

碳达峰约束下电动汽车在线充电调度算法

曹永胜, 刘洋, 王永全, 夏天

引用本文

曹永胜, 刘洋, 王永全, 夏天. 碳达峰约束下电动汽车在线充电调度算法[J]. 计算机科学, 2024, 51(3): 265-270.

CAO Yongsheng, LIU Yang, WANG Yongquan, XIA Tian. [Online Electric Vehicle Charging Algorithm Based on Carbon Peak Constraint](#) [J]. Computer Science, 2024, 51(3): 265-270.

相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

[基于伪标签的弱监督显著特征增强目标检测方法](#)

FeaEM: Feature Enhancement-based Method for Weakly Supervised Salient Object Detection via Multiple Pseudo Labels

计算机科学, 2024, 51(1): 233-242. <https://doi.org/10.11896/jsjcx.230500035>

[面向兴趣点推荐系统的自然噪声过滤算法](#)

Natural Noise Filtering Algorithm for Point-of-Interest Recommender Systems

计算机科学, 2023, 50(11): 132-142. <https://doi.org/10.11896/jsjcx.230400045>

[云环境中面向可靠性约束的工作流调度策略研究](#)

Reliability Constraint-oriented Workflow Scheduling Strategy in Cloud Environment

计算机科学, 2023, 50(10): 291-298. <https://doi.org/10.11896/jsjcx.220800039>

[基于分类不确定性最小化的半监督集成学习算法](#)

Classification Uncertainty Minimization-based Semi-supervised Ensemble Learning Algorithm

计算机科学, 2023, 50(10): 88-95. <https://doi.org/10.11896/jsjcx.230600048>

[基于序贯三支决策的半监督目标检测算法](#)

Semi-supervised Object Detection with Sequential Three-way Decision

计算机科学, 2023, 50(10): 1-6. <https://doi.org/10.11896/jsjcx.230600035>

碳达峰约束下电动汽车在线充电调度算法

曹永胜^{1,2} 刘洋^{1,2} 王永全¹ 夏天³

1 华东政法大学智能科学与信息法学系 上海 201620

2 上海交通大学智慧法院研究院 上海 200240

3 上海第二工业大学计算机与信息工程学院 上海 201209

摘要 随着电动汽车数量的增加,电动汽车充电对社区电网总负荷和碳排放量产生了很大的影响,导致社区电网不太稳定,降低了电能质量。文中基于碳达峰的约束条件,研究了未提前知晓电动汽车到达时间、出发时间和充电需求等情况下的电动汽车充电碳排放问题。首先,建立了电动汽车充电碳排放问题,并在未来信息未知的情况下进行了研究。针对电动汽车充电行为的不确定性,提出了一种改进型演员-评论家的智能充电碳排放算法。该算法采用的是连续碳排放动作的电动汽车充电碳排放策略,而不是离散近似碳排放动作。仿真结果表明,相比 OA 和 AEM 两个基准算法,所提算法能够降低电动汽车预期成本约 24.03% 和 21.49%。

关键词: 碳达峰;电动汽车;不确定性;在线充电;能量管理

中图分类号 TP311

Online Electric Vehicle Charging Algorithm Based on Carbon Peak Constraint

CAO Yongsheng^{1,2}, LIU Yang^{1,2}, WANG Yongquan¹ and XIA Tian³

1 Department of Intelligent Science and Information Law, East China University of Political Science and Law, Shanghai, 201620, China

2 Shanghai Jiao Tong University Intelligent Court Research Institute, Shanghai, 200240, China

3 School of Computer and Information Engineering, Shanghai Polytechnic University, Shanghai, 201209, China

Abstract With the increasing number of electric vehicles (EVs), EV charging significantly increases the total load of the community, greatly increases the carbon emissions of the community, brings great instability to the community power grid, and reduces the power quality of the community. This paper studies the problem of scheduling EV charging based on the constraints of carbon peak when the arrival time, departure time, and charging demand of EVs are not known in advance. First, we formulate and study the problem of charging EVs without knowing future information. Aiming to address the uncertainty of EV charging behavior, we propose an algorithm for intelligent charging carbon emissions using the actor-critic approach, which learns the optimal strategy for EV charging through continuous charging instead of using a discrete approximation of carbon emissions. Simulation results demonstrate that compared with the online charging algorithm and the AEM energy management algorithm, the proposed algorithm can reduce the expected cost by 24.03% and 21.49%.

Keywords Carbon peak, Electric vehicles, Uncertainty, Online charging, Energy management

1 引言

1.1 研究背景

随着人们对环境问题的日益关注,电动汽车成为了一种不直接消耗化石燃料、更环保且有前景的解决方案。然而

电动汽车间歇性的大功率充电需求会对公共电网的运行稳定性产生影响^[1]。电动汽车在碳减排和碳达峰方面扮演着关键角色,它们在道路交通领域能够显著减少温室气体的排放,有助于降低总体碳排放量。而电动汽车的并网给电力系统带来了一定压力。当大量电动汽车同时连接到电力系统进行

到稿日期:2023-08-08 返修日期:2024-01-24

基金项目:上海市科技创新行动计划启明星项目(扬帆专项)(22YF1411900);第四批中国博士后科学基金特别资助项目(2022TQ0210);国家自然科学基金(20&ZD199);教育部人文社会科学研究项目(20YJC820030);国家社科基金重大项目(21&ZD200);国家重点研发计划重点专项课题(2023YFC3306103,2023YFC3306105)

This work was supported by the Shanghai Science and Technology Innovation Action Plan Star Project (Sail special Project) (22YF1411900), Fourth Special Projects Funded by China Postdoctoral Science Foundation (2022TQ0210), National Social Science Foundation of China (20&ZD199), Humanities and Social Sciences Research Project of Ministry of Education (20YJC820030), Major project of National Social Science Fund (21&ZD200) and Key Research and Development Program of China (2023YFC3306103, 2023YFC3306105).

通信作者:曹永胜(yongshengcao@ecupl.edu.cn)

充电时,电网需要满足增加的负荷需求,这意味着电力系统必须具备足够的供电能力,以应对电动汽车充电所需,防止电网过载或电压波动等问题的发生。电动汽车本身也存在一定程度的不确定性,车辆的到达和离开时间可能会变化,导致充电需求波动。在充电调度算法中需要充分考虑这些不确定性因素,以确保充电过程的高效性和可靠性。此外,社区负载的不确定性也会影响电动汽车的充电行为。社区负载的不确定性指电网中各用户的用电行为可能发生变化,如用电峰谷的时间变动、用电量的不确定性等。这种不确定性将直接影响对电动汽车充电需求的预测和管理,因此需要考虑这些变化以实现有效的充电调度。

因此,在碳达峰约束下,研究电动汽车在线充电碳排放调度算法变得紧迫且必要^[2]。这样的算法可以优化电动汽车的充电策略,考虑到电力系统供电能力、社区负载的不确定性以及电动汽车本身的不确定性。通过合理调度和管理电动汽车的充电行为,可以最大程度降低充电对电力系统的影响,提高充电效率,同时满足电动汽车用户的需求。这将为实现碳达峰和碳减排目标提供重要支持,并推动可持续发展的能源和交通系统建设。如果能提前了解电动汽车未来的充电需求,就能安排合理的充电计划,使社区总负荷趋于平稳^[3]。随着中国碳达峰和碳中和目标的宣布,电力行业的发展路径受到广泛关注,我们需要不断探索符合中国国情的电力行业碳达峰和碳中和的发展路径^[4]。然而,电动汽车的出行方式和充电需求具有巨大的不确定性。在线充电策略成为了在不确定条件下实现电动汽车最优充电的有效方法。在线充电问题依赖于当前和过去的电动汽车配置信息,包括到达时间、出发时间和充电需求。物联网技术和智能交通系统的进步推动了电动汽车的快速发展^[5],共享的交通信息可以提高电动汽车之间的实时充电与路径决策调度,并为个体电动汽车做出智能决策,更容易预测电动汽车的行为,从而为适当数量的电动汽车提供充电决策。因此,本文在电动汽车未来信息未提前知晓的情况下设计了电动汽车充电在线算法。

1.2 研究现状

近年来,电动汽车随机充电需求下的在线充电算法已有一些研究。在线充电问题没有假设未来电动汽车的分布信息,现实场景中需要捕捉电动汽车充电行为的不确定性。文献[6]利用李雅普诺夫优化方法和改进的交替方向法提出了一种分布式在线算法,研究分布式数据中心和电动汽车的能源调度问题。文献[7]提出了一种基于进化算法的电动汽车在线能量管理框架。文献中[8]提出了一种考虑充电优先级和充电地点的基于实时电价的电动汽车集中充电策略。文献[9]提出了另外一种考虑时变需求响应和消费者响应的联合在线学习和定价算法,用来最小化运营电力成本。文献[10]提出了具有一定实际约束条件的商业建筑和电动汽车充电站潮流优化两阶段随机规划模型。然而,这些算法^[6-10]依赖于特定的模型或只在特殊的场景下工作。我们的目标是开发一种通用的方法来解决电动汽车充电问题,该方法对模型的依赖性较小,可以适用于各种实际场景。因此,本文采用无模型强化学习方法寻求电动汽车最优充电策略。

已有文献利用无模型强化学习框架和策略来处理能源

调度问题。例如,文献[11]提出了一种改进的Q强化学习方法,以最小化电费,减少用户的不适。文献[12]设计了一种基于Q强化学习的控制算法,以获得物理和网络不确定性下的最优控制。文献[13]研究了一种批量强化学习算法来调度可控制负载。然而,文献[11,13]都采用了强化学习方法,其中充电动作必须是离散值,这限制了模型的适用范围。由于电动汽车的充电量是一个连续的值,因此可以使用离散动作对电动汽车充电连续值进行近似,当采用离散充电动作时,在计算复杂度和性能之间很难取得良好的平衡。使用太少的离散动作空间可能会导致学习性能较差,但使用较大的离散动作空间会使训练变得困难,并导致计算开销较高^[14]。

标准的Q学习方法或深度Q神经网络方法无法推导出连续动作空间中的最优策略。演员评论家Actor-Critic算法同时使用了值函数网络和策略网络,它可以通过值函数的反馈来更稳定地训练策略网络,减少策略更新的方差,从而提高算法的稳定性。Actor-Critic算法的结构可以更容易地进行并行化实现,因为值函数网络和策略网络可以独立地更新。电动汽车充电量的问题通常涉及连续的充电动作空间,Actor-Critic算法可以直接输出一个连续动作的概率分布,其通过值函数网络(Critic)来估计状态的价值,值函数可以帮助评估不同充电量的好坏,并指导策略网络(Actor)的改进。通过值函数的反馈,算法可以更准确地估计充电量对未来回报的影响,从而更好地调整策略。因此,我们采用演员评论家方法^[15]来解决状态和动作连续的问题,利用深度神经网络对值函数进行估计,可以得到更好的逼近效果。已有学者利用演员评论家方法对调度问题进行了一些研究。文献[16]提出了一个并行演员评论家学习框架,以获得无限视界中多人非零博弈的接近最优反馈纳什均衡解。基于自适应动态规划策略,文献[17]提出了一种演员评论家方法来逼近性能函数。为此,本文设计了一种基于异步演员评论家方法的电动汽车在线充电策略,获得了电动汽车充电计划的接近最优解。

本文旨在设计一种电动汽车在线充电碳排放算法,在电动汽车未来状态未知的情况下,制定在线碳排放量策略。将电动汽车的碳排放量决策建模为马尔可夫决策过程,充电站根据电动汽车的过去和当前信息,包括到达时间、出发时间和充电需求,优化电动汽车的充电调度计划。在马尔可夫过程中,当前的碳排放量决策将影响下一个状态、碳排放量决策和未来的累积奖励。本研究在电动汽车充电行为不确定的情况下,研究电动汽车充电碳排放量策略以最小化总能量成本,提出了一种基于演员评论家的智能充电算法,该算法利用异步演员评论家算法确定每辆电动汽车的最优充电碳排放量。由于Actor-Critic算法同时学习策略网络和值函数网络,可以通过值函数的估计结果来指导策略的改进。值函数提供了对不同充电碳排放量的评估,可以帮助策略网络选择更优的充电策略,从而可以提高算法的收敛速度和性能。最后,将所提算法与其他两种算法进行比较。

1.3 本文贡献

随着电动汽车数量的增加,如何高效地进行低碳充电调度成为重要问题。本文研究了广泛应用于电动汽车充电管理系统的智能算法,有利于降低电网负荷和总体碳排放。本文

考虑了电动汽车充电情况的不确定性,提出的动态调度算法的鲁棒性好,更能适应复杂的充电环境。这对实际电网运行管理具有重要参考价值。本文提出了电动汽车在线充电碳排放调度问题,重点研究了在未来信息不确定的情况下如何选择最佳的充电策略以减少碳排放,具体贡献总结如下:

- 1)将电动汽车在线充电建模碳排放为马尔可夫过程,考虑了充电情况的不确定性,寻求最优的碳排放决策。
- 2)针对电动汽车在线充电碳排放决策问题,提出了一种基于改进演员-评论家算法的智能碳排放调度算法,该算法具有收敛性好的优点,可以求解最优充电碳排放策略。
- 3)利用所提算法求解连续碳排放行为的决策策略,并与离散动作空间的决策算法进行对比,证明了所提算法在计算与性能平衡上优于离散算法。

2 系统模型

本章构建了电动汽车实时充电碳排放调度模型,我们的目标是找到每辆电动汽车 k 的最优充电解决方案,以使电动汽车在操作区间 t 内的充电总成本最小。电动汽车每公里行驶所需电能约为 0.020 kWh 。假设电动汽车主要使用火力发电充电,而火力发电的碳排放强度为 $0.96 \text{ kg}/(\text{kWh})^{[18]}$ 。根据这些数据,可以计算得出电动汽车每公里行驶的充电过程中所产生的碳排放量为 0.198 kg 。设 $\Gamma(k, \tau) = [\tau, t_k^d]$ 为电动汽车 k 在充电站充电的时间指标集合。定义 EV k 在时间 t 的耗电量对应的碳排放为 $b_k(t)$,其中 $t \in \Gamma(k, \tau)$ 。当电动汽车在 τ 时隙到达充电站时,我们在不知道未来电动汽车需求的情况下优化充电碳排放量调度。将 $(b_1(t), b_2(t), \dots, b_k(t))$ 向量表示为 $\mathbf{b}(t)$,充电站的碳排放成本 $c(\mathbf{b}(t))$ 的计算式如下:

$$c(\mathbf{b}(t)) = \int_{l_b(t)}^{L(t)} \xi(\beta_0 + 2\beta_1 z) dz \quad (1)$$

在线碳排放量问题中,电动汽车的充电需求、到达时间和出发时间等信息都未提前知晓。充电站需要对当前各时间段的电动汽车进行充电调度,同时面临电动汽车未来充电信息的不确定性。根据文献[19],我们定义了在时刻 t 出现的外生信息变量 $\epsilon_k(t)$ 。

$$\epsilon_k(t) = (\tilde{t}_k^a, \tilde{t}_k^d, \tilde{D}_k(t)) \quad (2)$$

其中, $\tilde{t}_k^a, \tilde{t}_k^d, \tilde{D}_k(t)$ 分别表示 t 时刻电动汽车 k 的到达时间、驶离时间和电力需求。这些外生变量是实时观测的,给问题的解决带来了不确定性。

为了考虑除充电站充电外的地区基础负荷,我们需要综合计算 t 时刻区域的总用电所对应的碳排放量 $L(t)$,包括充电过程的碳排放量和基础负荷的碳排放量。

$$L(t) = \xi(\sum_k b_k(t) + l_b(t))$$

其中, $l_b(t)$ 为 t 时刻地区基础负荷, ξ 为碳排放系数。

我们的目标是找到最优充电碳排放量 $b = (b(1), \dots, b(T))$,使电动汽车在时间区间 T 内的碳排放总成本 C_T 最小。

$$C_T = \sum_{t=1}^T c(\mathbf{b}(t)) \quad (3)$$

本文构建了一个在线碳排放量优化调度问题:

$$\min \mathbb{E}[C_T]$$

$$\begin{aligned} \text{s. t. } \quad & \sum_{t \in Q(k, \tau)} b_k(t) = \frac{\tilde{D}_k(\tau)}{\eta}, k \in H(\tau), \\ & 0 \leq b_k(t) \leq b_{k, \max}, k \in H(\tau), t \in \tilde{T}(k, \tau), \\ & 0 \leq L(t) \leq L_{\max}, t \in \tilde{T}(k, \tau) \end{aligned} \quad (4)$$

其中, $\tilde{T}(k, \tau) = [\tau, \tilde{t}_k^d]$, $H(\tau)$ 为在 τ 时间停放在充电站的电动汽车集合,并在 t 时间点停留在充电站, $t \in \tilde{T}(k, \tau)$ 。在考虑充电过程的碳排放量和基础负荷的碳排放量时,需要注意 $L(t)$ 存在一个上限 L_{\max} ,这个上限满足碳达峰约束。碳达峰指国家或地区在一定时间内将碳排放量控制在峰值水平并逐渐减少的目标。具体来说,碳达峰政策可能要求电力系统在充电过程中的碳排放量不得超过特定的量,并且基础负荷的碳排放量也需要控制在一定范围内。这意味着电动汽车在充电时必须满足这些约束条件,以确保整体的碳排放量控制在可接受的范围内。我们通过优化电动汽车的充电策略,使其在碳排放较低的时间段进行充电,以最大程度地减少碳排放。

充电站将对在线充电优化问题求解最优解 $b_k^*(t)$,当新的电动汽车到达,或电动汽车完成充电,或来自社区的负载发生变化时,需要更新环境变量 $H(\tau), \tilde{T}(k, \tau), \tilde{D}_k(\tau)$,重新求解问题。

3 基于强化学习的在线碳排放算法

为解决电动汽车行为不确定性带来的问题,本文设计了一种基于强化学习的电动汽车在线碳排放算法,针对电动汽车充电量连续的问题,将基于值和基于策略的方法相结合,采用演员评论家算法解决电动汽车在线碳排放量问题。

将电动汽车充电碳排放决策建模为一个马尔可夫过程,充电站将根据过去和现在的外生信息制定充电计划,包括到达时间、出发时间,以及当前时间段在充电站中电动汽车的充电需求。状态 $\phi(t)$ 由电价 $p(t)$ 和电动汽车 k 的 SOC 组成。我们定义 $\text{SOC}_i(t)$ 为电动汽车 i 的电池容量的百分比。定义状态如下:

$$\phi(t) = (\text{SOC}_1(t), \dots, \text{SOC}_N(t), p(t)) \quad (5)$$

其中, $p(t)$ 是电力价格,根据文献[20-21],将电价建模为小区总负荷的线性函数 $p(t) = k(\sum_k b_k(t) + l_b(t)) = kL(t)/\xi$,随着电力负荷的增加,电价也会相应地增加。为了减少总体的碳排放量,我们需要将总负荷控制在一定的范围内,并确保碳排放量保持在峰值 L_{\max} 以下;通过优化充电策略和能源管理,使得电动汽车充电过程中的碳排放量最小化。在负荷增加时,我们动态调整充电时段和充电速率,将充电峰值分散到负荷较低、碳排放较少的时间段,以避免碳排放量过高。

充电碳排放动作空间由电动汽车 k 在 t 时刻的全部碳排放量 $b_k(t)$ 组成, $b_k(t)$ 为连续变量,其约束条件如下:

$$0 \leq b_k(t) \leq b_{k, \max}, k \in H(\tau), t \in \tilde{T}(k, \tau) \quad (6)$$

其中, $b_{k, \max}$ 为电动汽车 k 的最大碳排放量。充电碳排放量动作需要满足以下约束条件:

$$\sum_{t=\tau}^{\tilde{t}_k^d} b_k(t) = \frac{\tilde{D}_k(t)}{\eta} - \sum_{t=\tau}^{\tau-1} b_k(t) \quad (7)$$

其中, τ 是当前时刻, $\tilde{D}_k(t)$ 是电动汽车 k 的充电需求, η 是耗电量与碳排放量的转换系数, $\sum_{t=\tau_i}^{\tau-1} \tilde{b}_k(t)$ 是到达时间至当前时间段之间的电动汽车 k 的实际碳排放量。充电碳排放动作 $\tilde{b}_k(t)$ 的每个决策都会对状态产生影响, 反过来对未来时间段的电动汽车剩余充电需求 $\tilde{D}_k(t)$ 产生影响。

在设计状态空间时, 我们采取了两个阶段的算法优化策略。在第一个阶段, 算法致力于优化整个车队的充电计划, 以达到最佳的总充电量。状态空间为 $\phi(t) = (\text{SOC}_{ev}(t), \dots, \rho(t))$, 其中 $\text{SOC}_{ev}(t) = \sum \text{SOC}_i(t)$ 是 t 时刻充电站内全部电动汽车的整体充电量, 这意味着算法将车队总充电量作为一个动作, 从而有效地降低了状态空间的维度, 进而可以获得优化电动汽车车队充电碳排放动作。一旦完成了整个车队充电计划的优化, 我们就进入第二个阶段, 此时算法专注于为每辆电动汽车寻找接近最优的充电策略。在第二阶段中, 算法可以利用第一阶段得到的整体充电计划作为参考, 然后针对每辆电动汽车进行个体化的充电策略优化。此策略通过将车队总充电量作为动作, 显著降低了状态空间的维度, 从而减轻了算法的计算负担并提高了效率。

本文根据电价与负荷的关系, 利用电动汽车碳排放成本, 将奖励函数 $r(t)$ 设为:

$$- \sum_{k \in H(t)} (\beta_0 + 2\beta_1 b_k(t) + 2\beta_2 l_b(t)) b_k(t) \quad (8)$$

其中, $b_k(t)$ 是电动汽车 k 在时刻 t 的碳排放量。为了评估 $b_k(t)$ 的期望累积奖励, 使用策略 π_θ 根据状态 $\phi(t)$ 来选择充电碳排放动作, 设置值函数 $Q^{\pi_\theta}(\phi(t), b_k(t)) = \mathbb{E}_{\pi_\theta} [r(t) + \gamma Q^{\pi_\theta}(\phi(t+1), b_k(t+1))]$, 即在 π_θ 策略下, 在状态 $\phi(t)$ 下进行 $b_k(t)$ 的充电碳排放动作。其中 $\mathbb{E}_{\pi_\theta}[\cdot]$ 是期望函数, γ 是对预计期望决策的折扣因子。

策略函数是从状态到动作的映射, 它描述了代理在给定状态下应该如何表现。策略梯度方法旨在找到一个最大化期望回报的策略函数, 期望回报是随时间变化的奖励之和。策略梯度方法通常用于连续控制任务, 在这些任务中, 代理可以采取连续范围内的动作。我们假设策略梯度方法 $\pi_\theta(\phi(t), b_k(t))$ 是关于 θ 可微的, 且 θ 按如下方式更新:

$$\Delta\theta = \beta_a \nabla_\theta \mathcal{J}(\pi_\theta) = \beta_a \frac{\partial \mathcal{J}(\pi_\theta)}{\partial \pi_\theta} \frac{\partial \pi_\theta}{\partial \theta} \quad (9)$$

其中, β_a 是参与者的学习率, 应该设置得足够小, 以避免策略的振荡, 因为值函数的小更新将极大地影响策略的更新。根据最大熵原理, 我们可以利用高斯概率分布提供一个参数化策略来选择连续充电碳排放动作。

$$\pi_\theta(\phi(t), b_k(t)) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(b_k(t) - \varphi(\phi(t)))^2}{2\sigma^2}} \quad (10)$$

其中, $\varphi(\phi(t))$ 是充电状态的平均动作值, σ 定义了所有可能的充电碳排放动作的标准差。故 $\pi_\theta(\phi(t), b_k(t))$ 是在状态 $\phi(t)$ 中选择充电碳排放动作 $b_k(t)$ 的概率。根据 $Q^{\pi_\theta}(\phi(t), b_k(t))$ 可知充电碳排放动作 $b_k(t)$ 在状态 $\phi(t)$ 下的期望报酬。然后调整策略 π_θ 来进行电动汽车充电碳排放调度, 策略 π_θ 由高斯分布 $\mathcal{N}(\mu_\theta, \sigma_\theta^2)$ 生成连续动作。期望值 μ_θ 用多层感知机近似。利用时间差 (TD) 误差来表示近似值与真实值之间的误差。TD 误差定义为, $\delta_t = r(t+1) + \epsilon Q^{\pi_\theta}(\phi(t+1), b_k(t+1)) - Q^{\pi_\theta}(\phi(t), b_k(t))$, 其中 $r(t+1)$ 是下一个时间槽 $t+1$ 的奖励,

$r(t+1) + \epsilon Q^{\pi_\theta}(\phi(t+1), b_k(t+1))$ 是时间 t 的实际回报。

算法 1 给出了基于演员-评论家方法的智能充电碳排放 (OSC) 算法的描述。首先, 我们设置评论家学习率 β_c 、演员学习率 β_a , 以及折扣因子 ϵ 。评论家流程更新值函数及其参数, 然后演员过程利用评论家的输出来更新其策略参数。当演员学习过程与批评家过程输出呈正相关时, TD 误差可以指示学习过程是否比预期的更好或更差。我们的目标是减小 TD 误差, 使其能够在梯度方向上调整演员过程和批评家过程。根据当前状态、下一状态和当前奖励产生 TD 误差, 该 TD 误差将根据当前电动汽车充电碳排放动作和状态更新策略参数。

演员-评论家方法是一种无模型强化学习方法, 它使用两个独立的模型来学习策略和价值函数。演员模型学习策略函数, 该函数将状态映射到动作, 而评论家模型学习价值函数, 该函数将状态映射到值。批评家过程通过状态-动作值函数 $Q^{\pi_\theta}(\phi(t), b_k(t))$ 来评估策略, 演员过程具有以下策略梯度:

$$\nabla_\theta \mathcal{J}(\pi_\theta) \approx \int_{\phi} D^{\pi_\theta}(\phi(t)) \int_{b_k(t)} Q^{\pi_\theta}(\phi(t), b_k(t)) \nabla_\theta \pi_\theta(b_k(t) | (\phi(t), \theta)) db_k(t) d\phi(t) \quad (11)$$

根据演员参数 θ 和临界参数 θ^v 同时更新, 演员参数 θ 按照评论家输出决定的方向更新。当行动者-批评算法收敛时, 对两组参数进行优化。

该算法有一个策略 $\pi_\theta(b_k(t) | (\phi(t), \theta))$ 和一个状态的近似值函数 $Q(\phi(t), \theta^v)$, 并使用多步返回方法来更新策略和价值函数。在每一个 k_{\max} 动作之后, 策略和价值函数都会被更新, 直到满足了电动汽车的充电需求。参数通过以下方式更新:

$$d\theta^v = d\theta^v + \partial(R(t) - Q(\varphi(t), \theta^v))^2 / \partial\theta^v \quad (12)$$

演员过程参数通过以下方式更新:

$$d\theta = d\theta + \nabla_\theta \log \pi(b_k(t) | (\phi(t), \theta)) (R(t) - Q(\phi(t), \theta^v)) \quad (13)$$

更新过程中累积奖励的迭代公式为 $R(t+1) = r(t) + \epsilon R(t)$, 直接奖励为 $r(t) = - \sum_{k \in H(t)} (\beta_0 + 2\beta_1 b_k(t) + 2\beta_2 l_b(t)) b_k(t)$ 。算法在多线程中运行。线程中的梯度会累积, 参数会共享和更新。经过一定的迭代, 奖励趋于收敛, 得到最优的充电碳排放方案。

算法 1 基于演员评论家的在线充电碳排放算法

输入: 评论家学习率 β_c 和演员学习率 β_a , 折扣因子 ϵ

输出: 电动汽车充电碳排放决策 $b_k(t)$

1. 初始化: 线程步骤计数器 $t=1$, 全局共享计数器 $k=0$, $\phi(0) \sim d^{\pi_\theta}(\phi(t))$, 设置参数 θ, I ;
2. 线程开始: 复位梯度: $d\theta' = 0, d\theta^v' = 0$;
3. 同步线程参数: θ, θ^v ;
4. 设置算法起始时间, 得到状态 $\phi(t)$;
5. 循环操作, 其中每一步选择充电碳排放量: $b_k(k+1) \sim \pi_\theta(\phi(t), b_k(t))$;
6. 更新下一个状态: $\phi(t+1) \sim P(\phi(t), b_k(t), \phi(t+1))$;
7. 得到即时奖励 $r(t+1)$, 更新 k ;
8. 更新函数: $\Psi(\varphi(t), b_k(t)) = \nabla_\theta \ln \pi_\theta(\phi(t), b_k(t))$;
9. 更新: $I = \epsilon I$;
10. 更新函数: $Q^{\pi_\theta}(\phi(t+1), b_k(t+1)) = \theta^v T \nabla_\theta \ln \pi_\theta(\phi(t+1), b_k(t+1))$;
11. 更新评论家参数: $\theta_{t+1}^v = \theta_t^v + \beta_c \delta_t I$

12. 更新演员参数:

$$\theta_{i+1} = \theta_i + \beta_a \delta_i \nabla_{\theta} J(\pi_{\theta})$$

13. 更新: $\phi(t), b_k(t), Q^{\pi_{\theta}}(\phi(t), b_k(t))$

14. 使用 $d\theta', d\theta''$ 更新 θ', θ'' .

15. 直到 $k > k_{\max}$ 线程结束。

我们的目标是通过优化电动汽车的充电碳排放策略,使其充电过程中的碳排放量最小化;根据电网负荷曲线,动态调整电动车的充电时段安排;在电网负荷增加时,通过减小部分车辆的充电速率,将充电负荷峰值分散到负荷较低的时间段,将充电活动导向碳排放效率较高的那些时段,避免碳强度值超标;整体上降低电动车的充电过程中的平均碳强度,从源头上减少碳排放量。

4 实验与性能分析

本文采用持续两天的社区基础负载进行数值分析,电动汽车到达的统计数据(电动汽车到达分布、电动汽车到达时电池 SOC 的概率密度函数)可以从文献[19]中获得。

分别在两种情况下考虑两种类型的电动汽车: A 型,最大的充电率为 3.2 kW,电池容量为 36 kWh; B 型,最大充电量为 1.4 kW,电池容量为 16 kWh。我们通过蒙特卡罗模拟电动汽车到达和离开的分布和充电需求。

使用数据仿真实验来评估所提算法的性能,并将其与两种基准算法进行比较,基准算法 I 为滚动在线控制算法^[20],基准算法 II 为基于强化学习的 AEM 算法^[21]。AEM 算法中的充电碳排放动作是离散的。

图 1 给出了两天内的基准算法 I、基准算法 II 和 OSC 算法下的电动汽车充电碳排放结果。OSC(ϵ)算法是电动汽车充电行为不确定性下 OSC 算法的仿真结果。从图 1 中可以看出,在所有测试的算法中 OSC 算法下碳排放量峰值最低,算法 I、算法 II 以及 OSC 算法下碳排放量峰值分别为 68.01 kg, 67.51 kg 和 60.33 kg, OSC 算法下的碳排放量峰值相比算法 I 和算法 II 减少了 11.29% 和 10.64%。OSC 算法比算法 I 和算法 II 的波动更小。观察 OSC 算法和 OSC(ϵ)算法的结果可知,电动汽车充电行为的不确定性对充电调度算法的性能产生了影响。

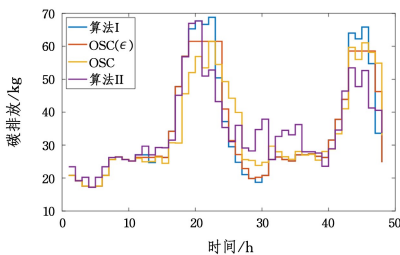


图 1 两天内电动汽车的充电碳排放量比较

Fig. 1 Comparison of carbon emissions from electric vehicle charging in two days

对于 A 型电动汽车,算法 I、算法 II 和 OSC 的电动汽车充电总成本分别为 200.69 元、127.54 元和 112.07 元。算法 I 和算法 II 的电动汽车碳排放成本比 OSC 算法的高 24.03% 和 21.49%,证明了 OSC 算法的性能优于算法 I 和算法 II。对于 B 型电动汽车,算法 I、算法 II 和 OSC 的充电总成本分别为 105.98 元、94.29 元、80.57 元,算法 I、算法 II 的电动

汽车充电碳排放成本分别比 OSC 算法高 23.97% 和 9.16%。

从图 2 中可以看出不同算法的不同电动汽车数量下每辆电动汽车的充电平均成本。算法 II 有 3 个级别的离散充电碳排放动作,分别是 33, 3300 和 33000 个离散值。仿真算法的平均成本都随着电动汽车数量的增加而稳步增加,OSC 在所有仿真中的平均碳排放成本最低。在不同数量的电动汽车的仿真实况下,我们验证了 OSC 算法的有效性。当电动汽车车辆数为 40 时,算法 I、算法 II-33、算法 II-3300,算法 II-33000、OSC 算法的每个电动汽车平均碳排放成本分别为 3.61 元、3.22 元、3.10 元、2.96 元、2.74 元。

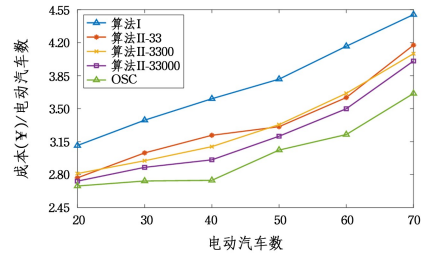


图 2 不同电动汽车数量下算法 I、算法 II 和 OSC 算法下每辆电动汽车的平均电力成本比较

Fig. 2 Comparison of average costs per EV of algorithms I, II, and OSC with EVs

上述现象主要是由社区负载的高峰需求和电动汽车用户的行为决策所致。例如,在晚上回家后,大量电动汽车同时连接到电网进行充电,这导致了负荷的集中和增加。本文提出的 OSC 算法的碳排放成本较其他两个基础算法更少,且更为平稳,结合图 1 电动汽车碳排放量可知,OSC 算法能较好地电动汽车充电行为进行调度,在满足用户充电需求的同时降低对电网稳定性的影响,减少碳排放。

折扣因子是降低时间差分误差的一个重要参数。从图 3 中可以看出,折扣因子会影响 OSC 算法的收敛性。我们仿真了一组折扣因子为 0.005, 0.01 和 0.05 下 OSC 算法的收敛过程,可以看到较低的值可以获得良好的性能。因此本文选择折扣因子 0.01 来平衡奖励和时间效率,实现合理的快速收敛和高奖励。

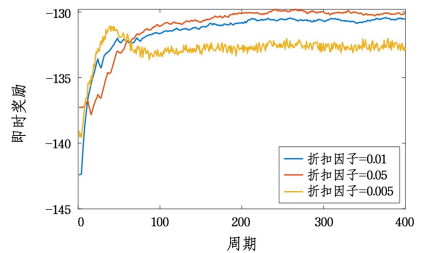


图 3 即时奖励的收敛性能

Fig. 3 Convergence performance of instant rewards

从图 4 中可以看出,当充电碳排放动作空间离散成更多的切片时,充电碳排放成本就会更低。有 40 辆电动汽车进行仿真时,OSC 算法的运行时间为 476 s。说明 OSC 算法具有较好的时间效率。与 OSC 算法相比,算法 II 离散充电碳排放空间切片越多,计算复杂度越大,但充电碳排放成本较低。当将充电动作空间离散成更少的切片时,算法 II-33 的计算复杂度与 OSC 算法接近。但是,算法 II 的充电动作空间被离散成

更多的切片,随着电动汽车数量的增加,计算复杂度迅速增加。而本文所提出的 OSC 算法的计算复杂度增加速率较小。

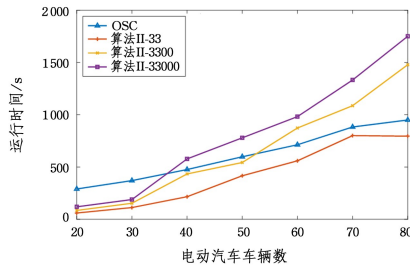


图4 算法的运行时间对比

Fig. 4 Comparison of running time of algorithms

结束语 本文在未来信息未知的情况下研究了一个电动汽车在线充电碳排放决策问题,以在碳达峰的约束条件下稳定碳排放量,降低电动汽车的充电碳排放成本。仿真结果表明,所提出的 OSC 算法下的充电碳排放成本性能分别比基准算法 I 和算法 II 高出 24.03% 和 21.49%,同时实现了良好的收敛性。OSC 算法下的碳排放量峰值相比算法 I 和算法 II 减少了 11.29% 和 10.64%,且 OSC 算法比算法 I 和算法 II 的波动更小。未来的工作中将考虑使用多智能体强化学习来协调多个充电桩的碳排放决策并考虑车辆到电网支持辅助服务的场景。

参考文献

- [1] YANG F, ZHANG J J. Current Status and Prospect of Low-carbon Development in China's Power Industry under the Carbon Peak and Carbon Neutrality Goals[J]. Environmental Protection, 2021, 49(17): 720-746.
- [2] TAN Q L, DAI M, MEI S F. Research on power system dispatch considering carbon quota and demand response of electric vehicles [J]. Power System and Clean Energy, 2021, 37(7): 79-86.
- [3] 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.
- [4] SHU Y B, ZHANG L Y, ZHANG Y Z, et al. Research on China's Path to Carbon Peak and Carbon Neutrality in the Power Sector [J]. Engineering Sciences in China, 2021, 23(6): 1-14.
- [5] HU Y, CHEN C, HE J, et al. IoT-based proactive energy supply control for connected electric vehicles [J]. IEEE Internet of Things Journal, 2019, 6(5): 7395-7405.
- [6] FU Q, DU W, WANG H, et al. Stability Analysis of DC Distribution System Considering Stochastic State of Electric Vehicle Charging Stations [J]. IEEE Transactions on Power Systems, 2022, 37(3): 1893-1903.
- [7] QI X, WU G, BORIBOONSOMSIN K, et al. Barth. Development and evaluation of an evolutionary algorithm-based online energy management system for plug-in hybrid electric vehicles [J]. IEEE Transactions Intell. Transp. Syst., 2017, 18(8): 2181-2191.
- [8] LI Y, WANG X, KANG Q, et al. An MCTS-Based Solution Approach to Solve Large-Scale Airline Crew Pairing Problems [J]. IEEE Transactions on Intelligent Transportation Systems, 2023, 24(5): 5477-5488.
- [9] LI P, WANG H, ZHANG B. A distributed online pricing strate-

gy for demand response programs [J]. IEEE Trans. Smart Grid, 2019, 10(1): 350-360.

- [10] QUDDUS M A, SHAHVARI O, MARUFUZZAMAN M, et al. A collaborative energy sharing optimization model among electric vehicle charging stations, commercial buildings, and power grid [J]. Appl. Energy, 2018, 229: 841-857.
- [11] NEMATKHAH F, BAHRAMI S, AMINIFAR F, et al. Exploiting the Potentials of HVAC Systems in Transactive Energy Markets [J]. IEEE Transactions on Smart Grid, 2021, 12(5): 4039-4048.
- [12] DUAN J, XU H, LIU W. Q-learning-based damping control of wide area power systems under cyber uncertainties [J]. IEEE Trans. Smart Grid, 2018, 9(6): 6408-6418.
- [13] RUELENSF, CLAESSENS B J, VANDAEL S, et al. Residential demand response of thermostatically controlled loads using batch reinforcement learning [J]. IEEE Trans. Smart Grid, 2017, 8(5): 2149-2159.
- [14] HU H X, WEN C, WEN G. A Distributed Lyapunov-Based Redesign Approach for Heterogeneous Uncertain Agents With Cooperation-Competition Interactions [J]. IEEE Transactions on Neural Networks and Learning Systems, 2022, 33(11): 6946-6960.
- [15] SUTTON R S, BARTO A G. Reinforcement learning: An introduction [J]. IEEE Trans. Neural Netw., 1998, 9(5): 1054.
- [16] ZHANG Z J, LIU F Z, LIU T, et al. A Persistent-Excitation-Free Method for System Disturbance Estimation Using Concurrent Learning [J]. IEEE Transactions on Circuits and Systems I: Regular Papers, 2023, 70(8): 3305-3315.
- [17] WANG X, DING D, DONG H, et al. Neural-network-based control for discrete-time nonlinear systems with input saturation under stochastic communication protocol [J]. IEEE/CAA Journal Automatica Sinica, 2021, 8(4): 766-778.
- [18] CHEN Z, LU Y, XING Q, et al. Analysis of Power System Scheduling Considering Carbon Quota of Electric Vehicles [J]. Automation of Electric Power Systems, 2019, 43(16): 44-51.
- [19] 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.
- [20] MA Z, CALLAWAY D S, HISKENS I A. Decentralized Charging Control of Large Populations of Plug-in Electric Vehicles [J]. IEEE Transactions on Control Systems Technology, 2013, 21(1): 67-78.
- [21] KIM B, PAIK M, KIM Y, et al. Distributed Electric Vehicle Charging Mechanism: A Game-Theoretical Approach [J]. IEEE Transactions on Vehicular Technology, 2022, 71(8): 8309-8317.



CAO Yongsheng, born in 1991, Ph. D., lecturer. His main research interests include data security, energy management, electric vehicle, and AI with big data.