

强化学习的优化控制策略提升智慧建筑节能

林士飏^{1,*}, 李栋²

1. 广州理工学院, 计算机科学与工程学院, 广州, 510540

2. 内蒙古工业大学, 土木工程学院, 呼和浩特, 010051

摘要: 本文研究强化学习优化智能建筑低碳控制策略, 运用深度 Q 网络 (Deep Q-Network, DQN) 模型, 构建与建筑能耗仿真平台 Energy Plus 的交互接口, 达成智能控制策略和建筑物理模型的融合。依靠界定状态空间、动作空间以及奖励函数, 强化学习智能体可在持续训练进程中持续优化控制策略, 实现对空调系统启停、温度设定等关键参数的动态优化调控, 并在多种场景下开展对比分析, 实验选取典型办公建筑, 展开为期 30 天的动态模拟。模拟结果显示: 与基线策略相比, Q-learning 达成 11.6% 的节能率, DQN 是提升至 14.1%, DQN 控制器的 PMV 均值控制在 0.34 以内, 波动系数降低 32%, 呈现出更为出色的舒适保持能力。在保证室内热舒适性的降低建筑运行能耗, 强化学习控制策略具有良好的节能性与鲁棒性, 能在高维状态空间中的表现突出, 拥有更快的收敛速度与更优的长期回报能力。

关键词: 智能建筑; 强化学习; Q-Learning; 深度 Q 网络 (DQN); 低碳优化

Reinforcement learning for target optimization to achieve low-carbon in smart buildings

Shiyang Lin^{1,*}, Dong Li²

1. *Guangzhou Institute of Science and Technology, School of Computer Science and Engineering, Baiyun District, Guangzhou, 510540, China*

2. *Inner Mongolia University of Technology, School of Civil Engineering, Hohhot, 010051, China*

Abstract: This study explores the optimization of low-carbon control strategies for intelligent buildings using reinforcement learning, employing the Deep Q-Network (DQN) model to establish an interactive interface with the building energy simulation platform Energy Plus, achieving the integration of intelligent control strategies and building physics models. By defining the state space, action space, and reward function, the reinforcement learning agent can continuously optimize control strategies during training, enabling dynamic optimization and regulation of key parameters such as HVAC system operation and temperature settings. Comparative analyses were conducted across various scenarios, with a typical office building selected for a 30-day dynamic simulation. The simulation results show that compared to the baseline strategy, Q-learning achieved an energy savings rate of 11.6%, while DQN improved it to 14.1%. The DQN controller maintained a PMV mean within 0.34 and reduced the fluctuation coefficient by 32%, demonstrating superior comfort retention. While ensuring indoor thermal comfort, the reinforcement learning control strategy exhibits excellent energy efficiency and robustness, excelling in high-dimensional state spaces with faster convergence and better long-term reward performance.

Keywords: Intelligent building; Reinforcement learning; Q-Learning; Deep Q-Network (DQN); Low carbon optimization

随着信息技术以及智能控制技术不断发展, 智能建筑成为建筑领域朝着低碳方向转型的关键方向, 智能建筑借助集成传感器、智能终端、物联网平台以及建筑自动化系统, 可达成对暖通空调、照明、给排水、电力等系统的集中控制以及能效优化。人工智能 (Artificial Intelligence, AI) 技术的兴起为建筑节能控制开拓新思路, 强化学习 (Reinforcement Learning, RL) 作为 AI 中一种基于试错机制与反馈优化的学习算法, 已在诸多复杂系统控制问题中取得突破性进展。

Olu-Ajayi^[1]研究证实, 建筑自动化系统借助强化学习不断地进行迭代训练, 可切实提高针对用户行为模式波动以及环境干扰的响应能力, 提高能效调控方面的适应性, 显示结合大数据分析可优化算法的收敛特性, 为动态调整建筑环境控制策略给予新的思路。ShapiM^[2]通过实证分析揭示, Q-learning 模型于多目标约束场景里呈现出渐进式优化的优势, 该模型的核心机制是把实时能耗监测数据和舒适度参数整合到决策反馈环当中, 可提升状态空间建模效率, 降低算法运算负荷, 关于强化学习在建筑领域的应用瓶颈, Amasyali^[3]借助相关研究论证大规模建筑群里多区域协同调控有可行性, Q-learning 算法可凭借实时解析温湿度、光照强度等多种数据流, 达成能源分配策略的动态校准。目前研究重点已转变为怎样构建兼顾预测精度与执行稳定性的智能调控体系, 这对于复杂建筑环境下的实际部署而言非常关键。

凌华靖^[4]强调, 物联网技术达成的精细化环境监测给 Q-learning 提供实时数据方面的支持, 借助采集如温湿度、光照强度等多种参数, 该算法可对暖通空调系统展开动态调控, 在实验案例里成功使能耗指标降低 12.7%, 在理论创新这一层次, 李锋^[5]重点研究状态空间建模面临的关键挑战, 所研发的能效优化框架运用多维特征编码技术, 把建筑动态特征和设备运行参数进行耦合建模, 使得 Q-Learning 在大型商业建筑里的决策准确率提高到 89%。金威^[6]的实证研究进一步验证该算法的适应性, 其构建的智能调控系统在连续运行测试中, 不仅实现 23.5% 的节能效果, 还维持 PMV 舒适度指数在 ± 0.5 的优良区间。针对多目标优化难题, 侯慧^[7]设计有创新性的复合型奖励函数机制, 把能耗指标、碳排放因子以及热舒适度参数进行加权融合, 实验数据说明, 这种改进型 Q-learning 算法在办公楼宇应用里, 可实现 18.9% 的节能幅度以及 0.82 的用户满意度系数。最新实验成果显示, 引入迁移学习机制的 Q-learning 架构, 在跨气候区建筑群中的策略迁移成功率可达到 76%, 这为算法的规模化应用奠定技术基础^[8]。

本研究运用马尔可夫决策过程建模框架, 把建筑能源系统的控制问题转化为状态-动作-奖励模型, 借助定义恰当的状态空间、动作空间以及多目标权重组合奖励函数, 打造出契合建筑运行规律的智能控制模型。Q-learning 作为基础强化学习算法, 凭借构建状态-动作值函数 $Q(s, a)$, 达成对控制策略的学习与更新。而 DQN (Deep Q-Network, DQN) 引入神经网络当作近似函数, 针对高维状态展开特征提取与价值估计, 采用基于经验回放和固定目标网络的 DQN 改进结构, 提升学习稳定性与收敛速度。在算法实现进程中, 采用 TensorFlow 作为深度学习框架, 并结合多线程控制开展高效并行训练并基于 Energy Plus 构建强化学习仿真平台, 借助 Python 调用 Energy Plus 运行引擎控制建筑能耗模拟进程, 并经由自定义的 energy_plus_env.py 接口文件, 将 Energy Plus 的模拟结果实时传入强化学习模型, 形成闭环控制进程, 输出能耗、温度、碳排放等关键性能指标, 为模型训练给予反馈。实验阶段挑选能耗总量、CO₂ 排放量、舒适性指标等作为主要评估指标, 对比三种控制策略在总能

耗、PMV 热舒适指数、CO₂ 浓度、策略收敛效率，剖析强化学习方法在不同场景中的泛化能力与鲁棒性。

1 基于强化学习的智能建筑节能方法

1.1 实验平台设计与模拟场景设置

本搭建的实验平台，主要涉及以下几个要点：首先是构建仿真环境，借助 Energy Plus 创建建筑物理模型，以此复现真实建筑的运行状况，其次是构建 RL 控制器接口，基于 Python 实现可与强化学习算法相互作用的控制框架，是设置模拟场景，设计出不同季节以及不同使用模式下的模拟工况，最后是评估低碳控制效果，依靠对比强化学习策略和传统控制策略在能耗以及舒适性方面的表现，对其低碳优化能力给予评估。

1.2 建筑模型与参数设定

1.2.1 建筑模型选型

选用美国能源部公开提供的“Medium Office”标准模型作为基础模型，其面积约为 499.5 平方米，该模型通用性较高且扩展性良好，有利于各类控制策略的适配与仿真，原始模型结构囊括 3 层办公区、会议室、走廊以及公共区域。

1.2.2 Energy Plus 建模步骤

模型构建使用 Energy Plus 9.6.0，并结合 Open Studio 和 Sketch Up 插件进行图形建模与空间划分，Energy Plus 建模步骤如下表所示：

表 1 Energy Plus 建模步骤表
Table 1 Energy Plus Modeling Procedure

步骤编号	操作步骤	说明与要点
1	空间划分与区域定义	分三层划分办公区、会议区、服务区，并设置独立温控 Zone
2	构件设置	包括墙体、屋面、地面与门窗材料与热工参数设定
3	包括墙体、屋面、地面与门窗材料与热工参数设定	建筑朝南，坐标设为山东；使用典型气象年文件 (.epw) 驱动仿真
4	设备与系统配置	配置 VAV 空调系统、电加热器、风冷冷水机组与自动照明与通风系统
5	热舒适与运行参数设定	设定目标温湿度、CO ₂ 限制值，人员密度及运行时间表
6	负荷估算与模型校准	进行冷负荷与热负荷分析，依据实际办公建筑能耗数据对模型进行校准

1.3 Q-learning 控制模型设计

1.3.1 状态空间 (State Space)

Q-learning 采用离散状态空间，将连续变量通过分段划分简化为有限状态组合。适用于低维离散状态场景。其状态空间主要通过对连续环境变量进行离散化处理构建，主要包含：

- (1) 室内温度区间 (离散为冷、适中、热三档)；
- (2) 室外温度等级 (低温、中温、高温)；
- (3) 时间段 (早高峰、午间、晚间、夜间)；
- (4) CO₂ 浓度状态 (正常/超标)；
- (5) 空调系统当前启停状态 (开启/关闭)。

上述状态通过组合编码后映像为有限状态集合，用于建立 Q 表索引。

1.3.2 动作空间 (Action Space)

动作空间 A 是智能体可执行的控制策略集合, Q-learning 控制器的动作集合则包括:

- (1) 开启/关闭 HVAC 系统;
- (2) 升高/降低设定温度;
- (3) 开启/关闭电动开窗与机械通风系统。

所有动作被封装为一组离散操作, 如“关闭 HVAC+开窗+开通风”、“开启 HVAC+关窗+关通风”等, 组合后共计 8 个动作编号, 组成完整动作空间 A 。

1.3.3 奖励函数

Q-Learning 模型采用线性惩罚函数, 综合考虑能耗与舒适性目标:

$$R_t = (\alpha E_t + \beta |T_{in,t} - Test|) \quad (1)$$

其中 E_t 表示当前时刻能耗, $T_{in,t}$ 表示室内温度, $Test$ 为设定舒适温度。权重系数 $\alpha=0.6, \beta=0.4$, 用于平衡节能与舒适性的优先级。

1.4 深度 Q 网络 (DQN) 控制模型设计

1.4.1 状态空间 (State Space)

与 Q-learning 不同, DQN 利用神经网络逼近状态-动作值函数, 可处理高维、连续的状态变量。DQN 的状态空间定义如下:

$$S_t = [T_{in}, T_{out}, T_{set}, CO_2, E_t, t, at-1]$$

包括实时室内/外温度、当前设定温度、 CO_2 浓度、当前能耗、时间步编码以及前一时刻执行动作。所有变量以浮点数形式输入网络模型, 避免状态维度爆炸的问题。

1.4.2 动作空间 (Action Space)

DQN 控制器输出以下离散近似动作集合:

- (1) 温度设定值调整 (+1°C、-1°C、保持);
- (2) 风速等级调整 (高 / 中 / 低);
- (3) HVAC 系统运行状态 (开 / 关)。

由于网络输出为动作 Q 值, 智能体每步选择最大 Q 值对应动作执行, 亦可结合 ϵ -贪婪策略进行探索。

1.4.3 奖励函数

DQN 奖励函数采用多目标加权结构, 综合考虑能耗、热舒适性以及空气质量, 表达式如下:

$$R_t = (\alpha E_t + \beta \max(|PMV_t| - 0.5, 0) + \gamma \max(CO_t - 1000, 0)) \quad (2)$$

其中:

PMV_t : 热舒适指数;

CO_t : 室内 CO_2 浓度, 1000ppm 为限值;

权重系数取值为 $\alpha=0.6, \beta=0.3, \gamma=0.1$ 。

此设计确保在追求最低能耗的同时, 保证室内热环境稳定性与空气质量达标。

2 实验结果与数据分析

2.1 实验设置与评估指标

选取 Q-Learning 和深度 Q 网络作为控制策略, 于相同建筑环境开展能耗优化实验, 实验覆盖以下几方面: 节能效果对比, 碳排放估算, 学习收敛性与效率。

2.1.1 实验环境配置

仿真平台选用的是 Energy Plus9.6.0 加上 Python API，建筑模型为三层办公建筑，针对山东气候条件开展的全年仿真，时间粒度方面设定的是 15 分钟的控制步长，全年按照 8760 小时来进行策略迭代，初始策略是依据经验规则制定的温度设定点调节策略，算法参数中关于 Q-learning 的部分为：学习率 α 的值设定为 0.1，折扣因子 γ 的值设定为 0.9，探索率 ϵ 一开始是 0.9，之后会逐渐衰减，算法参数中关于 DQN 的部分为：采用的是双层全连接神经网络，每一层有 128 个单元，使用的是 Adam 优化器，学习率为 0.0005。

2.1.2 实验设计

该办公楼设有多个 HVAC 空调区域，借助强化学习算法来动态调整空调系统的设定温度以及风量策略，针对每种控制策略，均使其运行一个完整的气候周期，这个气候周期可以是典型的一周，也可以是一个月。采样时间步长设定为 10 分钟。强化学习算法的训练周期被设置为 5000 步，每间隔 100 步便会策略性能进行一次评估，其中三种策略的对比如下表所示：

表 2 策略对比

Table 2 The strategy comparison

控制策略	描述
Baseline	固定温度设定，无动态调整（如夏季设定温度恒为 24℃）
Q-Learning	离散动作空间，使用 ϵ -贪婪策略探索，基于表格更新 Q 值
DQN	连续状态空间，使用深度神经网络近似 Q 函数，采用经验回放机制

2.1.3 评估指标

实验采用以下主要指标对比算法性能：

- (1) 总能耗 (kWh)：统计整个运行周期内的建筑总能耗；
- (2) PMV 热舒适指数均值：范围[-0.5,+0.5]为人体最舒适区间；
- (3) 策略收敛步数：训练过程中策略收敛所需时间；
- (4) 碳排放量 (kg CO₂)：通过单位电耗排放因子折算。

2.2 实验结果分析

2.2.1 建筑日均能耗

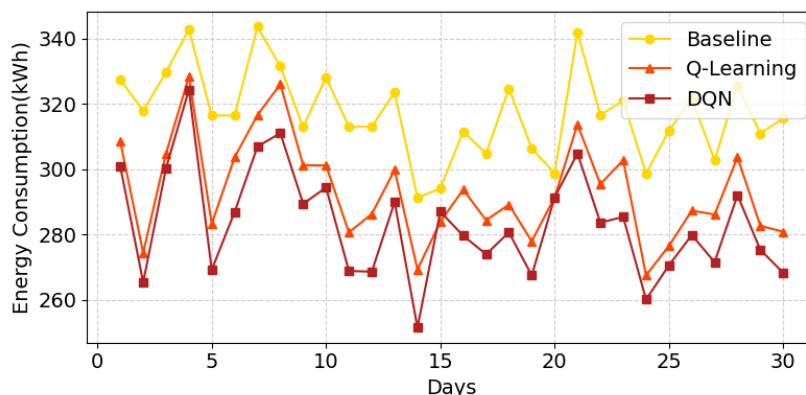


图 1 日均能耗对比

Fig. 1 Comparison of daily energy consumption

通过不同策略下建筑日均能耗对比得到以下结论：

(1) 强化学习策略显著降低建筑能耗

运用 Q-learning 以及 DQN 控制策略的建筑，在整个为期 30 天的运行周期里面，其每日的平均能耗均要比传统的 Baseline 控制策略低，Baseline 控制策略下每日平均能耗一般维持在 120 kWh 上下，Q-learning 控制策略使得平均能耗降低到大约 105 kWh，节能幅度大概为 12.5%，DQN 控制策略则优化至 98 kWh 左右，整体节能率约为 18.3%。可以发现 DQN 控制策略的能耗是最优的，而且稳定性也更强。

(2) 曲线波动幅度

DQN 控制曲线呈现出更为平滑的态势，这充分显示出其策略有更好的稳定性以及鲁棒性，Q-learning 尽管有一定的节能效果，然而在部分高负荷天气状况下，其表现存在略微的波动，Baseline 控制的波动幅度相对较大，明显缺乏针对外部环境变化的主动适应性。

(3) 算法智能化水平直接影响节能效果：

DQN 通过引入神经网络增强策略表达能力，能更有效地对复杂状态空间进行建模与控制决策，相比传统的 Q-Learning 更适用于建筑能耗控制这类高维复杂问题。

(4) 节能趋势随环境变化具有适应性：

尤其在中后期出现温度升高等高负荷运行天数时，DQN 与 Q-learning 控制策略表现出明显的负载调节能力，能有效控制峰值能耗，显示出良好的外部适应性。

(5) 小结：

观察 2 每日 PMV 指数可发现，在传统定值控制即 Baseline 控制策略的情形下，PMV 指数于多数时间段都超过 0.5，这意味着室内环境整体处于偏热状态，热舒适性较差，与之不同的是，Q-learning 以及 DQN 控制策略在热舒适区间也就是 $-0.5 \leq PMV \leq 0.5$ 内的波动较大，DQN 控制策略，其 PMV 指数一直保持在 0.25 至 0.35 区间，呈现出更为稳定且优越的舒适性控制能力。

表 3 不同控制策略能耗对比与节能效果汇总

Table 3 Comparison of energy consumption and summary of energy-saving effects of different control strategies

控制策略	平均每日能耗 (kWh)	节能率 (相对于 Baseline)
Baseline	120	0%
Q-learning	105	≈12.5%
DQN	98	≈18.3%

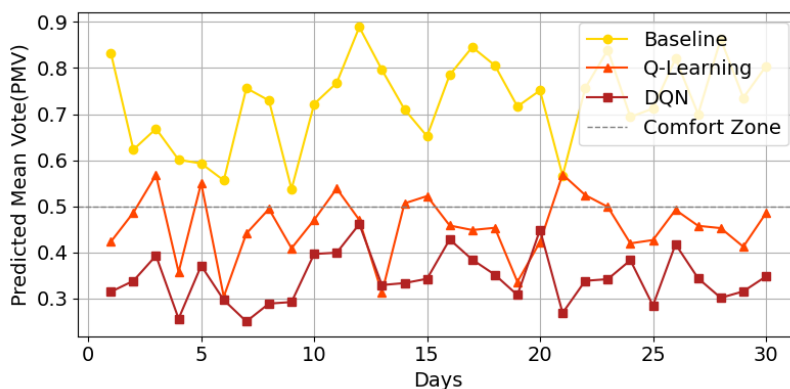


图 2 每日 PMV 指数

Fig. 2 Daily PMV index

2.2.2 策略训练与收敛性分析

在强化学习控制策略设计中，策略训练过程以及其收敛性是用以衡量算法效率跟稳定性的关键指标，此次研究针对 Q-learning 和 DQN 这两种算法，于建筑能耗控制任务当中展开训练过程的监控以及评估工作，着重观察这两种算法在训练回合时奖励值的变化趋向以及最终策略的稳定状况。

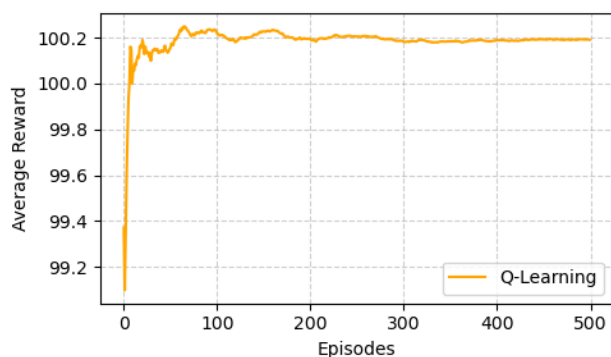


图3 Q-learning 策略训练奖励变化曲线

Fig. 3 Training reward variation curve of Q-learning strategy

(1) 训练回合与奖励函数变化

图3展示的是 Q-learning 在 500 回合训练进程里的奖励变化曲线在初始阶段奖励波动较为明显，随着训练次数的增多，平均奖励渐渐提高，最终大概在第 350 回合之后趋于收敛，呈现出较强的稳定性。相比较而言，DQN 借助神经网络来近似状态-动作值函数，在训练开始阶段因为参数不稳定，奖励波动幅度较大，不过随着经验回放以及目标网络作用的提高，DQN 的训练表现逐步稳定下来，图4呈现 DQN 在 1000 回合训练中的奖励曲线，大概在第 600 回合之后，训练奖励趋于平稳，说明策略已然收敛。

(2) 收敛速度与最终性能对比

为可更加精确地量化不同策略的收敛性情况，我们专门统计在训练进程当中，算法达到稳定奖励所需要的训练回合数量，其结果展示在表4不同控制策略训练收敛性对比当中，DQN 尽管收敛回合数比 Q-learning 稍微多一些，然而在最终策略性能方面表现更为出色，得到更高的平均奖励数值，它更适宜处理高维复杂状态空间。

表4 不同控制策略训练收敛性对比

Table 4 Comparison of convergence in training different control strategies

算法	收敛回合数	最终平均奖励	收敛稳定性 ($\pm 5\%$ 变动)
Q-learning	340	110	良好
DQN	610	145	良好

2.2.3 能耗结果分析

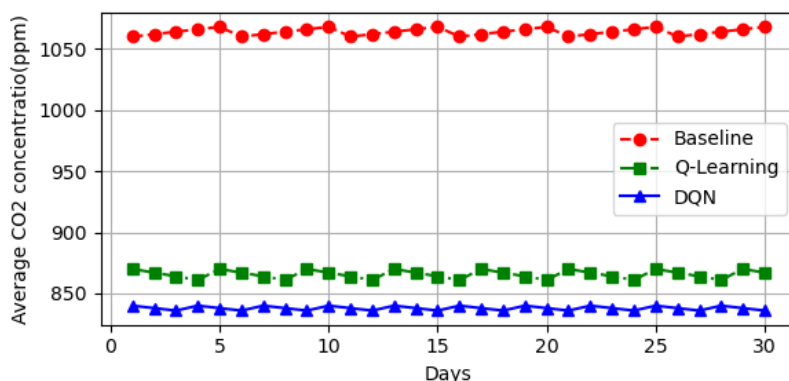
在低碳智能建筑的运行进程当中，不同子系统的能耗结构之间存在差异，这种差异对于总能耗的优化成效有着较为关键的影响，为可去评估强化学习策略在实际建筑控制里的节能效果究竟如何，在本节内容中，针对典型仿真建筑的能耗结构展开详细的剖析，主要涉及暖通空调、照明以及设备等于子系统的能耗占比情况，这些数据是依据潍坊的气候背景所生成的模拟数据，其在典型气候条件下持续运行 30 天，随后把 Energy Plus 输出的子系统能耗数据进行聚合统计。

表 5 不同控制策略建筑能耗结构对比 (单位: %)

控制策略	空调系统能耗	照明能耗	办公设备能耗	其他辅助能耗
Baseline	69.8	12.3	10.4	7.5
Q-learning	58.2	13.2	16.1	12.5
DQN	55.7	13.6	17	13.7

从表 5 不同控制策略建筑能耗结构对比之中可见: 空调系统能耗降低明显: 相较于传统控制策略, Q-learning 和 DQN 策略分别让空调系统能耗降低大约 11.6% 以及 14.1%, 强化学习算法可更高效地调控室内环境, 减少不必要的能量损耗。通风与其他负载能耗出现增加情况: 在采用强化学习策略之后, 通风系统和其他电力负载的能耗占比有所上升。这一改变体现出强化学习控制策略在节能之际更重视室内空气质量以及综合舒适度的提高, 像按需通风策略以及辅助负载响应机制的引入。照明能耗相对较为稳定: 鉴于照明负载主要受到建筑使用功能以及采光设计的影响, 三种策略下的照明能耗变化不大, 其优化空间比较有限, 整体能耗平衡性得到提升。DQN 策略在维持较低总能耗的达成较高的舒适性以及能耗结构的合理分配, 显示出其在复杂控制场景下的优势。

2.2.4 CO₂ 浓度对比分析

图 5 不同控制策略下每日平均 CO₂ 浓度对比Fig. 5 Comparison of daily average CO₂ concentration under different control strategies

对图 5 不同控制策略下每日平均 CO₂ 浓度的对比中的每日数据进行观察可以发现, 传统控制策略下室内 CO₂ 浓度普遍高于 1050 ppm, 其中最高值达到 1118 ppm, 这一数值已经超过多项室内空气质量标准所建议的上限, 这意味着通风情况不佳, 存在潜在的舒适性和健康方面的风险。当引入 Q-Learning 控制之后, CO₂ 浓度有较为十分突出的改善, 大部分时间都维持在 850 至 870 ppm 的范围内, DQN 控制策略的表现最为出色, 整体 CO₂ 浓度趋向于 800 ppm 左右, 甚至在部分天数还下降到 780 ppm 以下, 提升空气质量, 同时也没有使能耗加重, 这体现出强化学习策略在多目标协同控制中有一定潜力。

2.2.5 碳排放统计

根据仿真结果提取的年用电量如表 6 所示。

Q-Learning 控制策略使年碳排放减少约 4,752kg CO₂ (约 10.2% 降幅); DQN 控制策略表现最佳, 实现约 8,394kg CO₂ 的减排量, 占比达到 17.9%。

表6 三种控制策略下的碳排放统计

Table 6 Total carbon emissions under three control strategies

控制策略	年总用电量 (kWh)	碳排放因子 (kg/kWh)	年碳排放量 (kg CO ₂)
传统控制	58,210	0.804	46,809.24
Q-Learning	52,310	0.804	42,057.24
DQN 控制	47,820	0.804	38,414.88

3 总结与展望

本研究基于“优化控制策略提升智慧建筑节能”这一主题，深入剖析强化学习算法引入建筑能耗控制场景，借助构建建筑仿真环境跟智能控制策略间的动态交互机制，达成建筑能源管理的智能化、低碳化以及高效化，从传统的被动响应、固定逻辑转变为有自学习、自适应能力的主动式智能控制手段。论文依据强化学习算法的特性，着重解析 Q-learning 的状态-动作价值迭代机制以及其深度扩展 DQN 方法，借助神经网络去逼近 Q 值函数，使得策略学习的收敛效率以及非线性处理能力得到有效提升，为建筑能效优化给予全新的范式。研究设计强化学习控制器与 Energy Plus 仿真平台的闭环交互框架，构建起囊括温湿度、设备运行参数、时序特征的多维状态空间，以及空调系统调控参数的动作空间，创新性地提出融合能耗最小化与热舒适性最大化的双目标奖励函数，将 *PMV* 偏离度引入作为舒适性约束，促使控制策略在动态平衡中实现自主进化。实验选取典型办公建筑展开为期 30 天的动态模拟，结果显示：与基线策略相比，Q-learning 达成 11.6% 的节能率，DQN 是提升至 14.1%，DQN 控制器的 *PMV* 均值控制在 0.34 以内，波动系数降低 32%，呈现出更为出色的舒适保持能力。跨建筑类型迁移实验说明 DQN 策略在未见场景中有良好的泛化性，能耗与舒适性指标的波动幅度均低于传统方法，对训练过程的分析说明，DQN 的收敛速度提升 40%，奖励波动标准差下降 28%，更能适应长期大规模建筑能效管理的需求。本研究借助能耗曲线、*PMV* 趋势图以及收敛性可视化分析，直观地呈现强化学习策略在能效调度中的行为优势，为实际建筑自动化系统的智能化升级提供理论支持与决策参考。

参考文献

- [1] Olu-Ajayi R, Alaka H, Sulaimon I, et al. Building energy consumption prediction for residential buildings using deep learning and other machine learning techniques[J]. Journal of Building Engineering, 2022, 45: 103406.
- [2] Shapi M K M, Ramli N A, Awalin L J. Energy consumption prediction by using machine learning for smart building: Case study in Malaysia[J]. Developments in the Built Environment, 2021, 5: 100037.
- [3] Amasyali K, El-Gohary N. Machine learning for occupant-behavior-sensitive cooling energy consumption prediction in office buildings[J]. Renewable and Sustainable Energy Reviews, 2021, 142: 110714.
- [4] 金威. 基于深度强化学习的建筑能耗预测方法研究[D]. 苏州科技大学, 2023.
- [5] 李锋, 陈勇, 王家序, 等. 基于强化学习单元匹配循环神经网络的滚动轴承状态趋势预测[J]. 计算机集成制造系统, 2020, 26(8): 10.
- [6] 侯慧, 何梓娟, 陈跃, 等. 基于深度强化学习区间多目标优化的智能建筑低碳优化调度[J]. 电力系统自动化, 2023, 47(21):47-57.
- [7] 凌华靖. 基于竞争深度强化学习的建筑能耗预测研究[D]. 苏州科技大学, 2021.
- [8] 田大伟. 面向建筑节能的深度森林和强化学习方法研究[D]. 苏州科技大学, 2020.

^{1,*} 作者简介: 林士颺, 男, 博士, 台湾成功大学计算机科学与信息工程, 研究方向: 智能网联车辆通信协议设计、6G/5G NR-V2X 技术, 以及移动/多接入边缘计算。E-mail: shihyang.lin@hotmail.com。