在线 车辆到建筑系统的不确定性下的决策

RISHAV SEN，范德堡大学，美国YUNUO ZHANG，范德堡大学，美国FANGQI LIU，范德堡大学，美国JOSE PAOLO TALUSAN，范德堡大学，美国AVA PETTET，日产先进技术中心—硅谷，美国YOSHINORI SUZUE，日产先进技术中心—硅谷，美国AYAN MUKHOPADHYAY，范德堡大学，美国ABHISHEK DUBEY，范德堡大学，美国

arXiv:2601.03476v1 [eess.SY] 7 Jan 2026

车网楼（V2B）系统将智能建筑和电动汽车（EV）等物理基础设施，与连接至建筑充电桩的设备相结合，并通过数字控制机制管理能源使用。通过将电动汽车作为灵活的能源储备，建筑可以在分时电价和需量电费政策下动态地为其充电和放电，从而优化能源使用并降低成本。这种设置形成了V2B优化问题：建筑需要协调电动汽车的充电和放电，在满足用户充电需求的同时，最大限度地降低总电费。然而，V2B优化问题面临以下挑战：（1）波动的电价，包括能源费用（美元/千瓦时）和需量电费（美元/千瓦）；（2）较长的规划时域（通常超过30天）；（3）具有不同充电速率、可控性和充放电方向（即单向或双向）的异构充电器；以及（4）用户特定的电池电量离开时的状态，以确保满足用户要求。不同于现有方法通常将这一场景建模为单次组合优化问题，我们指出了以往研究中的关键局限，并将V2B优化问题建模为马尔可夫决策过程（MDP），即一种随机控制过程。由于状态空间和动作空间规模庞大，求解由此产生的MDP颇具挑战。为应对大规模状态空间带来的挑战，我们采用在线搜索；同时，通过使用领域专用启发式方法剪枝缺乏前景的动作，缓解动作空间带来的问题。我们与日产先进技术中心—硅谷合作验证了该方法。利用其电动汽车测试平台的数据，我们表明，所提出的框架显著优于最先进的方法。

其他关键词和短语：蒙特卡洛树搜索（MCTS）、优化、车网楼（V2B）、电动汽车充电

1 引言

电动汽车（EV）正推动全球能源格局向更可持续的能源解决方案转变 [17]。管理电动汽车的能源需求既带来了机遇，也带来了挑战。在本文中，我们聚焦于其中一种机遇——车辆到建筑充电（V2B），即对电动汽车和智能建筑的能源管理进行协同优化。V2B 充电背后的核心理念是利用对充电速率的控制能力，并将双向电动汽车作为储能设施，从而增强智能建筑的韧性和需求响应能力 [31]。例如，可以在能源价格较低且能源供应充足时为电动汽车充电，并在能源成本较高时放电，为建筑供能，从而促进能源优化利用，减少峰值功率需求（计费周期内的最高瞬时功率使用量），大幅降低建筑的能源成本和需求费用 [26]。随后，可以将部分节省的费用直接或通过优惠充电的方式分享给电动汽车车主，由此为建筑和电动汽车车主构建一个 双赢 框架。

尽管车辆到建筑框架看似简单，但在实际运行中仍面临诸多挑战。虽然可以通过提供低价（或免费）充电来激励电动汽车车主参与此类计划，但仅以优化能源成本为目标的策略，可能会导致车主离开智能建筑时电动汽车的电量低得任意，从而给建筑所有者带来不便。因此，车辆到建筑优化框架必须确保电动汽车车主在离开时间时，车辆电量达到预先规定的水平，同时应对建筑负荷波动、电动汽车到达（和离开）以及能源价格变化等外生不确定性。由于一天中会有许多电动汽车陆续到达，且各车辆的充电需求不同，优化问题会变得极其复杂。现代建筑还使用异构电动汽车充电器（其品牌、型号、充电速率和充放电方向各不相同），这进一步增加了不确定性下序贯决策过程的复杂性。此外，

作者联系方式：Rishav Sen，范德堡大学，美国纳什维尔，rishav.sen@vanderbilt.edu；Yunuo Zhang，范德堡大学，美国纳什维尔，yunuo.zhang@vanderbilt.edu；Fangqi Liu，范德堡大学，美国纳什维尔，fangqi.liu@vanderbilt.edu；Jose Paolo Talusan，范德堡大学，美国纳什维尔，jose.paolo.talusan@vanderbilt.edu；Ava Pettet，日产先进技术中心—硅谷，美国圣克拉拉，ava.pettet@nissan-usa.com；Yoshinori Suzue，日产先进技术中心—硅谷，美国圣克拉拉，yoshinori.suzue@nissan-usa.com；Ayan Mukhopadhyay，范德堡大学，美国纳什维尔，ayan.mukhopadhyay@vanderbilt.edu；Abhishek Dubey，范德堡大学，美国纳什维尔，abhishek.dubey@vanderbilt.edu。

车辆到建筑框架必须能够适应电力公司的复杂定价政策，例如分时电价（每千瓦时（kWh）的费用）和需量电费（每千瓦（kW）的费用）；至关重要的是，需求成本是在相对较长的时间跨度内计算的，例如，对于大多数美国电力公司而言，这涉及计算一个月内的最大瞬时功率利用率。

此前的研究已经探讨了车到建筑问题中的底层优化问题；例如，Tanguy 等人 [29] 为这一问题场景提出了最全面的优化框架之一。其模型简洁且极易求解——决策问题（即确定何时以及以多大程度为每辆车充电或放电）被建模为线性规划，可在多项式时间内高效求解，因此从设计上便具有可扩展性。然而，这种固有的可扩展性是以若干不符合实际的假设为代价的：1）一次性优化方法假设车辆的到达和离开时间均已预先知晓，并且一次性计算所有车辆到充电桩的分配，无法适应动态变化，例如能源价格或建筑负荷的变化；2）该公式的线性特性隐含地依赖于同质充电桩配置；3）该公式仅考虑能源成本，却无法纳入需求成本，而需求成本在实际中是决定车辆到建筑策略的重要因素。我们通过将车到建筑问题表述为随机控制过程，并将其建模为马尔可夫决策过程（MDP），来解决这些局限。我们的公式基于这样的观察：在实际中，必须在外生不确定性来源的影响下，依次决定何时为车辆充电或放电，以及充放电量。

然而，在我们的问题设置下，为该马尔可夫决策过程计算最优策略非常具有挑战性，尤其是由于复杂的信用分配和维数灾难 [28]。具体而言，鉴于决策必须以相对较高的频率计算（例如，每15分钟一次），而需求成本只有在计费周期结束时（通常为一个月）才能观测到，决策者的动作会产生长期影响，但奖励或反馈可能要过很久才能获得。此外，随着规划时域的增加，可能的状态和动作数量会呈指数增长。这种增长会导致组合爆炸，使得考虑所有可能的动作和状态序列在计算上变得不可行。为应对这些挑战，我们提出了一种在线方法，专注于为系统的当前状态计算近似最优动作，而不是试图学习一种策略。我们的方法基于蒙特卡洛树搜索（MCTS） [9], ，这是一种用于随机控制过程的通用在线算法。

尽管蒙特卡洛树搜索通过采用基于多臂老虎机的策略，在决策空间中探索有潜力的轨迹，从而应对车辆到建筑场景中的部分挑战，但它受到动作空间维数灾难的影响，使其无法适用于我们的问题场景（如下文所示）。为应对这些挑战，我们提出了*DG-MCTS*（领域知识引导的蒙特卡洛树搜索）。该方法采用动作生成框架，利用领域知识以及V2B优化问题的底层结构。该动作生成框架大幅缩减了决策空间。我们借鉴成熟的启发式方法，并通过加入随机生成的动作来扩充动作集合，以确保搜索树不会局限于这一经过截断的集合。我们与日产先进技术中心—硅谷（NATC-SV）合作，在真实世界数据上开展了大量仿真，结果表明，我们的方法优于现有方法。为进一步确保我们的方法能够扩展到规模极大的问题实例，我们还提出了一种分布式搜索算法，该算法以牺牲（小幅）*small* 程度的性能为代价，换取计算时间的节省。总而言之，我们的贡献如下：

- 我们提出了一种决策理论形式化方法，用于描述车辆到建筑问题，将电动汽车与智能建筑之间的交互建模为随机控制过程。
- 我们展示了如何利用基于领域专用启发式方法的动作生成模块和随机增强步骤对 MCTS 进行扩展，从而为该 MDP 计算出近似最优动作。
- 我们还提出了该算法的去中心化版本，其运行速度显著更快，但代价是性能略有下降。

- 我们与 NATC-SV 合作使用真实世界数据，并展示了所提出的方法优于具有竞争力的基线方法。

2 相关工作

我们强调了解决车到建筑问题的四项主要挑战，即：1）电动汽车到达和离开时间所带来的不确定性；2）分时电价下动态变化的能源成本和建筑负荷；3）需量电费和长期奖励；以及4）充电桩的异质性；

固有不确定性：线性规划（LP）和混合整数线性规划（MILP）等传统优化方法广泛应用于车辆到建筑系统，以在满足约束条件的同时降低成本 [11, 29]。然而，这些方法假定已准确掌握功率使用量和电动汽车的行程安排，这在现实应用中并不切实际。为应对这种不确定性，研究人员探索了随机优化 [20] 和鲁棒控制方法 [16]。利用循环神经网络（RNN）等模型进行概率预测，可以捕捉时间依赖关系 [34]。强化学习（RL）也已被用于学习自适应充电策略 [2, 25]。尽管取得了这些进展，长期不确定性的管理仍然面临挑战。深度学习模型难以处理长期依赖关系，而强化学习算法在复杂环境中则面临样本效率低和收敛问题 [1]。对极端天气、费率突然变化或电动汽车行为变化等罕见事件的接触有限，也会进一步降低适应能力，导致次优解。

分时电价：分时电价下能源价格的波动增加了车辆到建筑优化的复杂性，需要制定能够适应动态电费的充电计划，以尽可能降低费用 [6]。启发式方法和机器学习算法，以及深度强化学习，在学习能够响应分时电价波动的最优充电策略方面发挥了作用 [13]。然而，由于信用分配困难，以及智能建筑和电动汽车用户呈现的能源使用模式具有非平稳性，这些方法面临挑战 [12]。

需量电费与长期奖励：预测时域与计费周期不匹配，使车辆到建筑优化十分困难。现有采用模型预测控制进行需量电费预测的研究，受到较长计费周期所带来的计算复杂度限制。因此，这些研究转而采用仅涵盖单日的较短预测时域，导致决策并非最优，尤其是在峰值需求出现在预测时域之外时 [15, 19]。

充电器异质性：智能建筑通常会逐步部署电动汽车充电基础设施，因此充电器种类繁多，其额定功率和充放电方向各不相同。这进一步增加了优化动作空间的复杂性。已有一些研究处理了异质性问题 [7]。然而，这些研究将规划限制在单日范围内，或完全忽略需量电费，因此牺牲了长期奖励。

蒙特卡洛树搜索的作用：蒙特卡洛树搜索（MCTS）等基于搜索的算法，为应对车辆到建筑系统中固有的不确定性提供了一种有前景的替代方案。MCTS 可以通过模拟大量可能的未来情景来处理随机环境，因此非常适合在不确定性下进行规划 [3, 18]。与传统的强化学习方法不同，MCTS 能够构建搜索树，纳入未来状态和奖励，从而有效地进行长时域规划 [21]。这一能力使其能够处理长期依赖关系，并通过持续利用新信息更新搜索树，适应电动汽车行为的突然变化 [14]。

3 问题描述

我们首先介绍问题设定。回顾一下，在我们的设定中，电动汽车车主按照规律的模式到达和离开，且电池使用曲线具有可预测性。虽然通常无法提前获知建筑能耗，但我们假设可以使用一个基于历史数据预测建筑能耗的预测模型，而且该模型能够非常可靠地完成预测 [30, 33]。

3.1 规格

时段：所考虑的总时间（计费周期）被划分为有限个离散时段。时段 $t \in \mathrm{T}$从 t^{start} 开始，到 t^{end}结束。每个时段的持续时间为 $\delta_t = t^{\text{end}} - t^{\text{start}}$小时。根据电力公司的定价政策，一天中的时间可划分为高峰 时段和非高峰 时段。电力公司通常会在高峰时段 $\mathrm{T}^{peak} \subset \mathrm{T}$ 内，根据长度为 τ的时段聚合（即$t_j^\tau = \{t_i,\ t_{i+1},\ \ldots\ldots,\ t_{i+\tau}\}$表示第 j^{th} 个聚合时段，而 i是对应时段的索引）来计算需求成本（稍后解释）。每个聚合时间槽的持续时间为 $\delta_a = \delta_t \cdot \tau$。随后，高峰时段被划分为 Ω 个聚合时段 $\{t_1^\tau,\ t_2^\tau,\ \cdots\ t_\Omega^\tau\}$。我们使用记号 T_j^τ 表示所有这类 $t_i \in t_j^\tau$集合。例如，如果高峰时段为凌晨 12 点至凌晨 1 点，每个时段为 5 分钟，且 τ为 15 分钟，那么

第 1 个聚合　　时段（$j = 0$），即凌晨 12 点至凌晨 12:15，对应表示为 $t_0^\tau = \{t_0,\ t_1,\ t_2\}$，则 $\mathrm{T}_0^\tau = \{0,\ 1,\ 2\}$。

分时电价：电力公司通常根据两项费用或收费来计算成本——能源费用和需量电费。有些公司采用可变的能源费用费率，即使用分时定价模型。能源费用（美元/千瓦时）可能因高峰时段和非高峰时段而异，并由集合 W 表示，其中 w_t^e 是时段 $t \in \mathrm{T}$ 的分时电价。

需量电费：成本计算的另一部分是需量电费 w^d（美元/千瓦）。这是一项费用，根据计费周期内某一特定时间段中最高的用电速率确定。通常只考虑高峰时段内的用电量。

车辆：我们用 V 表示电动汽车的集合。假设车辆 $v \in \mathrm{V}$ 的电池容量为 $[e_{min}^v, e_{max}^v]$，其在时间 t 的荷电状态记为 e_t^v。每辆车辆均可在其到达和离开时段 $[t_{start}^v, t_{end}^v]$ 之间充电；由于我们的随机问题建模，t_{end}^v 对求解器而言是未知的。相反，电动汽车用户会提供一个离开时间窗口 T_{end}^v。用户到达时的荷电状态（SOC）为 $e_{t_{start}^v}^v$，并要求在离开时达到 e_{req}^v 的荷电状态，其中只有 $e_{t_{start}^v}^v$ 在到达时已知。用户离开时的荷电状态可能不同于其要求值 e_{req}^v，并记为 $e_{t_{end}^v}^v$。我们将在后文说明，当用户的要求未得到满足时如何对其进行补偿。我们的电动汽车合作伙伴的车辆支持双向运行（充电和放电），也可以对该问题进行修改，使其仅适用于单向（仅充电）场景。电动汽车的充电速率表示为 c_t^v。电动汽车在一个聚合时间槽内用于充电的功率表示为 $c_{t_j^\tau}^v = \sum_{i=1}^{\tau} c_{t_i}^v$。

建筑物功耗：我们用 $\mathrm{B}^e = \{b_1^e, b_2^e, \cdots\ b_{|\mathrm{T}|}^e\}$（单位：千瓦时）表示建筑的能源消耗。该数值用于计算电力公司的能源费用。高峰时段的功率消耗表示为 $\mathrm{B}^p = \{b_0^p, b_1^p, \cdots\ b_\Omega^p\}$，对应每个聚合时间槽 t_j^τ，该时间槽涵盖 $t_i, \cdots\ t_{i+\tau}$ 个时段，$b_j^p = 1/\tau \cdot \sum_{i=1}^{\tau}(b_i^e/\delta_t)$，并用于计算需求成本。充电：充电器的控制模式（可控或不可控）、充放电方向（单向或双向）以及最大功率（例如，10 千瓦时、20 千瓦时）各不相同。可控充电器可根据需要开始和停止充电，而不可控充电器则持续充电直至充满。单向充电器仅向车辆供电，而双向充电器还可放电，从而实现车辆到建筑充电。每种充电器类型由这些属性的组合定义（例如，20 千瓦时的可控单向充电器）。所有充电器类型的集合表示为 K，共有 N 个充电器。每种类型 $k \in \mathrm{K}$ 有 $|k|$ 个可用充电器，因此 $\sum |k| = N$。充电器采用线性充电曲线，在每个时间步均匀地为电动汽车充电，对优化 [27] 的影响很小。我们将在消融研究中将其与分段充电曲线进行比较。

充电速率：我们将充电速率定义为每个时段输送给电动汽车的功率，单位为千瓦时。我们假设电动汽车能够以充电器支持的最大功率进行充电和放电，该功率范围为 $[q_k^{min}, q_k^{max}]\ \forall k \in \mathrm{K}$；其中，当充电器为双向充电器时，$q_k^{min}$ 可以为负值，表示已连接电动汽车正在放电。充电器的效率用 η 表示，适用于充电和放电。我们还定义了电动汽车与充电器类型的占用函数 $\zeta\colon \mathrm{V} \times \mathrm{T} \to \mathrm{C}$，其中 $\zeta^v(t) = k_i$，表示电动汽车 v 在时间 t 连接到类型为 k_i 的充电器。一旦充电器分配给某辆电动汽车，在其离开之前就不能更换充电器。

峰值功率使用量：聚合时间槽 t_j^τ 内的峰值功率记为 π_j。回顾一下，一个聚合时段 t_j^τ 由若干较小的时段 $\{t_i, t_{i+1}, ..., t_{i+\tau}\}$ 组成；峰值功率通过对每个聚合时间槽 t_j^τ，$\pi_j = b_j^p + \sum_{i=1}^{N} c_{t_j^\tau}^v$ 内的功率消耗（建筑和充电桩所用功率之和）取平均值来计算。因此，高峰时段内使用的峰值（最大）功率为 $P^{max} = \max(\pi_j)$。

需求费用：电力供应商会根据客户在特定周期（通常为一个月份）内的峰值用电功率（通常以千瓦（kW）计量）收取需求费用。该费用不同于能源成本，后者取决于使用的能源总量（千瓦时（kWh））。需求费用等于聚合时间槽内的峰值功率与需量电费的乘积，计算方式为 $w^d \cdot P^{max}$。由于需求费用的评估具有不确定性且缺乏灵活性，模型难以针对较长计费周期内的需求费用进行优化和规划。因此，需要以非短视的方式计算需求费用。

能源成本：时段 s 内使用的总能源量，是车辆再充电所用能源与满足建筑能源需求所用能源之和。我们使用 $g_t = w_t \cdot \sum_{v \in \mathcal{V}}(c_t^v + b_t^e)$ 表示该时段的能源成本。将计费周期内所有时段的能源成本相加，即可得到总能源使用成本。在电价较高的时段，通过对已连接电动汽车进行放电，我们可以在满足建筑能源需求的同时，有效降低总体电费。这些电动汽车随后会在电价较低的时间重新充电。

SoC缺失成本：这是车辆到建筑（V2B）中的一项关键指标，用于衡量每辆电动汽车在离开时荷电状态的实际值与要求值之间的能源缺额

问题。在我们的设定中，由于每辆电动汽车的离开时间未知，因此这项指标对于了解用户要求的满足程度至关重要。我们按 w^s的费率，以货币价值计量缺失电量，单位为美元/缺失千瓦时。

总账单：计费周期内的总电费是需求成本与能源成本之和。直观的策略是在整个计费周期内尽量降低峰值功率，从而减少需求成本，并确保高峰时段的能源成本保持在满足电动汽车充电至所需荷电状态水平这一约束条件所需的最低水平。其表示如下：

$$\sum_{t\in\mathcal{T}} w_t^e \cdot (b_t^e + c_t^v) + w^d \cdot P^{max} + \sum_{v\in\mathcal{V}} w^s \cdot |e^v_{t^v_{end}} - e^v_{req}| \tag{1}$$

解空间： 我们将解定义为充电桩与汽车的分配，以及相应的充放电计划。令 H 为解的集合；对于每项充电分配（k， t），恰好分配一辆电动汽车 $v \in$V，使得 v， k， t $\in$H，这符合实际使用情况：一辆电动汽车只连接到一个充电桩，不会在不同充电桩之间切换。对于每项分配$h_{v,\ k,\ t} \in$H，我们分配一个充电速率 $c_t^v \in C$。我们始终假设电动汽车在整个停留期间都连接在充电桩上。我们按照先到先服务原则分配汽车，优先依据汽车的到达时间，以及充电桩的充电速率、充放电方向和可控性进行分配。例如，一种优先顺序可以是先将汽车分配给 20 千瓦时双向充电桩，再分配给 20 千瓦时单向充电桩（均为可控充电桩），最后分配给不可控充电桩。我们只能为数量不超过可用充电桩的汽车提供充电。没有空闲充电桩时到达的多余汽车，在剩余停留期间都无法充电。

3.2 马尔可夫决策过程（MDP）

我们基于 Shi 和 Wong 的先前研究，将车到建筑问题建模为马尔可夫决策过程（MDP） [22]。MDP 由一个四元组 $\{\mathcal{S}, \mathcal{A}, \mathcal{P}\ \rho\}$刻画，其中 $\mathcal{S}$ 是状态集合，$\mathcal{A}$ 是可能动作的集合，$\mathcal{P}$ 是状态-动作转移模型，而 ρ 是用于刻画智能体效用（或成本）的奖励函数 [8]。

决策时刻：每个离散的决策事件都会产生一个决策时刻， $t \in$T。在事件之间，环境以连续时间运行，充电桩为电动汽车充电或放电。在每个决策时刻，决策者采取一个动作，使状态从决策前状态转变为决策后状态，随后获得即时奖励。

状态：我们将状态集合表示为 S $\in\{\{b_t^e\}, \hat{}P^{max}, \{q_k^{min}, q_k^{max}\}\ \forall k \in$K, $\{e_t^v, t^v_{start}, \hat{}t^v{}_{end}, e^v_{min}, e^v_{max}\} \forall v \in$V$\}$，时段 t的状态在决策前时刻表示为 s_t。其中包括当前建筑负荷（b_t^e）、所有充电器功率（q_k^{min}, q_k^{max}）、电动汽车详情（$e_t^v, t^v_{start}, \hat{}t^v{}_{end}$, e^v_{min}, e^v_{max}），其中 $\hat{}\ t^v{}_{end}$是根据用户离开时间窗口 $\mathrm{T}^{\ v}_{end}$估算的离开时间。我们还纳入峰值功率估计值 $\hat{}P^{max}$，以便做出更好的决策。

动作：我们将时段 t 的所有可行动作集合记为 A_t。A_t 中的一个动作对应于所有充电桩 K 的充电速率组合。充电速率受建筑总功率消耗的限制，其中包括为车辆充电所提供的功率；也就是说，电动汽车的放电量不能超过建筑当前的功率使用量。充电速率被限制在最大和最小允许值之间的离散值中。

状态转移：在时段 t，决策者可以采取一个动作，使状态 s 转移到 s'。在这一转移过程中，系统会经历许多不同的随机过程。首先，电动汽车可能会在常规到达和离开时间之外到达或离开，其情况由某种停留时长分布决定。其次，电价可能在几乎没有预警的情况下发生变化，例如在紧急负荷削减计划（ELRP）事件期间，消费者会因减少能源消耗而获得经济激励。最后，建筑的功率消耗可能随一天中的时刻、一年中的月份甚至天气而变化。虽然这些变化遵循预测分布，但仍会给状态转移带来随机性。我们不详细讨论时间转移和状态转移概率的数学模型及表达式，因为我们的框架仅依赖生成式世界模型，而不依赖对这些转移的显式估计。

奖励：马尔可夫决策过程中的奖励通常包含两个部分：执行动作时获得的一次性即时奖励，以及随着过程演进而产生的连续时间奖励；奖励还高度依赖于具体领域。对于充电桩优化，我们关注的是在满足用户一定服务质量的同时，最大限度地降低总体能源使用成本。我们用 $\rho(s,\ a)$表示在状态 s下执行动作 a所获得的奖励。奖励既包括中间奖励 r_t^i，也包括回合奖励 r_t^e。我们在轨迹展开过程中使用回合奖励。中间奖励为 $r_t^i = g_t + w^s \cdot \sum_{v\in\mathcal{V}} \left|e^v_{t^v_{end}} - e^v_{t^v}\right|$，如果 $t = t^v_{end}$；否则为 $r_t^i = g_t$。回合奖励 r_t^e根据公式（1）计算，只对需求成本的计算做了细微修改：使用峰值功率的估计值 $\hat{}P^{max}$。具体如下：

:

$$r_t^e = \sum_{t\in\mathcal{T}} w_t^e \cdot (b_t^e + c_t^v) + w^d \cdot \hat{P}^{max} + \sum_{v\in\mathcal{V}} w^s \cdot |e^v_{t^v_{end}} - e^v_{req}| \tag{2}$$

4 在线方法

我们提出一种在线方法来管理充电控制。首先，我们采用启发式电动汽车分配策略，例如 先到先服务 和双向优先策略，在考虑公平性以及双向充电器削峰能力的同时，将电动汽车分配给充电桩。为应对电动汽车到达与离开的不确定性，我们通过采样数据来估计未来系统状态，并利用离线求解器推导最优动作、确定性能指标的上界（见附录 A 节）。A。

为适应环境不确定性，我们采用在线蒙特卡洛树搜索（MCTS），以实现动态且稳健的决策。鉴于 MCTS 在实时场景中的运行时间面临挑战，我们引入领域知识引导（DG-MCTS），利用启发式动作和需量电费预测来缩减并调整动作空间，从而提高探索效率。我们的在线 DG-MCTS 求解器利用电动汽车运营合作伙伴日产提供的回合数据，构建前瞻性搜索树。为降低计算复杂度，我们将较长的计费周期（月度）分解为更短且易于管理的每日规划时域。算法 1 概述了 DG-MCTS 方法的工作流程。下一节将详细介绍该方法的各个组成部分，包括电动汽车分配、未来状态估计，以及如何将领域知识引导的探索融入 MCTS。

4.1 情节采样与不确定性处理

情节样本：我们使用来自美国加利福尼亚州圣克拉拉的日产先进技术中心硅谷分部 (NATC-SV) 的真实世界数据对泊松分布进行建模。
车辆与充电桩分配：我们根据到达时间，按照先到先服务策略为电动汽车分配充电桩。如果多辆汽车在同一时段到达，则根据 e^v_{req}对它们排序以打破平局。我们根据充电桩的充放电方向、最大功率和可控性为电动汽车分配充电桩。

离开时间的不确定性：鉴于每位用户仅提供离开时间窗口 T^v_{end}，我们的目标是估算该区间内的离开时间。由于 T^v_{end} 表示的是一段可能离开时间的范围，而非固定时刻，我们将离开时间视为 T^v_{end}内的随机变量，以此对这种不确定性进行建模。我们从区间 T^v_{end} 上的均匀分布中抽取样本，以近似得到一个具有代表性的离开时间。

令 $\hat{}\,t^v_{end}$ 表示用户 v 的估计离开时间，它是 T^v_{end} 上均匀分布的随机变量的一次实现。我们从 $\hat{}\,t^v_{end} \sim \mathrm{U}(\min(\mathrm{T}^v_{end}), \max(\mathrm{T}^v_{end}))$ 中采样，其中 $\min(\mathrm{T}^v_{end})$ 和 $\max(\mathrm{T}^v_{end})$ 分别是离开时间窗口 T^v_{end} 的下限和上限。通过这种方式采样，我们可以选择一个可行的离开时间，且不会偏向区间内的任何特定时刻，从而确保对整个时间窗口进行公平估算。这可作为用户离开行为的代理。

估算峰值功率阈值：对于固定的电动汽车到达和离开轨迹，我们通过求解混合整数线性规划，在满足充电需求的同时最大限度地降低成本，从而确定最优充电决策。此过程还会估算月度峰值功率阈值。我们求解与评估数据不同的抽样回合，并根据最优动作计算峰值功率。样本中峰值功率的 99%置信水平为月度估算值提供了稳健依据。这种基于置信水平的阈值将需量电费纳入抽样过程，使其与混合整数线性规划解得出的最优充电结果相一致。

4.2 处理指数级动作空间

在我们的问题中，可行动作对应于所有充电器的一组充电器功率，因为充电器可能各不相同，并且可以关闭，也可以为已连接的电动汽车充电或放电。此外，充电器功率之和必须受该时刻当前建筑负荷的约束。因此，对于充电器数量众多且配置各异的建筑，如果我们为每个充电器考虑 p 个离散动作，那么给定状态下可能的动作数量将呈组合式增长，N^p。在在线场景中探索如此庞大的动作空间并不可行。为应对这一挑战，我们引入一种启发式方法，以便从可行动作集合中找出有前景的动作。

最小松弛度优先启发式方法： 我们的目标之一是为电动汽车充电，使其离开时达到期望的荷电状态。为指导充电决策，我们采用最小松弛时间优先（LLF）启发式方法 [32]。该启发式方法优先处理距离离开时间最短的电动汽车，确保即将进入离开时间窗口的车辆在充电分配中优先。在每个决策时刻，我们根据当前状态 s 和时间 t 计算可用功率差额 $= \hat{} P^{\max} - b_t^p$，其中 $\hat{} P^{\max}$ 是峰值功率阈值估计值，而 b_t^p 是当前建筑负荷。每辆电动汽车 v 的涓流充电率定义为涓流充电率 $= (e^v_{\text{req}} - e^v_t)/(\hat{} t^v_{\text{end}} - t)$，其中 e^v_{req} 是所需能量，e^v_t 是当前能量，而 $\hat{} t^v_{\text{end}}$ 是预计离开时间。随后，我们计算当前时段所有电动汽车的涓流充电率总和。如果该总和小于可用功率差额，则我们有能力进行过量充电。在这种情况下，我们将每辆电动汽车的充电速率设为其涓流充电率，并按照松弛度的反向顺序（最小松弛度优先），为配备双向充电器的电动汽车分配额外充电，直至功率差额全部用尽。如果涓流充电率总和超过功率差额，则我们让配备双向充电器的电动汽车放电，并优先选择离开时间最远的车辆，以填补差额，然后恢复涓流充电。

我们通过引入基于两个参数（功率差额和可行动作的边界）的直观动作，改进有前景的动作。如果可用功率差额为正，表明存在富余容量，我们便添加离散的增量充电动作，以探索过量充电的选项。否则，如果功率差额为负（即建筑负荷超过峰值功率估计），则加入额外的放电动作，以缓解过剩需求；我们将这些额外动作称为 β。最后，我们在所选动作的最小值与最低可行充电速率之间添加一个动作，$q_k^{\min}$，确保其不低于电动汽车所要求的最低荷电状态$e^v_{\min}$。同样，我们在所选动作的最大值与最高可行充电速率之间添加另一个动作，$q_k^{\max}$，确保其不超过电动汽车的最高允许荷电状态 $e^v_{\max}$；我们将动作空间中的这一变化称为 “ 偏移量 ”。

这种由 LLF 引导的动作具有双重作用：它通过缩小搜索空间并加快决策速度，提升了可追溯性。这使搜索过程保持较高的计算效率。通过根据当前状态（包括每辆电动汽车的紧迫程度）为动作排序，最小松弛时间优先启发式使我们能够快速调整充电速率，并将探索集中在最相关的动作上。

算法1 领域知识引导的蒙特卡洛树搜索（DG-MCTS）输入：当前状态 S_t、迭代次数 I、探索范围 β输出：最佳动作 $[A_k^* 对于 k \in K]$，1 $D \leftarrow$ 估计需量电费(S_t) // 估计需量电费2 $[A'_k 对于 k \in K] \leftarrow$LLF启发式方法($S_t$，$D$) // 获取启发式动作 // 动作剪枝 3 如果 $BuildingLoad < D$ 则 4 对每个 $k \in K$ 执行 // 鼓励充电5 $Space(A_k) \leftarrow$ 获取邻居($[A'_k - \beta$、$A'_k + \beta +$ 偏移量$] \cup$ 近边界动作) 6 否则 7 对每个 $k \in K$ 执行 // 鼓励放电8 $Space(A_k) \leftarrow$ 获取邻居($[A'_k - \beta -$ 偏移量、$A'_k + \beta] \cup$ 近边界动作)9 $ActionSpace \leftarrow [Space(A_K).Clip[q_k^{min}, q_k^{max}]$ 对于 $k \in K$]10 对于 $n \leftarrow 1$ 到 I 执行 // 建立树 11 $Sample \leftarrow$ 生成样本(EVArrTime、EVDepTime、BuildingLoad) // 移动到下一个状态，应用剪枝后的动作 12 $S'_{t+1} \leftarrow$ 从中选择(S_t、$ActionSpace$、$Sample$) // 移动到下一个状态，应用剪枝后的动作 13 $S'_{t+2} \leftarrow$ 扩展(S'_{t+1}、$Sample$) // 添加新的子节点14 $v \leftarrow$ 模拟(S'_{t+2}、$Sample$、涓流速率) // 模拟展开 15 反向传播(S'_{t+2}、v) // 更新树统计信息 16 $A^* \leftarrow$ 最佳动作(S_t) // 选择价值最高的动作

时间分解： 我们将较长的计费周期（通常为一个月）分解为更短且计算上易于处理的规划时域，例如一天。然而，这种分解带来了一个重大挑战：只有在完整计费周期结束时，即在较长的规划时域内，才能准确计算需量电费。为此，我们利用车网互动马尔可夫决策过程（MDP）的两个关键性质。

首先需要注意，电动汽车的到达与离开时间独立于智能体的充电动作——无论现有车辆如何充电，电动汽车都会到达建筑并离开。这种独立性使我们能够预先采样电动汽车的轨迹

，使用基于历史数据的生成模型生成电动汽车到达与离开轨迹，而无需根据搜索树中采取的动作对采样进行条件设定。

其次，我们利用前文所述的峰值功率阈值估计值。基于这一峰值功率估计值，我们引入时间分解的核心概念：将完整的规划时域划分为较短的时段，并增加约束，要求每个时段的总耗电量低于固定采样轨迹对应的需量电费估计值。

每天开始时（即较短的分解后规划时域开始时），我们从生成模型中采样多条电动汽车到达和离开的轨迹。对于每条轨迹，训练好的模型 f 会估计较长规划时域（例如一个月）内的需量电费。随后，我们将这一估计电费纳入搜索树中，作为回合奖励的一部分；搜索树则在较短的每日规划时域内运行。

4.3 蒙特卡洛树搜索（MCTS）评估

采用混合整数线性规划（MILP）等方法离线求解车辆到建筑问题时，未能考虑现实世界中的随机性。相反，这些方法依赖对系统的完整了解，以便最优地选择最佳动作。因此，我们采用蒙特卡洛树搜索（MCTS）；这是一种随时算法，已广泛应用于博弈场景 [23]。

MCTS 将规划建模为一棵树，其中状态是节点，动作是边。它以非对称方式探索这棵树，利用应用于树的上置信界（UCT）算法 [9], , 在利用与探索之间取得平衡。节点值通过模拟展开 进行估算，采用简单的默认策略，通常是随机动作选择。随着搜索的推进，节点值估算会逐渐改善。这种方法能够高效地探索庞大的动作空间。MCTS 需要生成式环境模型、用于导航的树策略，以及用于估算价值的默认策略。

我们利用收集到的历史数据，在树向未来扩展的过程中采样生成新的情节。我们使用标准的应用于树的上置信界（UCT） [10] 来遍历搜索树，并决定扩展哪些节点。扩展节点时，我们会针对给定状态采样有前景的动作。在 MCTS 树之外通过模拟展开估算动作价值时，我们依赖默认策略。该策略会模拟至指定的时间范围，并将效用沿树向上传播。我们的默认策略是涓流充电速率策略——按照预计的离开时间，为每辆车充入满足其目标荷电状态所需的电量。

根节点并行化： 鉴于电动汽车到达与离开以及建筑功率需求的时间都存在高度不确定性，即使遵循已知分布，抽取一个情节也可能无法代表电动汽车未来的实际行为。我们通过根节点并行化来应对这一问题：抽取多个情节，并为每个情节实例化一棵新的蒙特卡洛树搜索树，以其电动汽车到达与离开以及建筑功率需求作为根节点。每棵树并行探索；执行后，对各棵树中共同根节点的每个动作所得分数取平均值。随后，选择所有树中平均分数最高的动作。

4.3.1 集中式蒙特卡洛树搜索。 在集中式多智能体方法中，单棵搜索树代表车辆到建筑系统中所有智能体组合后的决策空间。不同于各智能体独立运行，这棵统一的树整合了所有智能体的动作，从而实现联合优化。这种方法在车辆到建筑场景中具有优势，因为单辆电动汽车的决策会影响整体建筑负荷和需量电费，因此需要协调以降低成本。然而，集中式蒙特卡洛树搜索也面临计算挑战：由于动作空间维度高、规划时域长，它必须在每个决策点同时考虑所有智能体的动作。我们在算法 1中展示这一过程。

集中式蒙特卡洛树搜索在每一步都明确考虑智能体之间的交互，这在需要协同配合的场景中尤其具有优势。例如，集中式蒙特卡洛树搜索可以直接纳入一辆电动汽车的充电动作对所有电动汽车及建筑需量电费的影响。然而，集中式蒙特卡洛树搜索必须应对两个主要挑战：**(1)** 在单一树结构中高效探索巨大的动作空间，以及**(2)**管理与V2B优化问题相关的长期时间跨度。我们通过利用领域专用启发式方法来解决这些挑战：这些方法能够简化动作空间的探索，并在蒙特卡洛树搜索框架内引入较短时间跨度的策略规划；我们将其命名为领域知识引导的蒙特卡洛树搜索（DG-MCTS）。

4.3.2 去中心化蒙特卡洛树搜索。 尽管DG-MCTS能够应对大规模状态空间，但在面对高维动作空间和较长时间跨度时仍然存在困难。随着汽车数量增加，单棵蒙特卡洛树搜索树可能会变得过大。为此，我们采用多智能体蒙特卡洛树搜索的去中心化方法（dMCTS） [4, 18]。

在 dMCTS 中，每个智能体都会构建自己的搜索树，以找到近似最优的动作。这缩小了每个智能体的搜索空间，并有助于系统扩展。然而，这也带来了两个挑战：**(1)** 在探索过程中，每个智能体都必须考虑其他智能体的动作。如果我们依赖代理来预测其他智能体的动作 [18], ，就有可能产生高昂的需求费用；**(2)**将动作空间分散化并不能解决长时间范围的问题。

我们利用车辆到楼宇（V2B）问题的结构，通过一种启发式方法应对这些挑战：**关键性评分**：我们根据智能体充电需求的紧迫程度对其排序。我们假设总能量消耗有限，因此优先安排必须在较短时间内满足较高充电要求的智能体。对于每辆已连接电动汽车 v，我们使用最小松弛度优先 [32], $c_{score} = (t^v_{end} - t) - (e^v_{req} - e^v_t)/q^{max}_{\zeta^v}$ t计算其关键性评分() 其中，t^v_{end} 是 v离开的时段， t是当前时段， e^v_{req} 是所需荷电状态（SOC）， e^v_t是当前荷电状态，而 $\zeta^v(t)$是与 v相连的充电器。松弛度较小（达到所需充电量后剩余时间较少）的智能体优先级较高。

流程：每个充电桩使用 dMCTS 进行决策，无需集中式控制器。这种设置降低了计算成本，因为每个智能体只关注自己的动作空间，同时仍会考虑其他智能体的选择。算法2详细说明了动作选择、模拟展开和反向传播的步骤。算法3 展示了如何根据关键性评分对汽车排序。这样的排序可确保在共享能源分配中优先满足需求紧迫的汽车，即使每个智能体运行的都是自己的本地搜索。

算法2 分散式蒙特卡洛树搜索（dMCTS） 输入：当前状态 S_t、迭代次数 N、探索范围 β输出：最佳动作 $[A^*_k \text{ for } k \in K]$ for 所有充电桩 $D \leftarrow$ 估算需量电费(S_t) // 估算需量电费 2 foreach $k \in K$ 按最小松弛度优先排序 do 3 $A'_k \leftarrow$ 最小松弛度优先策略(S^k_t, D) // 获取启发式动作 $Space(A_k) \leftarrow$ 获取邻居($[-\beta + A'_k$ + 噪声 $\beta + A'_k$ + 噪声$]$) for $n \leftarrow 1$ to N do // 为每个充电桩建立树 $Sample \leftarrow$ 生成样本(EVArrivalTime, EVDepartureTime, 建筑负荷) 6 $S^k_t \leftarrow S_t$7 $S^k_{t+1} \leftarrow$ 从($S^k_t, Space(A_k)$ $Sample$)中选择 // 移至下一个状态，应用剪枝后的动作 8 $S^k_{t+2} \leftarrow$ 展开($S^k_{t+1}, Sample$) // 添加新的子节点 $v \leftarrow$ 模拟展开($S^k_{t+2}, Sample, LLFPolicy$) // 模拟展开 10 反向传播($S^k_{t+2}$, v^k) // 更新树状态 for k11 $A^*_k \leftarrow$ 最佳动作(S_t) // 为 k选择最佳动作 12 $S_t \leftarrow$ 更新(S_t, A^*_k) // 更新下一个充电桩的状态

13 **返回** $[\mathcal{A}^*_k$对于 $k \in K]$

算法3 按关键性对汽车排序

输入： 汽车集合 V

输出： 基于关键性排序的汽车列表 V′

1 $\mathcal{V}^{critical} \leftarrow \emptyset$// 初始化关键性评分列表

2 **对于** $v \in \mathcal{V}$ **执行**

3 $c_v \leftarrow$计算关键性(v) // 计算关键性评分

4 $\mathrm{V}^{critical} \leftarrow \mathcal{V}^{critical} \cup \{(v, c_v)\}$ // 添加汽车 v及其关键性评分

5 $\mathcal{V}' \leftarrow$ 排序($\mathcal{V}^{critical}$,按 c_v降序) // 按关键性评分对汽车排序

6 **返回**$\mathcal{V}'$ // 返回已排序的汽车列表

5 实验与分析

5.1 设置

数据收集 为评估我们所提出方法的有效性，我们采用了来自 NATC-SV 研究实验室的数据。优化仅限于工作日，因为周末上班的员工很少，且需量电费通常不包括周末。硅

谷电力公司在计算需量电费时不计入周日。我们的设置包括 10 个单向充电器（20 千瓦时）和 5 个支持 ±20 千瓦时的双向充电器（以 20 千瓦时充电并以 20 千瓦时放电）。所有电动汽车均具备双向充电能力。

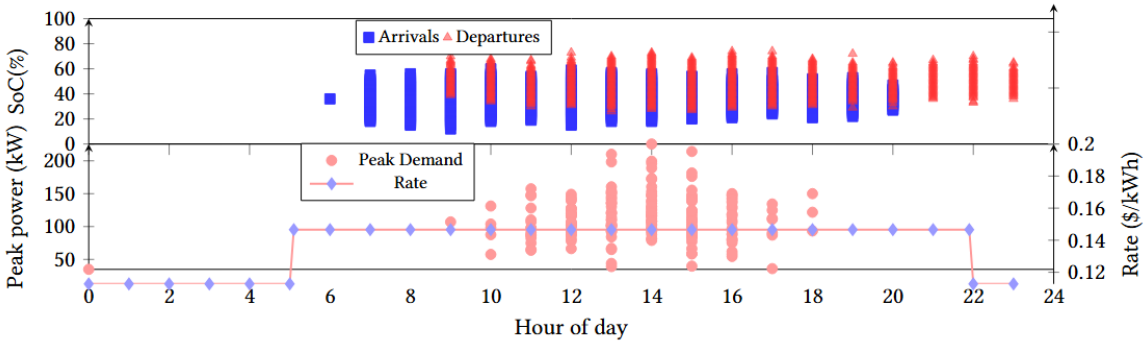


图 1。（上）8 个月内电动汽车到达与离开时间与到达时间及所需荷电状态的对比。（下）建筑峰值功率需求与一天中的时间及分时电价的对比。

我们收集了来自加利福尼亚州圣克拉拉 NATC-SV 的真实世界数据，涵盖 2023 年五月至 2024 年一月为期九个月的建筑用电量、电动汽车充电桩使用情况和电动汽车遥测数据。建筑负荷波动可以使用标准模型可靠地预测 [33], ，而电动汽车行为（到达、离开和荷电状态要求）则根据历史数据使用泊松分布进行建模。该方法捕捉了电动汽车到达率和荷电状态需求的变化，并使用每小时平均值来反映真实的使用模式。工作日到达的双向电动汽车数量每天都会波动，带来了固有的不确定性。车辆到达和离开是用户特有的行为，主要由工作时间决定，并且独立于建筑负荷或充电动作。我们与建筑运营商合作确认，到达和离开模式由用户决策决定，而不是由系统级控制决定。这种独立性使我们能够预先采样到达和离开时间，而不会影响优化。此外，当前的充电决策不会影响未来用户的到达或离开时间；优化过程必须在制定高效充电计划的同时，考虑用户行为的不可预测性以及现有限制条件。图 1 展示了到达和离开时间及荷电状态要求，以及峰值功率需求的分布和对应时间。我们从 2023 年五月至十二月采样了 110 个月度计费时段，以构建具有代表性的数据集。

电价采用圣克拉拉硅谷电力公司的费率：高峰时段（上午 6 点至晚上 10 点，周日除外）为每千瓦时 $0.147，非高峰时段为每千瓦时 $0.113；高峰时段的需量电费为每千瓦 $9.62。

峰值功率估计。 为了提高行动效果，我们根据混合整数线性规划求解器生成的最优行动序列，将每个情节对应的月度峰值功率估计纳入考量，以反映工作日条件的变化。我们采用混合整数线性规划数据中 99%置信区间的下界，作为需量电费的保守估计。此输入特征会在强化学习训练过程中进一步优化。

超参数调优。 为优化 DG-MCTS 框架的性能，我们采用了先进的超参数优化库 Optuna，该库使用高效的采样策略来探索超参数空间。它支持树结构帕岑估计器（TPE）等算法，可在超参数空间中自适应地采样有前景的区域，同时舍弃效果较差的配置。在 Optuna 中，每次试验都代表一组独特的超参数组合，并根据目标函数进行评估；结果则存储在研究数据库中，以供进一步分析。Optuna 的目标是最小化一个自定义分数，该分数根据公式（1）计算。超参数搜索空间包含若干关键参数：迭代次数、最大深度、未满足荷电状态要求和功率差额超限的惩罚，以及满足特定 SoC 目标的奖励。此外，我们还探索了正则化参数，如 C和奖励折扣因子 γ，以及峰值功率估计的容差参数（ϵ）。基于试验的方法生成了超参数重要性图，如附录图2所示，揭示了影响性能的最重要因素，并为后续模型改进提供指导。参数设置如表 1所示。

表 1. 蒙特卡洛树搜索配置的超参数

超参数	**限制**	**值**
迭代次数（仿真步数）	[50, 500]	200
最大树深度	[10, 80]	70
探索系数 (C)	[0.5, 2]	1.414
折扣因子 (γ)	[0.9, 1]	1
未达到荷电状态的惩罚	[0.1, 5]	0.5
超出功率缺口的惩罚	[0, 20]	5
满足所需荷电状态的奖励	[0.01, 1]	0.1
最大化荷电状态的奖励	[0.001, 1]	0.01
估计峰值功率调整（ϵ）	$[-10, 10]$	0
探索样本（并行树）	[5, 20]	10

表 2。月总成本（越低越好）。

策略	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC
DG-MCTS	5466.96 ± 24.4	**6032.16 ± 88.8**	6021.98 ± 42.8	**8512.24 ± 81.4**	**6357.79 ± 36.8**	**6744.54 ± 75.7**	**5806.73 ± 79.4**	**5195.53 ± 127.1**
dMCTS	5534.54 ± 84.2	6050.26 ± 60.0	6123.93 ± 73.1	8550.6 ± 53.3	6467.52 ± 75.9	6819.85 ± 62.6	5849.69 ± 56.9	5317.49 ± 157.0
RL	**5416.3 ± 32.8**	6067.6 ± 152.1	**5913.83 ± 21.4**	8571.87 ± 87.9	6403.82 ± 105.3	6852.93 ± 129.3	5898.55 ± 71.9	5432.92 ± 135.0
LLF	5515.86 ± 30.7	6068.35 ± 45.0	6045.75 ± 37.1	8637.07 ± 44.9	6364.37 ± 37.0	6802.74 ± 48.8	5831.15 ± 33.1	5245.1 ± 140.6
EDF	5521.67 ± 38.32	6076.35 ± 58.8	6047.83 ± 37.6	8637.67 ± 45.2	6369.2 ± 41.1	6810.84 ± 51.8	5831.53 ± 33.4	5292.71 ± 152.6
ReqCharge	5577.22 ± 43.8	6147.51 ± 47.1	6123.04 ± 35.0	8743.39 ± 53.7	6465.7 ± 46.0	6852.65 ± 56.9	5939.52 ± 42.2	5254.94 ± 65.5
MaxCharge	6827.5 ± 188.6	7710.79 ± 228.91	7577.06 ± 207.8	9402.73 ± 144.4	8259.19 ± 249.7	8348.78 ± 204.7	7081.51 ± 184.7	7888.98 ± 291.8

表3. 各策略的 SoC 缺额（越低越好）。

策略	DG-MCTS	dMCTS	RL	LLF	EDF	ReqCharge	MaxCharge
均值 ±标准差	113.62 ± 83.86	164.6 ± 174.24	141.92 ± 58.17	59.66 ± 84.13	58.86 ± 82.29	84.87 ± 80.73	**27.56 ± 60.55**

表 4. 按策略统计的 SoC 缺失车辆数（越低越好）。

策略	DG-MCTS	dMCTS	RL	LLF	EDF	ReqCharge	MaxCharge
均值 ±标准差	43.82 ± 30.98	73.16 ± 52.88	207.34 ± 60.29	40.47 ± 30.67	41.04 ± 31.36	80.57 ± 37.85	**4.61 ± 10.22**

硬件使用　　编。 所有实验均在一台配备 128 GB RA 的 32核 4.5 GHz 机器上进行并计时　　M.

基线方法。 我们通过与多种方法进行比较来评估在线方法的性能，包括现实世界中的充电流程、若干智能启发式方法以及一种基于强化学习的策略。下面简要介绍这些基线方法。

• **MaxCharge**：此方法模拟当前现实世界中的充电方式，即所有已连接电动汽车都以最快速度充电至 e^v_{max}。
• **ReqCharge**：与*MaxCharge*类似，但仅以尽可能快的速度为所有已连接电动汽车充电至 e^v_{req}。
• **最小松弛时间优先（LLF）**：采用第4.2节所用的相同启发式策略。如果可能，会充电超过 e^v_{req}，然后利用多余电能降低峰值功率需求。否则，采用涓流充电为电动汽车充电至 e^v_{req}。
• **最早截止时间优先（EDF）**：EDF 优先为离开时间最近的电动汽车充电，遵循最早截止时间优先调度方法 [24]。与 LLF 类似，如果预计建筑负荷较高，则可能超额充电，并在离开前放电。否则，采用涓流充电达到 e^v_{req}。

• **强化学习（RL）**：我们使用深度确定性策略梯度（DDPG）算法管理充电动作。状态包括时间、建筑负荷、充电状态、电动汽车荷电状态和预计离开时间。奖励函数综合了需求成本、能源成本和荷电状态偏差惩罚。DDPG 模型采用由两层多层感知机（MLP）构成的演员网络和评论家网络，每个网络包含 96 个神经元。动作掩码可防止未连接电动汽车时进行充电，确保在离开前达到 SoC 目标，并释放多余能量。策略指导将由 MILP 生成的最优动作融入

表 5. 所有月份的削峰（越低越好）

策略	DG-MCTS	dMCTS	RL	LLF	EDF	ReqCharge	MaxCharge
平均值 ±标准差	**-5.17 ± 13.67**	3.14 ± 11.80	18.16 ± 22.81	1.77 ± 9.02	2.77 ± 10.52	11.22 ± 6.15	94.02 ± 46.60

表6。消融研究 | 敏感性分析，基于2023年8月的10个测试情景（越低越好）

月份	DG-MCTS	MCTS/P	MCTS/H	MCTS/C	MCTS(ME)	MCTS(LE)	MCTS(BLF)
AUG	**8512.24 ± 81.42**	8690.19 ± 86.53	8705.76 ± 63.60	8521.80 ± 85.13	8601.91 ± 52.99	8625.59 ± 41.88	8670.58 ± 76.77

表7. 较大的离开时间窗口（3小时）下各策略的总成本（越低越好）

月份	DG-MCTS	LLF	EDF	ReqCharge	MaxCharge
AUG	**8632.47 ± 73.45**	8682.06 ± 49.67	8682.98 ± 50.74	8742.56 ± 55.19	9402.73 ± 144.43

训练集，以进行超参数。. 每个月份都使用 60 个模拟样本训练一个单独的模型 .

5.2 结果

我们使用四项指标评估所有方法：

(1) **总成本**：根据公式（1），计费周期内的电费、需量电费和 SoC缺失成本之和。(2) **SoC 缺额**：指离场车辆的荷电状态相对于所需荷电状态的能源缺额，每千瓦时将产生 20 美分的惩罚，比电网最高能源成本高 42%。这会抑制建筑依赖双向电动汽车来减少电网购电量。(3) **未达到所需 SoC 的车辆数**：SoC 缺额车辆的数量，可反映模型是否满足所有所需的 SoC 目标，或优先保障部分车辆而牺牲其他车辆。(4) **削峰**：指以下两种情况下需量电费的差值：(i) 建筑的功率使用量（不进行任何充电）；(ii) 根据相应策略为电动汽车充电。正值表示该策略通过控制充电动作降低了需量电费。

我们使用 2023 年五月至十二月的数据评估了在线方法。随机选取 10 个回合作为测试集，其余 100 个回合则用作根节点并行化中的探索样本。策略无法获取车辆的实际离开时间，因此依赖离开时间窗口。

表 2 比较了不同策略在八个月内的月度账单。领域知识引导的蒙特卡洛树搜索（DG-MCTS）在六个月中优于其他启发式方法，其成本效益高于现实世界的基线方法 MaxCharge 和 ReqCharge，以及 LLF 和 EDF 等智能启发式方法。

DG-MCTS 通过优化削峰有效降低了月度成本，同时漏掉的荷电状态值与启发式方法相当。每次决策平均耗时 23.75 秒，而 dMCTS 更快，为 15.38 秒。

基于强化学习的策略在八个月中的两个月优于 DG-MCTS，但仍有许多电动汽车荷电状态要求未得到满足（表 3）。表 4 显示，由于动作选择不确定性，强化学习未能满足平均 207 辆汽车的荷电状态目标。尽管我们的方法在某些情节中导致更多荷电状态未达标的情况，但在未达到规定荷电状态的电动汽车数量方面，其表现与智能启发式方法相近。

结果表明，我们的方法通过在不确定性下优化动作，优于启发式方法。智能启发式方法虽然优于 MaxCharge 和 ReqCharge，但无法预见未来事件，因此性能受限。即使采用 MaxCharge，如果部分汽车停留时间过短，也无法达到其 SoC 目标。

对不确定性的鲁棒性。 最后，我们通过以下方式进一步提高不确定性程度，并展示我们的方法优于所有启发式方法：**1)** 扩大潜在离开时间窗口，如表 7 所示；以及 **2)**应对每日电动汽车到达数量意外增加的情况，将每日汽车数量提高到约 25 辆，如表 8 所示。与其他策略相比，DG-MCTS 能更好地应对环境变化，每次决策耗时 48.97 秒，而 dMCTS 每次决策耗时

表8. 每日车辆到达量意外增加情况下各策略的总成本（越低越好）

月份	DG-MCTS	dMCTS	LLF	EDF	ReqCharge	MaxCharge
AUG	**8589.0 ± 74.19**	8650.75 ± 64.98	8691.76 ± 53.1	8696.32 ± 57.44	10345 ± 100	11395.53 ± 432.54

Ta表 9. 总成本与执行时间　　不同探索次数下的每项决策时间　　样本数时的每次决策时间（越低越好）

参数	DG-MCTS (10ES)	MCTS-5ES	蒙特卡洛树搜索-15ES	蒙特卡洛树搜索-20ES
总成本（美元）	8512.24 ± 79.98	8537.95 ± 96.19	8511.85 ± 77.3	8511.53.13 ± 75.65
Time	23.75	22.37	24.39	25.01

平均为 25.46 秒。这表明，在计算时间方面，dMCTS 的扩展能力比 DG-MCTS 更强。表8 还显示，在扩展规模的过程中，dMCTS 优于其他基线方法。

5.3 消融研究

我们通过消融实验评估所用方法中关键技术的影响。我们选取建筑峰值负荷最高的月份——2023年8月，并评估其对总电费的影响。测试的消融项包括：**1) MCTS/P**，该方法仅使用当前峰值需求替代预测峰值需求。**2) MCTS/H**，该方法移除用于缩小动作空间的启发式方法。**3) MCTS/C**，该方法使用分段线性电池曲线，而非简化线性模型。表6 展示了各项特征的影响。

峰值需求预测：我们考察**MCTS/P**，该方法移除了峰值需求预测，改用当前功率需求作为阈值，其性能大幅下降。这表明，对于规划时域内任意长度的预测，准确的峰值需求预测模型都至关重要。

用于动作剪枝的启发式方法：**MCTS/H** 方法不再使用启发式方法进行动作剪枝，因此性能有所下降。这表明 MCTS 的表现取决于生成的树的质量。移除动作剪枝会扩大动作空间，使得难以选择“良好”的动作。增加迭代次数可以改善解，但可能会增加计算时间。

分段线性充电曲线：采用分段线性充电曲线的**MCTS/C**可提供类似的解决方案，但会增加复杂性。大多数电动汽车制造商不会披露充电曲线，这使荷电状态曲线分析变得复杂。正如 Sundström 和 Binding [27] 所指出的，尽管分段线性曲线可能提高准确性，但缺少这些曲线并不会影响该方法的有效性。

5.4 敏感性分析

本节通过扰动 MCTS 探索样本，探讨改变电动汽车行为和建筑负荷预测准确度所产生的影响。我们评估了三种采用调整后探索样本的情景：（1）更多电动汽车（ME）；（2）更少电动汽车（FE）；（3）更大的建筑负荷波动（BLF）。我们在表 2中的 10 个未作改动的 8 月测试回合上测试策略的性能。表 6右半部分的结果评估了策略对不确定性增加的响应，以及这对总成本的影响。当 DG-MCTS 与包含更多电动汽车的探索样本结合时，记为 MCTS(ME)，探索样本中的汽车平均比测试回合多 25%。相比之下，使用较少电动汽车的样本时，记为 MCTS(FE)，探索样本中的汽车平均少 25%。尽管这两种情景的总电费都高于 DG-MCTS，但其总成本仍低于表现最佳的领域专用启发式方法 LLF。在 MCTS(BLF) 情景中，我们使用建筑负荷波动更大的探索样本测试 DG-MCTS，具体做法是向预测负荷添加固定扰动 ±10%。这考虑了建筑负荷预测模型可能存在的误差，并评估规划器应对此类变化的能力。如表 6 所示，总电费高于 DG-MCTS 和 LLF。随着预测误差增加，DG-MCTS 相较于最佳启发式方法的性能优势逐渐消失。这种影响在建筑负荷预测方面更为明显：准确度下降时，性能会迅速下滑。这可能是因为 MCTS 中的决策树高度依赖预测的建筑负荷来估算奖励。

5.5 探索样本的影响

我们加入表 9，以说明探索样本数量不同时性能如何变化。我们以 5 为步长，将探索样本数量从 5 逐步增加到 20。MCTS-5ES 使用 5 个样本，MCTS-15ES 使用 15 个样本，MCTS-20ES 使用 20 个样本，其中 ES 表示探索步数。其他实验中使用的 DG-MCTS 版本包含 10 个探索样本，记为 DG-MCTS (10ES)。表中结果显示，改变探索样本数量对整体性能影响极小，结果基本一致，偏差很小。这表明所选探索样本足以获得稳定的结果。

6 结论

本文提出了一种领域知识引导的蒙特卡洛树搜索（DG-MCTS）方法，通过优化充电控制来应对智能建筑中的车辆到建筑（V2B）挑战。在在线决策场景中，我们面临诸多挑战，例如电动汽车到达与离开的随机性，以及决策时间有限。为应对这些挑战，DG-MCTS 利用领域专用启发式方法引导动作选择并剪枝动作空间，使其能够在合理的计算时间内，在降低成本和优化需量电费方面优于传统启发式算法。虽然 DG-MCTS 表现更优，但在汽车和充电桩数量较多的场景中，去中心化蒙特卡洛树搜索（dMCTS）具有更高的扩展效率。我们使用来自电动汽车制造商日产北美及其硅谷实验室的数据，对这两种方法进行了评估；该实验采用了一个配备 15 个充电桩的智能建筑模拟 V2B 场景。结果表明，DG-MCTS 能够有效处理电动汽车离开时间的不确定性，同时实现最低月度总成本并满足荷电状态要求，展现出其在动态条件下管理在线电动汽车充电的能力。与此同时，dMCTS 提供了一种可扩展的替代方案，在高维多智能体场景中表现良好。

7 致谢

本材料基于日产先进技术中心硅谷分部资助的研究成果。本文所呈现的结果基于由美国国家科学基金会支持的Chameleon测试平台获得。本文所表达的任何观点、研究结果、结论或建议均代表作者的意见，并不一定反映日产的观点。

参考文献

[1] Moloud Abdar、Farhad Pourpanah、Sadiq Hussain、Dana Rezazadegan、Li Liu、Mohammad Ghavamzadeh、Paul W. Fieguth、Xiaochun Cao、Abbas Khosravi、U. Rajendra Acharya、Vladimir Makarenkov 和 Saeid Nahavandi。2020。《深度学习中的不确定性量化综述：技术、应用与挑战》。《信息融合》76（2020），243–297。[2] Heba M. Abdullah、Adel Gastli 和 Lazhar Ben-Brahim。2021。《基于强化学习的电动汽车充电管理系统——综述》。《IEEE Access》9（2021），41506–41531。[3] Aijun Bai、Feng Wu 和 Xiaoping Chen。2013。在蒙特卡洛树搜索中基于贝叶斯混合建模与推断的汤普森采样。载于《神经信息处理系统》。[4] Daniel Claes、Frans Oliehoek、Hendrik Baier 和 Karl Tuyls。2017。面向多机器人仓库调度的去中心化在线规划。载于《国际自主智能体与多智能体系统会议》（AAMAS）。ACM。[5] IBM ILOG Cplex。2009。V12. 1：《CPLEX 用户手册》。国际商业机器公司46、53（2009），157。[6] Leonardo de A. Bitencourt、Bruno S.M.C. Borba、Renan SilvaMaciel、Márcio Zamboti Fortes 和 Vitor Hugo Ferreira。2017。考虑动态定价的电动汽车最优充放电控制。2017 IEEE曼彻斯特电力技术会议（2017）。[7] Mahdi Ghafoori、Moatassem Abdallah 和 Serena Kim。2023。利用机器学习和车辆到建筑（V2B）系统实现商业建筑电力削峰。《应用能源》340（2023），121052。[8] Mykel J. Kochenderfer。2015。《不确定条件下的决策：理论与应用》。麻省理工学院出版社。doi:10.7551/mitpress/10187.001.0001[9] Levente Kocsis 和 Csaba Szepesvári。2006。基于赌博机的蒙特卡洛规划。载于《欧洲机器学习会议》。施普林格，282–293。[10] Levente Kocsis 和 Csaba Szepesvári。2006。基于赌博机的蒙特卡洛规划。载于《第17届欧洲机器学习会议》（德国柏林）（ECML'06）。施普林格出版社，柏林，海德堡，282–293。[11] Yanqing Kuang、Mengqi Hu、Rui Dai 和 Dong Yang。2017。电动汽车与建筑集成的协同决策模型。《能源学报》105（2017），2077–2082。[12] Jae Hyun Lee、Eunjung Lee 和 Jinho Kim。2020。动态定价方案中基于数据驱动方法和强化学习的电动汽车充放电算法。《能源》（2020）。[13] Karol Lina López、Christian Gagné 和 Marc-André Gardner。2019。利用深度学习进行需求侧管理，实现电动汽车智能充电。《IEEE智能电网汇刊》10（2019），2683–2691。[14] Simon M. M. Lucas、Spyridon Samothrakis 和 Diego Perez Liebana。2014。面向蒙特卡洛树搜索的快速进化适应。载于 EvoApplications。[15] Jianing Luo、Yanping Yuan、Mahmood Mastani Joybari 和 Xiaoling Cao。2024。开发一种带有车辆到建筑模式的光伏—建筑—电动汽车集成系统基于预测的调度控制策略。《可再生能源》224，C（2024）。

[16] Hoang Tien Nguyen 和 Dae-Hyun Choi。2023。具备车辆到电网／车辆到车辆能力的智能电动汽车充电站的分布鲁棒模型预测控制。《*IEEE*智能电网汇刊》14（2023），4621–4633。[17] Björn Nykvist 和 Måns Nilsson。2015。电动汽车电池组成本的快速下降。《自然·气候变化》5，4（2015），329–332。[18] Geoffrey Pettet、Ayan Mukhopadhyay、Mykel Kochenderfer、Yevgeniy Vorobeychik 和 Abhishek Dubey。2020。关于智能互联社区应急响应系统中的算法决策程序。*arXiv*预印本 *arXiv:2001.07362*（2020）。[19] Michael J Risbeck 和 James B Rawlings。2019。面向时变成本和峰值需量电费优化的经济模型预测控制。《*IEEE*自动控制汇刊》65，7（2019），2957–2968。[20] Tetsuya Sato、Takayuki Shiina 和 Ryunosuke Hamada。2022。基于随机规划的电动汽车公交车充电计划优化。*2022*年第*12*届国际先进应用信息学大会（*IIAI-AAI*）（2022），627–632。[21] Liam Schramm 和 Abdeslam Boularias。2024。通过状态占用正则化实现蒙特卡洛树搜索中的可证明高效长时域探索。载于《第*41*届国际机器学习大会论文集》（《机器学习研究论文集》，第*235*卷），Ruslan Salakhutdinov、Zico Kolter、Katherine Heller、Adrian Weller、Nuria Oliver、Jonathan Scarlett 和 Felix Berkenkamp（编）。PMLR，43828–43861。[22] Wenbo Shi 和 Vincent WS Wong。2011。不确定电价下的实时车辆到电网控制算法。载于*2011 IEEE*智能电网通信国际会议（*SmartGridComm*）。IEEE，261–266。[23] David Silver、Thomas Hubert、Julian Schrittwieser、Ioannis Antonoglou、Matthew Lai、Arthur Guez、Marc Lanctot、Laurent Sifre、Dharshan Kumaran、Thore Graepel 等。2018。通过自我对弈掌握国际象棋、将棋和围棋的通用强化学习算法。《科学》362，6419（2018），1140–1144。[24] John A. Stankovic、Krithi Ramamritham 和 Marco Spuri。1998。《实时系统的截止期限调度：*EDF*及相关算法》。Kluwer Academic Publishers，美国。[25] SJ Sultanuddin、R Vibin、A Rajesh Kumar、Nihar Ranjan Behera、M Jahir Pasha 和 KK Baseer。2023。电动汽车车队改进型强化学习智能充电策略的开发。《能源存储期刊》64（2023），106987。[26] Yongjun Sun、Shengwei Wang、Fu Xiao 和 Diance Gao。2013。商业建筑中利用不同冷能储存设施进行峰值负荷转移控制：综述。《能源转换与管理》71（2013），101–114。[27] Olle Sundström 和 Carl Binding。2010。电动汽车车队充电规划的优化方法。载于《控制、通信与电力工程国际会议论文集》。Citeseer，28–29。[28] Richard S Sutton 和 Andrew G Barto。2018。《强化学习：导论》。麻省理工学院出版社。[29] Kevin Tanguy、Maxime R Dubois、Karol Lina Lopez 和 Christian Gagné。2016。利用车辆到建筑进行协同充电的优化模型与经济评估。《可持续城市与社会》26（2016），496–506。[30] Hao Tu、Srdjan Lukic、Abhishek Dubey 和 Gabor Karsai。2020。COVID-19期间建筑用电负荷的基于长短期记忆网络的在线预测方法。载于《*PHM*学会年会》。[31] 美国能源部。2023。用于移动储能的双向充电与电动汽车。[32] Yunjian Xu、Feng Pan 和 Lang Tong。2016。电动汽车充电的动态调度：一种优先级规则。《*IEEE*自动控制汇刊》61，12（2016），4094–4099。[33] Qing Yin、Chunmiao Han、Ailin Li、Xiao Liu 和 Ying Liu。2024。基于人工神经网络的建筑能耗预测模型研究综述。《可持续性》16，17（2024），7805。[34] Xian Zhang、Ka Wing Chan、Hairong Li、Huaizhi Wang、Jing Qiu 和 Guibin Wang。2020。采用新型排队模型的电动汽车充电负荷深度学习概率预测。《*IEEE*控制论汇刊》51（2020），3157–3170。

A 离线方法

对于任意给定的情节（即电动汽车到达与离开的一条轨迹，以及单个计费周期内的建筑负荷），我们可以使用离线优化程序求解车到建筑问题，并得到该周期的精确需求成本。因此，我们建立一个混合整数线性规划（MILP），以计算任意情节的最优需求成本。

A.1 精确解

混合整数线性规划的目标是最小化方程1中的总奖励多目标加权和，同时满足车到建筑问题的约束条件和要求。这些要求包括确保每辆电动汽车在停留期间分配给同一充电器，并使操作保持在充电器的限制范围内。我们使用 CPLEX [5] 求解该混合整数线性规划。

决策变量：我们使用第3.1节中定义的同一组变量。此外，z_v 用于表示汽车所需荷电状态与其离开时荷电状态之间的差值。

约束条件：为了匹配蒙特卡洛树搜索的车辆与充电桩分配策略（该策略遵循先到先服务的分配原则），我们预先计算车辆与充电桩的分配。当$h_{v,\ k,\ t} = 1$时，表示在时间 t将车辆 v 分配给充电器 k；车辆在停留期间始终保持该分配。

我们将分配变量与充电变量耦合如下。

$$\forall v \in \mathcal{V}, \forall t \in \mathcal{T} : c_t^v \leq \sum_k q_k^{max} \cdot \eta \cdot h_{v,k,t} \tag{3}$$

$$\forall v \in \mathcal{V}, \forall t \in \mathcal{T} : c_t^v \geq \sum_k q_k^{min} \cdot \eta \cdot h_{v,k,t} \tag{4}$$

我们跟踪　　电动汽车离开充电站前的荷电状态，并跟踪其与所需荷电状态之间是否存在差距。

$$\forall v \in \mathcal{V} : z_v = \left| e_{req}^v - e_{t_{end}^v}^v \right| \tag{5}$$

此外，对于电动汽车的 i^{th} 时段 t_i，我们可以使用 $e_{t_{n+1}}^v$求出剩余电量：

$$\forall v \in \mathcal{V}, t \in \mathcal{T}, \text{ if } i = 0 : \quad e_{t_i}^v = e_0^v + c_{t_i}^v, \tag{6}$$

$$\text{otherwise: } e_{t_i}^v = e_{t_{i-1}}^v + c_{t_i}^v \tag{7}$$

其中 e_0^v 是电动汽车到达时的初始电量。

需求成本可按如下方式计算（请记住，Ω 是聚合时段的数量）：

$$\forall j \in t_j^\tau, j \in \{0, 1, \cdots, \Omega\} : \pi_j = \sum_{v \in \mathcal{V}} \sum_{\forall i \in \mathcal{T}^\tau} c_{t_j}^v + b_j^p \tag{8}$$

$$P_{max} \geq \pi_j \tag{9}$$

我们需要确保车辆在停留期间始终使用同一个充电器，并通过在整个停留期间保持车辆与充电器的连续分配来实现：

$$\forall v \in \mathcal{V}, \forall k \in \mathcal{K}, \forall t \in \mathcal{T} \setminus t_{end} : a_{v,k,t} - a_{v,k,t+1} = 0 \tag{10}$$

目标 在给定分时电价的情况下，我们希望将能源成本和需求成本降至最低，同时缩小离开时所需的荷电状态与实际离开时荷电状态之间的差距。

$$\min \sum_{t \in \mathcal{T}} w_t \cdot \sum_{v \in \mathcal{V}} \left(c_t^v + b_t^e \right) + P_{max} \cdot w^d + w^s \cdot \sum_{v \in \mathcal{V}} z_v \tag{11}$$

表10. 强化学习（深度确定性策略梯度）超参数及选定值。

参数	描述	范围
演员网络	每层的单元数	[96, 96]
评论家网络	每层的单元数	[96, 96]
Γ	未来奖励的折扣因子	1
演员与评论家学习率	用于更新演员网络的学习率和评论家网络	10^{-5}, 10^{-3}
缓冲区大小	用于获取转移的批次大小来自经验回放缓冲区	64
批次大小	经验回放缓冲区大小	10^{6}
动作噪声	动作探索过程中添加的噪声过程	normal(0,0.2)
策略引导率 引入	策略引导的概率ance	0.5 或 0.7
λ_S、 λ_B、 λ_D	荷电状态要求的惩罚系数、电费成本和需量电费	1、1、3
随机种子	演员和评论家的随机种子网络初始化	0-5
训练步数, 更新步数	训练步数和每次更新的步数目标网络的	5, 5

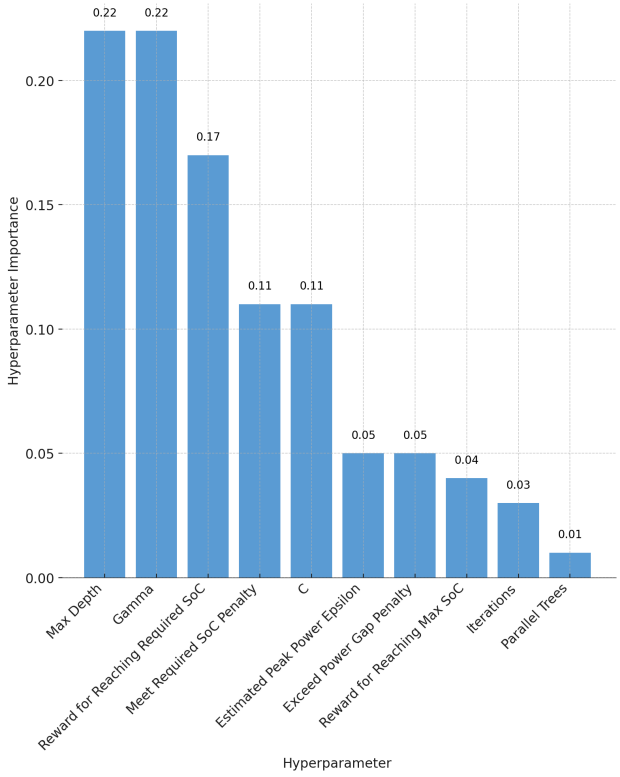


图2. 超参数探索中各因素的重要性