, 6(5): 2312-2324.
- [33] WANG S, BI S, ZHANG Y A. Reinforcement learning for real-time pricing and scheduling control in EV charging stations[J]. IEEE Transactions on Industrial Informatics, 2021, 17(2): 849-859.
- [34] VANDAEL S, CLAESSENS B, ERNST D, et al. Reinforcement learning of heuristic EV fleet charging in a day-ahead electricity market[J]. IEEE Transactions on Smart Grid, 2015, 6(4): 1795-1805.
- [35] CHIŞ A, LUNDÉN J, KOIVUNEN V. Reinforcement learning-based plug-in electric vehicle charging with forecasted price[J]. IEEE Transactions on Vehicular Technology, 2016, 66(5): 3674-3684.
- [36] WAN Z, LI H, HE H, et al. Model-free real-time EV charging scheduling based on deep reinforcement learning[J]. IEEE Transactions on Smart Grid, 2019, 10(5): 5246-5257.
- [37] LI H, WAN Z, HE H. Constrained EV charging scheduling based on safe deep reinforcement learning[J]. IEEE Transactions on Smart Grid, 2020, 11(3): 2427-2439.
- [38] DA SILVA F L, NISHIDA C E H, ROIJERS D M, et al. Coordination of electric vehicle charging through multiagent reinforcement learning[J]. IEEE Transactions on Smart

- Grid, 2020, 11(3): 2347-2356.
- [39] 李航, 李国杰, 汪可友. 基于深度强化学习的电动汽车实时调度策略[J]. 电力系统自动化, 2020, 44(22): 161-167.  
LI Hang, LI Guojie, WANG Keyou. Electric vehicle real-time scheduling strategy based on deep reinforcement learning[J]. Automation of Electric Power Systems, 2020, 44(22): 161-167.
- [40] CAO Y, WANG H, LI D, et al. Smart online charging algorithm for electric vehicles via customized actor-critic learning[J]. IEEE Internet of Things Journal, 2022, 9(1): 684-694.
- [41] JIN J, XU Y. Optimal policy characterization enhanced actor-critic approach for electric vehicle charging scheduling in a power distribution network[J]. IEEE Transactions on Smart Grid, 2021, 12(2): 1416-1428.
- [42] YAN L, CHEN X, ZHOU J, et al. Deep reinforcement learning for continuous electric vehicles charging control with dynamic user behaviors[J]. IEEE Transactions on Smart Grid, 2021, 12(6): 5124-5134.
- [43] SHIN M J, CHOI D H, KIM J. Cooperative management for PV/ESS-enabled electric vehicle charging stations: a multiagent deep reinforcement learning approach[J]. IEEE Transactions on Industrial Informatics, 2020, 16(5): 3493-3503.
- [44] QIU D, YE Y, PAPADASKALOPOULOS D, et al. A deep reinforcement learning method for pricing electric vehicles with discrete charging levels[J]. IEEE Transactions on Industry Applications, 2020, 56(5): 5901-5912.
- [45] DING T, ZENG Z, BAI J, et al. Optimal electric vehicle charging strategy with Markov decision process and reinforcement learning technique[J]. IEEE Transactions on Industry Applications, 2020, 56(5): 5811-5823.
- [46] LI S, HU W, CAO D, et al. Electric vehicle charging management based on deep reinforcement learning[J]. Journal of Modern Power Systems and Clean Energy, 2022, 10(3): 719-730.
- [47] ZHANG F, YANG Q, AN D. CDDPG: a deep reinforcement learning-based approach for electric vehicle charging control[J]. IEEE Internet of Things Journal, 2020, 8(5): 3075-3087.
- [48] 赵星宇, 胡俊杰. 集群电动汽车充电行为的深度强化学习优化方法[J]. 电网技术, 2021, 45(6): 2319-2327.  
ZHAO Xingyu, HU Junjie. Deep reinforcement learning based optimization method for charging of aggregated electric vehicles[J]. Power System Technology, 2021, 45(6): 2319-2327.
- [49] ZHANG C, LIU Y, WU F, et al. Effective charging planning based on deep reinforcement learning for electric vehicles[J]. IEEE Transactions on Intelligent Transportation Systems, 2020, 22(1): 542-554.
- [50] QIAN T, SHAO C, WANG X, et al. Deep reinforcement learning for EV charging navigation by coordinating smart grid and intelligent transportation system[J]. IEEE Transactions on Smart Grid, 2020, 11(2): 1714-1723.
- [51] LIU J, GUO H, XIONG J, et al. Smart and resilient EV charging in SDN-enhanced vehicular edge computing networks[J]. IEEE Journal on Selected Areas in Communications, 2020, 38(1): 217-228.
- [52] 袁红霞, 张俊, 许沛东, 等. 基于图强化学习的电力交通耦合网络快速充电需求引导研究[J]. 电网技术, 2021, 45(3): 979-986.  
YUAN Hongxia, ZHANG Jun, XU Peidong, et al. Fast charging demand guidance in coupled power- transportation networks based on graph reinforcement learning[J]. Power System Technology, 2021, 45(3): 979-986.
- [53] CHEN L, YANG F, WU S, et al. Electric vehicle charging navigation strategy based on data driven and deep reinforcement learning[C] // Proceedings of the 5th International Conference on Control Engineering and Artificial Intelligence, January 14-16, 2021, Sanya, China: 16-23.
- [54] ZHANG W, LIU H, WANG F, et al. Intelligent electric vehicle charging recommendation based on multi-agent reinforcement learning[J]. arXiv preprint arXiv: 2102.07359, 2021.
- [55] LIU C H, DAI Z, ZHAO Y, et al. Distributed and energy-efficient mobile crowdsensing with charging stations by deep reinforcement learning[J]. IEEE Transactions on Mobile Computing, 2021, 20(1): 130-146.
- 
- 收稿日期: 2021-10-28; 修回日期: 2022-03-12  
作者简介:  
张延宇(1980—), 男, 通信作者, 博士, 副教授, 硕士生导师, 研究方向为车联网、能源互联网; E-mail: zyy@henu.edu.cn  
饶新朋(1997—), 男, 硕士研究生, 研究方向为电动汽车充电调度、深度强化学习; E-mail: xinpengrao@henu.edu.cn  
周书奎(1996—), 男, 硕士研究生, 研究方向为机器学习、电动汽车充电调度。E-mail: 104754190912@henu.edu.cn  
(编辑 姜新丽)