

文章编号: 1671-6612 (2025) 05-623-09

基于强化学习的 空调系统水泵在线自适应变频控制方法

张雯琦¹ 余 勇¹ 袁中原¹ 高 波²

(1. 西南交通大学机械工程学院 成都 610031;

2. 四川省建筑科学研究院有限公司 成都 610081)

【摘 要】 空调系统节能减碳十分重要, 系统运行控制是关键。针对中央空调系统水泵变频控制中传统定压差、定温差及经验式 PID 方法响应慢、能耗高且难以协调系统耦合的问题, 提出了一种基于经典强化学习算法 Q-learning 的在线自适应变频控制策略。该方法首先利用历史运行数据构建空调系统主要组成部分的数据驱动仿真模型, 并在该模型集上进行强化学习预训练以获得初始控制策略; 随后将预训练模型部署到实际系统中, 通过每分钟在线数据采集和整点控制机制, 不断更新 Q 表, 实现对水泵频率的自适应调节。该方法经过一个月的实地部署, 在真实空调系统中进行水泵运行频率的自适应调节。实验结果表明在 2025 年 5 月 20 日至 6 月 20 日的在线部署中, 相较于 35 Hz 固定频率基准策略, 系统在满足建筑使用人员用冷需求的前提下实现节能 4.59% 的优化效果。此外, 讨论了传感器误差对于本方法的影响以及修正策略。该方法能够一定程度上突破传统局部最优控制的局限, 实现跨设备、跨工况的全局能效优化, 为中央空调系统节能控制提供了新的技术路径。

【关键词】 强化学习; Q-learning; 在线自适应; 水泵变频; 节能控制

中图分类号 TU831.3 文献标志码 A

Reinforcement Learning-Based Online Variable Frequency Control Method for HVAC System Pumps

Zhang Wenqi¹ Yu Yong¹ Yuan Zhongyuan¹ Gao Bo²

(1.School of Mechanical Engineering, Southwest Jiaotong University, Chengdu, 610031;

2.Sichuan Institute of Building Research, Chengdu, 610081)

【Abstract】 Energy conservation and carbon-emission reduction in HVAC systems hinge critically on operational control. For variable-frequency pump control in central air-conditioning systems, conventional strategies-constant differential pressure, constant temperature difference, and empirical PID control-exhibit slow response, high energy consumption, and poor coordination of system couplings. This study proposes an online adaptive variable-frequency control strategy based on the classical reinforcement-learning algorithm Q-learning. Historical operating data are first used to construct data-driven simulation models of the pumps and terminal heat exchangers; reinforcement-learning pre-training on these models yields an initial control policy. The pretrained model is then deployed on a real system, where minute-level data acquisition and hourly control actions continuously update the Q-table, enabling adaptive regulation of pump frequency. During a one-month field deployment (20 May – 20 June

基金项目: 四川华西集团科技项目《基于碳中和目标的绿色零碳建筑技术体系研究及示范》(HXKX2021/024)

作者简介: 张雯琦(1999-), 男, 硕士研究生, E-mail: 903981050@qq.com

通讯作者: 袁中原(1983-), 男, 博士研究生, 副教授, E-mail: zhongyuan.yuan@home.swjtu.edu.cn

收稿日期: 2025-06-30

2025), the proposed method achieved a 4.59 % energy-savings relative to a 35 Hz fixed-frequency benchmark while fully meeting occupants' cooling demands. The impact of sensor errors on the learning method and corresponding correction strategies are also discussed. The approach partially overcomes the local-optimum limitations of traditional control schemes, achieving global energy-efficiency optimization across devices and operating conditions, and offers a new technological pathway for energy-saving control of central air-conditioning systems.

【Keywords】 Reinforcement learning; Q-learning; Online adaptation; Variable-frequency pump; Energy-saving control

0 引言

据统计,我国空调系统所消耗能源占我国社会总能耗的 21.7%^[1],空调系统节能减碳是当前全球气候变暖大背景下的大势所趋^[2]。中央空调系统的运行控制是空调系统节能减排的关键,系统层面的全局最优控制是相关研究的最终目标。但在实际工程中,设备级自动控制相对成熟而设备间、系统级的智能控制仍有很大优化空间。中央空调系统设备的独立自控无法适应其他的设备的运行状态,控制目标往往只是达到适应温度、温差、压差等特定参数,也并非实现满足运行需求下的全局能耗最优。这种局部最优的控制策略忽视了系统中设备间的动态耦合特性,难以协调整体运行状态,进而导致系统整体能耗的上升。因此,仅以满足个别参数设定为目标的传统控制方法,难以适应复杂运行工况和动态负荷变化的实际需求。

其中,以空调系统中的水泵变频控制为例,目前主流的控制策略多采用定压差、定温差或基于经验规则的 PID 控制方法^[3]。这类控制方式具有结构简单、工程实现方便等优点,但在面对负荷动态变化、水力系统复杂耦合及冷热源侧多因素干扰时,表现出明显的局限性。具体而言:定压差控制常常无法适应末端设备负荷变化,容易导致能耗浪费;而定温差控制对系统响应较慢,难以快速匹配真实需求^[4]。此外,基于静态设定点的控制策略缺乏对运行环境的感知能力,难以实现系统能效的最优协调。

为实现在保障用户需求前提下的空调系统全局能耗最小化,有必要探索一种具备自适应能力、能够动态优化控制策略的先进水泵控制方法。在此背景下,强化学习(Reinforcement Learning, RL)^[5]作为一种基于环境交互学习最优策略的人工智能方法,近年来在复杂系统控制中的应用受到广泛关注。强化学习能够在不依赖系统精确模型的前提下,基于历史数据和实时反馈不断优化控制策略,具有良好的泛化能力和动态适应性。将强化学习方

法引入空调系统水泵变频控制,有望突破传统控制方式的局限,实现跨设备、跨工况的协同优化控制,进而提升系统整体能效水平。

因此,本文旨在研究基于强化学习的空调系统水泵变频控制策略,通过构建系统状态感知与反馈机制,实现对不同工况水泵运行的智能优化控制,为中央空调系统节能控制提供新思路和技术路径。

1 系统介绍

本文涉及的空调系统服务四川省成都市的某办公建筑,建筑面积约 15 万 m³,共有 13 层,建筑始建于 1989 年,于 2017 年翻新改造,单位面积年能耗由 38.44kWh/m³降低到 36.19kWh/m³,建筑配套的空调系统设计年平均性能系数(Coefficient of Performance, COP)为 3.35;设计供回水温度夏季为 7/12℃,冬季为 45/40℃^[6]。

本文研究对象为该办公建筑的空气源热泵空调系统,系统流程示意图如图 1 所示。该系统以 12 台模块式空气源热泵作为冷热源,分 3 组布置,每组设置 4 台。模块式空气源热泵机组和立式单级离心泵的主要铭牌参数如表 1、表 2 所示。冷热输配系统动力来源于 3 台立式单级离心泵并联运行,运行调节方式为手动变频调节,水泵频率设定值仅依靠运维人员经验,无变频控制逻辑。

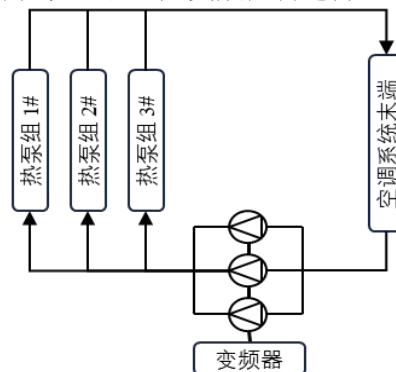


图 1 本文空调系统布局

Fig.1 Layout of the Proposed HVAC System

表 1 模块式空气源热泵机组铭牌参数

Table 1 Nameplate of the Modular Air-Source Heat

Pump Unit			
参数类型	参数值	参数类型	参数值
名义制冷量	130kW	名义制热量	132kW
压缩机数量	4 个	水流量 (冷/热)	373.3/384.8L/min
额定频率	50Hz	性能系数	3.35
冷媒种类	R410a	额定功率	38.8/39.2kW

表 2 立式单级离心泵铭牌参数

Table 2 Nameplate Data of the Vertical Single-Stage

Centrifugal Pump			
参数类型	参数值	参数类型	参数值
流量	100m ³ /h	必须汽蚀余量	3m
扬程	32m	配套功率	15kW
转速	1480r/min	——	——

2 方法介绍

2.1 经典强化学习算法 Q-learning

Q-learning^[7]是一种经典的强化学习算法,其核心思想是通过与环境的反复交互,学习在不同状态下采取何种动作能够获得最大的长期收益。在强化学习问题中,智能体在某一状态 s_t 下选择一个动作 a_t ,环境则根据该动作反馈一个即时奖励 r_t ,并转移至下一个状态 s_{t+1} 。Q-learning 的目标就是学习一个动作价值函数 (即 Q 函数),表示在某一状态下采取某一动作所能获得的预期累积奖励。在与环境的交互过程中,产生的周期性样本序列如式 (1) 所示。

$$\{s_0, a_0, r_0, s_1, a_1, r_1, \dots, s_t, a_t, r_t, \dots, s_{T-1}, a_{T-1}, r_{T-1}, s_T\} \quad (1)$$

在此基础上,将从该时刻开始到周期结束的折扣奖励之和作为 t 时刻的回报 R_t 。

$$R_t = r_t + \gamma r_{t+1} + \gamma^2 r_{t+2} + \dots + \gamma^{T-t-1} r_{T-1} = \sum_{k=0}^{T-t-1} \gamma^k r_{t+k} \quad (2)$$

动作值函数被定义为:

$$Q(s, a) = E \left\{ \sum_{k=0}^{T-t-1} \gamma^k r_{t+k} \mid s_t = s, a_t = a \right\} \quad (3)$$

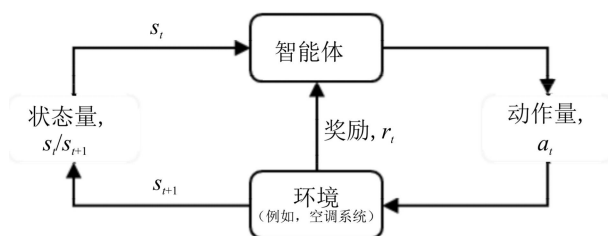


图 2 强化学习范式

Fig.2 Reinforcement Learning Paradigm

Q-learning 不依赖于环境的先验模型,通过经验不断更新 Q 函数,其更新公式:

$$Q(s, a) = Q(s, a) + \alpha \left[r_t + \gamma \max_{a'} Q(s', a') - Q(s, a) \right] \quad (4)$$

式中: s 和 a 分别表示当前状态和动作; r_t 是执行该动作后获得的即时奖励; s' 是动作后环境的新状态。 α 是学习率,控制每次更新的幅度,也即每个新的“经验”对现有策略的影响程度; γ 是折扣因子,表示对未来奖励的重视程度。

表 3 Q-learning 伪代码

Table 3 Q-learning Pseudocode

Q-learning 算法	
1:	初始化状态-动作值函数为 0
2:	for episode $i = 1, \dots$, do
3:	for step $t = 1, \dots$, do
4:	根据当前策略 π 在状态 s_t 下选择动作 a_t , 获得实时奖励值 r_t , 进入 s_{t+1} ;
5:	使用当前策略 π 在状态 s_{t+1} 下选择动作 a_{t+1} ;
6:	计算 $Q^\pi(s, a)$ 和目标值函数 $Q_{t \arg et}^\pi(s, a) = r_t + \gamma \max_{a'} Q(s', a')$;
7:	梯度更新 $Q^\pi(s, a) = Q^\pi(s, a) + \alpha [Q_{t \arg et}^\pi(s, a) - Q^\pi(s, a)]$;
8:	end for
9:	end for

2.2 控制问题的强化学习定义

本文涉及的空气源热泵空调系统的主要耗能设备有空气源热泵机组、循环水泵和末端风机盘管。其中,空气源热泵机组负责制取冷热量,循环水泵将储存在空调水系统中的冷热量输送至各末端风机盘管,而末端设备则根据室内用户的使用习惯和热舒适需求进行设定,以营造满足用户需求的室内环境。本项目的数据集参数包括空调系统逐时运行参数—空气源热泵机组、末端风机盘管和循环水泵总能耗、系统供回水温度、流量,以及室外干球温度和室外相对湿度,办公楼固定运行时间为周一至周五,空调启停时间通常为 8:00 至 21:00。

状态空间中的参数是控制问题的控制依据,描述了智能体所处的“环境”特征,为智能体的决策提供所有必需且完整的基本信息,涵盖了解决目标问题所涉及的所有环境属性。本文选择的状态空间参数为当前时刻供冷量 Q , 室外干球温度 T_d , 供水温度 T_s 。

$$f_{\text{pump}} \in \{30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50\} \quad (9)$$

奖励函数以降低空调系统总能耗, 约束不安全、不稳定动作为设计目的。

$$\text{reward} = \begin{cases} -(P_{\text{ASHP}} + P_{\text{pump}} + P_t), & \text{if } P_t \leq k_1 \text{ and } \text{abs}(Q - CL) \leq k_2 \\ -p_1, & \text{if } P_t > k_1 \\ -p_2, & \text{if } \text{abs}(Q - CL) > k_2 \end{cases} \quad (10)$$

式中: k_1 、 k_2 和 p_1 、 p_2 为待定系数, 也可以理解为人为引入的强化学习训练过程所需的超参数。这些超参数的具体数值或取值范围应在实际应用中探索。

2.3 本文所提出方法工作流程

如图 3 所示, 本文所提出的强化学习方法在空调系统水泵变频控制中的工作流程可简要概括为获取空调系统历史运行数据、建立数据驱动模型、进行模型上的预训练以及进行在线部署自学习四个部分。首先收集系统运行过程中的关键参数, 如水泵频率、供回水温差、压差、能耗等, 为后续建模提供数据基础; 基于收集到的历史数据, 构建系统的仿真数据驱动模型, 用于模拟不同控制策略下的系统响应; 在数据驱动模型上进行离线训练, 使用强化学习算法学习初始策略, 以降低在线训练风险并提升初始控制性能; 最终将预训练模型部署到实际系统中, 并继续通过实时交互不断调整和优化

$$S = \{Q, T_d, T_s\} \quad (5)$$

当前时刻供冷量反映了系统在当前时刻对冷负荷的响应能力, 是衡量末端需求是否被满足的重要指标。水泵的主要作用是保障冷冻水的有效输送从而实现供需平衡, 因此供冷量可以直接体现水泵运行是否满足控制目标, 是学习控制策略的重要依据。值得注意的是, 本文认为办公建筑热惰性显著且水泵控制时间步长较短, 认为负荷产生显著变化的时间远大于水泵频率控制的时间步长, 因而并未在此使用可能引入预测误差的负荷预测值。由于 Q-learning 天然的状态、动作空间离散性质, 本文状态、动作参数:

$$Q = \{50, 100, 150, \dots, 900, 950, 1000, \dots\} \quad (6)$$

$$T_d = \{\dots, 23, 24, 25, 26, 27, \dots, 38, 39, 40, 41, 42, \dots\} \quad (7)$$

$$T_s = \{\dots, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, \dots\} \quad (8)$$

动作空间参数为水泵变频器设定参数:

$$f_{\text{pump}} \in \{30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50\} \quad (9)$$

策略, 实现动态环境下的自适应能耗优化控制。其中, 数据驱动模型的建立和预训练生成初始控制策略的方法参见前述研究^[8], 各数据驱动模型的输入与输出参数如表 4 所示。

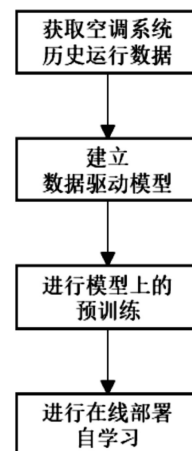


图 3 控制策略生成与应用工作流程

Fig.3 Workflow of Control Strategy

表 4 数据驱动模型输入输出参数

Table 4 Input and Output Parameters of Data-Driven Models

模型名称	输入参数	输出参数
水泵能耗-流量模型	水泵开启台数、水泵统一设定频率、冷负荷值	水泵总能耗、系统总流量
空气源热泵能耗模型	室外干球温度、供水温度、冷负荷值、系统总流量	空气源热泵总能耗
末端总能耗模型	供水温度、系统总流量、冷负荷值、时刻	末端总能耗
末端换热模型	供水温度、供水流量、末端总能耗	回水温度

最终将预训练得到的初始控制策略部署到设备中, 由于全年大部分时间水泵开启台数为两台, 且系统启停控制由本地运维人员执行, 预训练控制策略默认水泵开启台数恒定为 2。在线自学习过程中, 在线控制程序执行逻辑如图 4 所示, 程序启动后定时每分钟采集数据, 通过读取变频器当前频率实现对水泵启停状态的监测。水泵开启后需要等待一个预先设定的缓冲时间段, 以等待空调系统启动时水温的响应。当满足控制条件后, 系统将当前状态信息作为输入, 更新环境状态向量, 并随机探索

或利用: 若处于探索状态, 则执行预设的探索动作, 用于增强策略的全局搜索能力; 否则通过已训练的 Q 函数评估当前状态下各可选动作的 Q 值, 选取最优动作。在动作选择阶段, 控制策略采用 ϵ -贪婪策略 (ϵ -greedy), 以一定概率随机选取动作以保持探索能力, 其余时间选择具有最大 Q 值的贪婪动作, 确保策略的收敛性与稳定性。最终, 根据选定动作调整控制变量 (水泵变频器设定频率), 并将控制效果反馈 (能耗值) 收集起来用于更新内置 Q 表, 形成闭环学习过程。

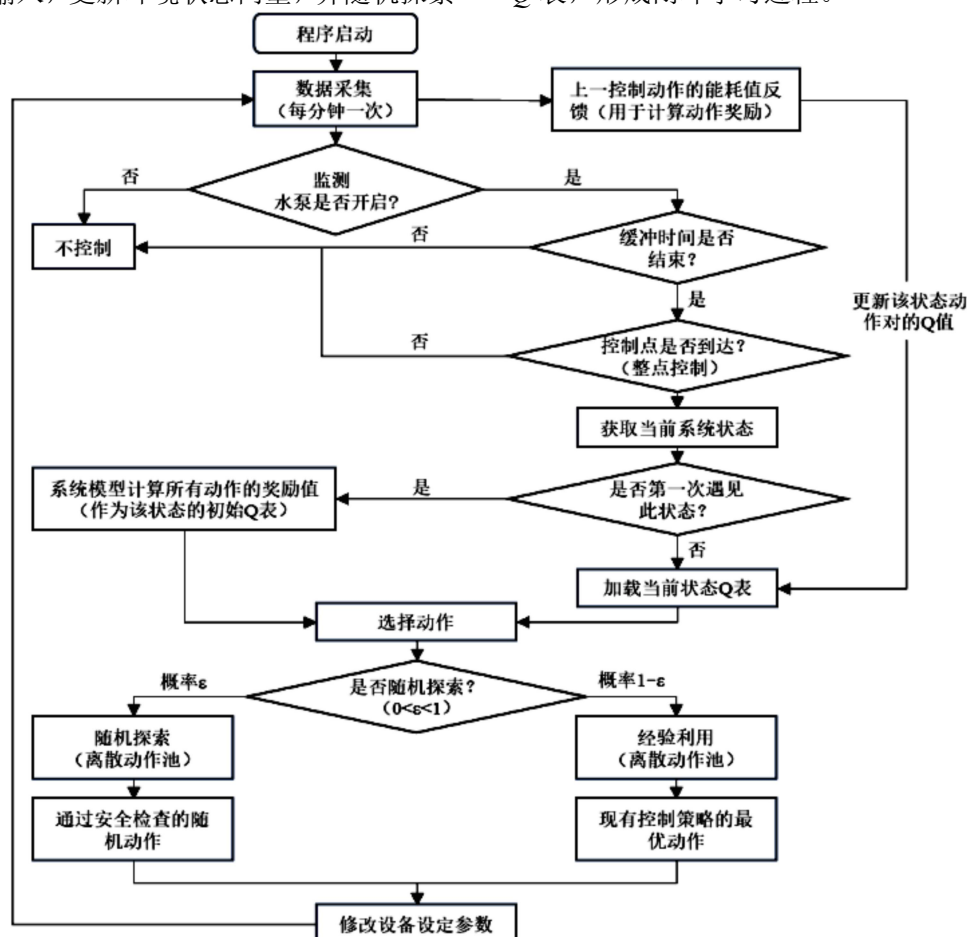


图 4 在线控制程序执行逻辑

Fig.4 Workflow of the Online Control

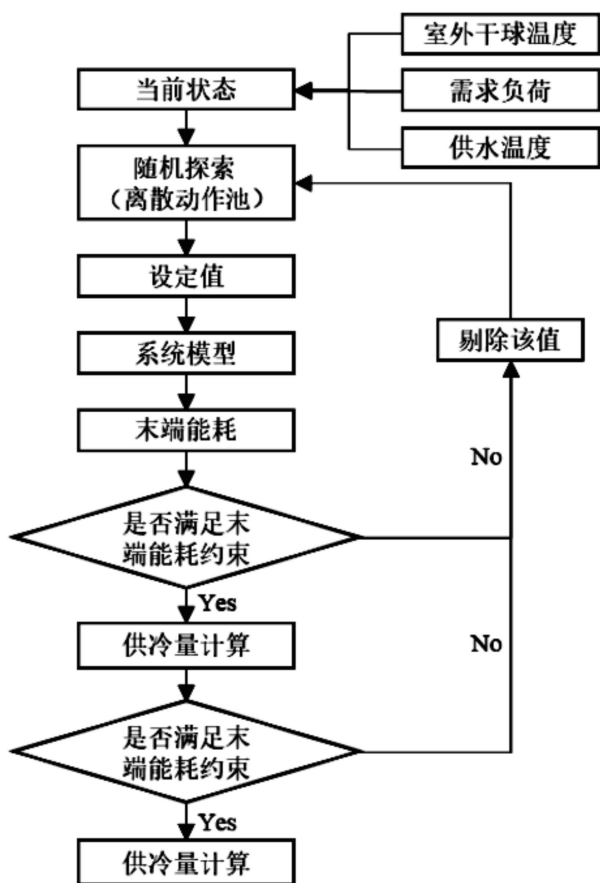


图5 安全检查逻辑

Fig.5 Safety Check

为确保强化学习算法在空调系统中的实际部署后在线自学习过程中的安全性与可行性,本文在策略探索阶段随随机探索机制引入了约束,其决策逻辑如图5所示。该机制在执行探索动作前,利用系统状态及程序内调用的数据驱动模型对所生成的探索值进行逐级筛选与校验,以规避潜在的不合理动作。具体而言,算法首先基于当前环境状态(包括室外干球温度、供水温度及系统冷负荷)生成随机探索动作,即给出一个完全随机、脱离历史经验的控制设定值。随后,该设定值被输入至数据驱动模型中,用以估算其对应的末端能耗。末端设备天然存在的能耗上限等同于末端换热能力的上限,为保障末端设备的正常运行,系统首先判断该动作设定值需要的末端出力是否在末端设备的能力范围内;若不能满足,则视为无效探索动作,予以剔除。若满足末端能耗约束,算法进一步判断当前动作的供冷能力是否满足冷负荷需求,以确保满足建筑冷

量需求。只有同时满足这两级约束条件的探索动作,才被认为是“安全”的,并可参与后续的强化学习策略更新过程。这样的设计体现了强化学习在复杂物理系统控制中的工程适应性考虑。一方面,通过引入硬约束校验机制,有效避免因盲目探索带来的控制偏差或末端冷热失衡问题,保障了系统运行的稳定性与舒适性;另一方面,该机制通过合理裁剪动作空间,显著提升了探索效率,有助于提升强化学习策略的收敛速度与收敛质量。

3 实验结果

3.1 强化学习预训练获得初始控制策略

表5和表6对Q-learning中主要的超参数进行了简要阐述与取值。将历史2个制冷季节共2190个样本的状态参数组合作为强化学习预训练的输入引导控制策略的学习,图6展示了2190个状态为一回合的回合奖励变化情况。

表5 Q-learning 主要超参数及释义

Table 5 Q-learning Hyperparameters and Descriptions

超参数	解释
Learning rate	学习率,表示每次更新 Q 值时新信息融合的比例
γ	折扣因子,表示对未来奖励的重视程度
Start exploration rate	初始探索率
Exploration fraction	探索周期比率,初始探索率衰减到最终探索率经历的回合数占比
Final exploration rate	终止探索率
k_1	奖励函数中,对末端能耗值的约束,单位 kWh
k_2	奖励函数中,对负荷供应与需求之间绝对差值的约束,单位 kW
p_1/p_2	违反约束时,智能体的“惩罚”奖励值

表6 Q-learning 训练设置

Table 6 Q-learning Pre-training Settings

超参数	取值	超参数	取值
Learning rate	0.5	γ	0.99
Start exploration rate	1	Final exploration rate	0.01
Exploration fraction	0.3	k_1	23
k_2	100	p_1/p_2	500

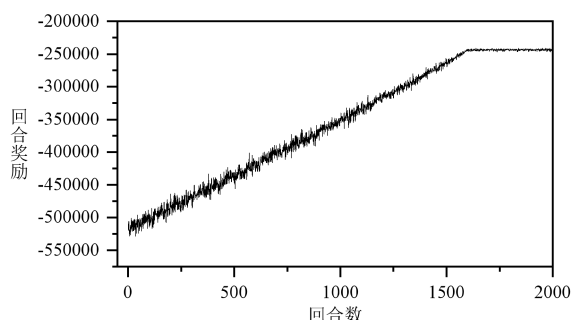


图 6 回合奖励变化曲线

Fig.6 Episode Reward Curve

3.2 实地部署结果

为验证本文提出的基于 Q-learning 的水泵变频控制策略在真实环境中的节能效果与控制性能, 本文于 2025 年 5 月 20 日至 6 月 20 日实施为期一个月的在线部署实验, 将设备电表能耗读数与仿真模型中传统固定频率设定值控制方式 (35Hz) 的能耗结果进行了对比。表 7 展示了不同控制策略下空调系统能耗的分项情况及节能表现。从实验结果可见, 在整个运行周期内, 采用 Q-learning 控制策略的空调系统总能耗为 23358.3kWh, 较基准策略的 24430.9kWh 下降了 1072.6kWh, 实现了 4.59% 的整体节能效果。其中, 空调系统中能耗占比较高的空气源热泵与水泵部分分别实现了 10.79% 和 6.59% 的能耗下降, 表明强化学习策略在系统供能设备的联动控制上具有明显的节能优势。需要指出的是, 末端设备由于独立于空调机房由室内使用人员控制, 常与空调机房错时运行, 造成了机房非供冷时间段额外的设备能耗浪费。尽管如此, 从全局能耗角度来看, 该策略仍能有效协调水泵和冷源侧设备运行, 实现系统综合能效的提升, 验证了其在实际空调系统中的部署可行性与节能潜力。

表 7 控制策略在线部署能耗情况及仿真对比

Table 7 Energy Consumption of the Online-Deployed Control Strategy versus Simulation

空调系统能耗 分项 (kWh)	空气源热 泵能耗	水泵能 耗	空调末端 能耗	总能耗
Q-learning 控制 策略	16692.8	2562.8	4102.7	23358.3
基准设定值 (35Hz)	18494.7	2731.8	3204.4	24430.9
能耗变化值	1801.9	169.0	-898.3	1072.6
能耗变化率	10.79%	6.59%	-21.89%	4.59%

在实现系统整体能耗优化的同时, 本文进一步分析了部署期间各关键监测点的室内温度表现, 以验证控制策略在保障末端用能需求的有效性。本文在三个典型楼层选取了换热最不利温度点的走廊区域布设温度传感器, 对空调运行时段内重点监测其室内空气温度变化情况。图 7 展示了部署期内各监测点温度数据的统计分布, 包括其分位数、上下四分位距。其中, 4 楼与 8 楼在空调运行时段内的温度分布整体更为集中, 主要维持在 24.5~27.5℃ 之间, 满足室内办公人员热舒适需求。而 13 楼受外围护结构影响天然具有高冷负荷值, 水泵的动态调节带来的换热能力改变远小于水温动态调节带来的换热能力改变, 故本文所讨论的方法对这一症状缓解程度不大。由于监测温度位于末端换热的最不利位置, 实际办公区域的温度普遍低于最不利温度点 1~3℃, 在控制策略部署阶段 13 楼办公人员勉强达到热舒适。

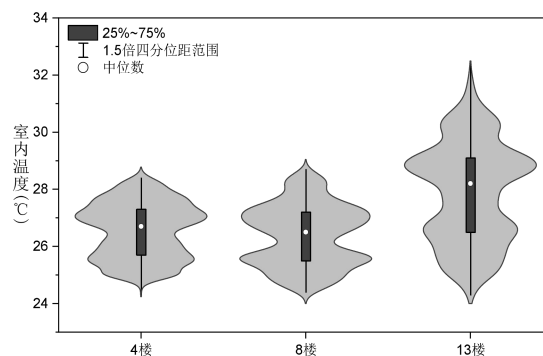


图 7 空调运行时室内室温分布情况

Fig.7 Indoor Temperature Distribution During HVAC Operation

3.3 控制动作频率变化

为进一步分析强化学习控制策略在实际运行过程中的行为特征与响应能力, 本文选取部署期内具有代表性的一天, 对空调系统的关键状态参数与控制动作变化情况进行了追踪与可视化, 如图 8 所示。可以看到, 策略在多数时段内连续执行了相同的 30Hz 频率动作, 尤其在系统负荷变化较缓或处于平台期时, 相邻控制点之间缺乏明显的动作微调。这一现象与 Q-learning 控制策略中状态空间和动作空间均采用离散化处理密切相关。具体而言, 为降低 Q 表的维度复杂度与训练难度, 本文在策略设计阶段对供冷量、供水温度、外界温度等状态变量进行了分箱离散化处理, 并设定了有限的动作集合。

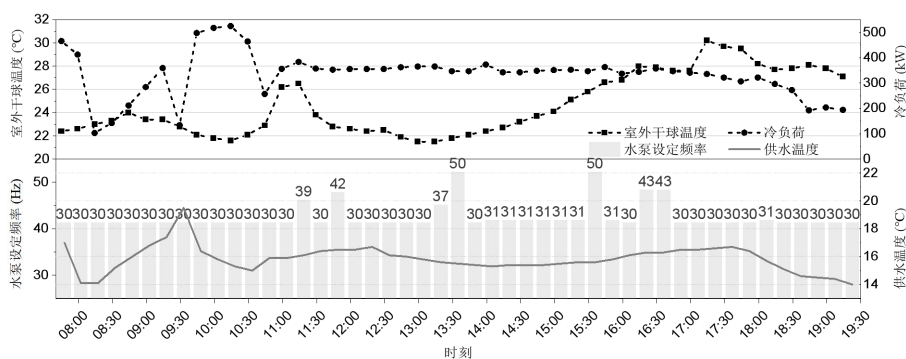


图8 典型日策略动作执行情况

Fig.8 States and Actions on a Typical Day

虽然这样的离散化设计能够有效提升学习效率和策略稳定性,但也可能造成策略在实际部署中难以捕捉状态变量的细微波动,从而出现连续重复动作的情况。这样的动作选择可能降低了控制精度,但冷负荷测量本身存在一定噪声和误差,状态变量的不确定性使得过度敏感的动作调整反而可能引发系统不稳定。因此,离散化处理在一定程度上也起到了稳定调节的正面作用。此外,控制动作中偶发的高水泵频率值一方面是 Q-learning 不考虑任何既有“经验”的随机探索,另一方面又可能是该状态在预训练过程中并没有找到最优动作,仍然停留在次优动作值上。

3.4 传感器偏差带来的影响

在线部署阶段,本文中有三个参数存在显著的测量差异,分别为室外干球温度、水流量、供(回)水温度。对于室外干球温度,由于预训练阶段使用的是全球气象参数公开数据集而部署阶段使用的是本地气象站监测数据,这一参数并没有可供误差校准的测点,因此仅能通过控制策略的在线学习机制逐渐适应传感器的偏差;对于水流量,尽管流量计对水系统总管流量的监测存在显著差异,本文考虑对部署阶段流量计读数乘以值为 1.3 的修正系数,修正后的读数差异在 5%以内,视为工程上可接受的读数误差。下面讨论供回水温度测点的误差对本方法的影响。

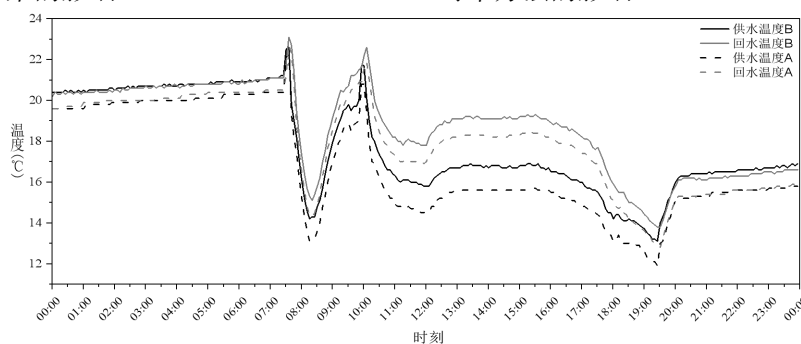


图9 不同传感器温度测量值曲线

Fig.9 Temperature Measurement Curves from Different Sensors

图9为预训练数据来源的传感器A和在线部署阶段传感器B在同一天读数的变化情况,图10为其对应的逐时测温偏差(A-B)。可以看到,不同时刻对同一温度测点的测量出现了0.3~1.4℃范围内不同程度测量偏差。一方面,供回水温度的测量偏差直接影响供冷量计算结果,直接导致状态值的偏差;更重要地,该误差对于本文方法的影响主要体现在数据驱动模型中末端换热模型对末端换热效果的估计偏差上:在低负荷工况下(供水温度

16℃,流量120m³/h,末端能耗5kWh),0.3~1.4℃的供水温度测量偏差会直接导致0.27~1.01℃的回水温度高估,从而低估3.7~12.0kW(2.88%~9.63%)供冷能力;在高负荷工况下(供水温度10℃,流量180m³/h,末端能耗20kWh),0.3~1.4℃的供水温度测量偏差会直接导致0.17~0.73℃的回水温度高估,从而低估25.5~56.7kW(3.25%~7.46%)供冷能力。由此可以看到,若水温偏差值浮动不定,未来的在线学习阶段可以考虑按照一定百分比修

正被低估的供冷能力,使得控制策略更加匹配地进行在线自学习。

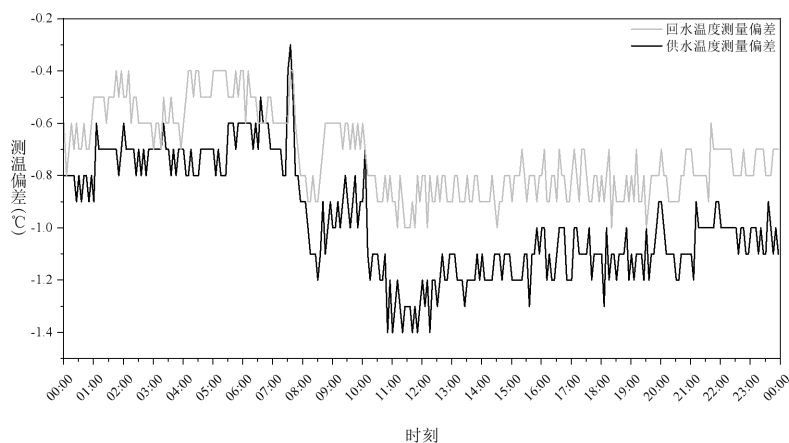


图10 测量偏差逐时分布曲线

Fig.10 Measurement Deviations Curves from Different Sensors

4 结论

本文针对设备间参数耦合的中央空调系统水泵控制问题,系统地设计并实现了基于 Q-learning 的在线自适应变频控制框架,引入强化学习以期在无精确模型前提下实现控制策略的动态优化。该方法通过历史数据训练的黑箱模型完成离线预训练,降低了在线学习风险的同时提升了初始策略性能;在线部署阶段,结合 ϵ -贪婪探索与安全约束机制,实现了对水泵频率的实时调整与闭环优化。一个月的实地验证表明,该策略在制冷季保证人员用冷需求的同时在总能耗、关键设备能耗方面均取得一定改善,同时针对传感器偏差这一常见工程问题给出了一定的修正思路。

然而,研究仍存在以下不足:一是在为期一个月的部署周期中,控制策略未能充分展示策略的在线演进过程,可能的原因是学习时长不足或学习效率偏低;二是所用仿真模型的精度与泛化能力对在线学习效果影响显著,亟需结合高精度数字孪生技术以提升策略快速收敛能力。未来工作可沿以下方向展开:首先,可以将在线自适应变频控制方法推广至水冷冷却水泵和冷冻水泵场景,以适应冷塔、冷机与负荷需求的多变工况;其次,研究有效的策略成长度量与在线性能评估指标,用于定量捕捉强化学习控制策略的适应性与成长趋势;最后,探索以数字孪生技术营造虚拟环境的虚实结合空调系统强化学习控制策略,以在增强控制策略的鲁棒性和泛化能力同时,高效快速地对空调系统控制策略

进行演化更新。

参考文献:

- [1] 清华大学建筑节能研究中心.中国建筑节能年度发展研究报告 2021[M].北京:中国建筑工业出版社,2021.
- [2] 李先庭,赵阳,魏庆芑,等.碳中和背景下我国空调系统发展趋势[J].暖通空调,2022,52(10):75-83,61.
- [3] Wang S, Burnett J. Online adaptive control for optimizing variable-speed pumps of indirect water-cooled chilling systems [J]. Applied Thermal Engineering, 2001,21(11):1083-1103.
- [4] 王寒栋.空调冷冻水泵变频控制方式分析与比较[J].制冷空调与电力机械,2004,(1):16-21.
- [5] Sutton R S, Barto A G. Reinforcement learning: An introduction [M]. MIT press, 2018.
- [6] Gao B, Zhu X, Ren J, et al. Multi-objective optimization of energy-saving measures and operation parameters for a newly retrofitted building in future climate conditions: A case study of an office building in Chengdu [J]. Energy Rep, 2023,9:2269-85.
- [7] Watkins C J C H. Learning from delayed rewards [D]. Kings College University of Cambridge, 1989.
- [8] Zhang W, Yu Y, Yuan Z, et al. Data-driven pre-training framework for reinforcement learning of air-source heat pump (ASHP) systems based on historical data in office buildings: Field validation [J]. Energy and Buildings, 2025,332.