文章编号：1008－8857(2025)01－0027－06 **DOI**: 10.13259/j.cnki.eri.2025.01.004

基于深度强化学习策略的空气源热泵除霜能效提升

张 博，田煦嘉
（北京北辰实业股份有限公司，北京 100100）

摘 要：热泵除霜技术在空调等制冷系统中发挥着重要的作用，但传统的除霜方法存在能耗高、效率低、系统性能下降等问题。采用先进的强化学习技术，对热泵除霜过程进行了模型优化，采用 Q−learning 算法及 epsilon−greedy 策略学习一个最优的动作值函数，同时在探索和利用之间进行权衡。结果表明：基于 Q−learning 算法构建的模型能够准确预测热泵除霜过程中的参数变化，并可实现热泵系统更高的能源利用效率。研究可为热泵除霜技术的优化提供理论基础和实际指导。

关键词：热泵；除霜；强化学习；Q−learning 算法；epsilon−greedy 策略；能耗；效率
中图分类号：TU83 **文献标志码**：A

Efficiency improvement of air source heat pumps using deep reinforcement learning-based defrosting strategy

ZHANG Bo，TIAN Xujia
(Beijing North Star Company Ltd., Beijing 100100, China)

Abstract: The defrosting technologies of heat pumps are pivotal in air conditioning and refrigeration systems. Nonetheless, conventional defrosting techniques are plagued by high energy consumption, limited efficiency, and reduced system performance. This research optimized the heat pump defrosting process model using advanced reinforcement learning techniques, emphasizing exploration-exploitation equilibrium and deriving an optimal action-value function via the Q-learning algorithm and the epsilon-greedy approach. Findings demonstrate that the Q-learning algorithm-based model accurately forecasts parameter variations in the heat pump defrosting process, improving energy utilization efficiency within the heat pump system. It offers both theoretical underpinning and practical insights for optimizing heat pump defrosting technologies.

Keywords: heat pump; defrosting; reinforcement learning; Q-learning algorithm; epsilon-greedy approach; energy consumption; efficiency

收稿日期：2023－09－26
第一作者：张 博(1982—)，男，硕士，工程师。研究方向：热能与动力工程。E−mail：gikisun2@126.com

空气源热泵在能源领域的应用中有着显著的能量效率优势。除霜是热泵系统中一个重要的过程，其目的是移除热泵蒸发器表面的结霜或冰层，以改善热交换效果。传统的除霜方法存在的能耗高、效率低、系统性能下降等问题，一直是限制其更广泛应用的主要技术难题。近年来，研究人员深入探讨了空气源热泵在结霜区的运行性能和面临的挑战，并提出了一些新方法以减少热泵除霜造成的不必要的能源浪费。例如，王沣浩等[1]在研究补气增焓技术的同时探讨了双级压缩式热泵循环系统的潜力，并提出复叠式循环也是一种较好的解决方案。董建锴[2]提出了通过改变室外风机的风向和风量来延缓结霜的新方法。李刚等[3]发现预先降低空气湿度有助于延缓结霜。刘启芬等[4]建立了一个模糊除霜控制系统。该系统不仅能根据大气温度、翅片温度和风机电流进行动态调整，还能根据除霜效果自适应地调整控制规则。王世权等[5]利用图像识别技术对空气源热泵结霜程度进行直接监测，有效解决了室外光照变化影响测霜准确性的问题。钱付平等[6]提出了空气源热泵空调系统的节能措施，即应选用高效率的压缩机，并采用强化传热措施提高传热系数，减小传热温差，同时还应注意改善热泵机组周围环境。

随着人工智能的发展，Eom 等[7]对全连接深度神经网络(FCDNN)、卷积神经网络(CNN)和长短时记忆(LSTM)模型进行了比较，发现 FCDNN 模型在高准确性和数据收集方面的表现更为优越。Shin 等[8]利用诸如人工神经网络、支持向量机、随机森林和 K−最近邻模型等机器学习技术，开发了热泵系统的性能预测模型。通过揭示输入值和输出值之间的关系，该模型能在无需额外安装昂贵的性能测量设备或各种监测传感器的情况下实现热泵操作期间的实时性能参数测量和监控，进而促进系统优化，提升故障诊断和运行效率。另一项研究[9]通过收集和分析来自 11 个不同制造商的 33 台大型热泵的系统配置和性能的综合数据集，突显出大数据在大规模热泵系统性能分析和改善方面的重要性，从而为热泵技术的进一步优化提供了宝贵的数据支持。本文专注开展热泵除霜过程的模型优化工作，而非构建全新的理论框架。研究核心在于运用先进的 Q−learning 算法，对现有热泵除霜流程进行精细化调整与改进。

本文通过深入分析历史环境数据和热泵运行数据，结合 Q−learning 算法的强大学习能力，旨在发掘一种更高效、更节能的除霜策略，以期显著降低热泵除霜过程中的能耗，同时提升系统的整体性能和稳定性，从而在实际应用中实现更高的能源利用效率和更好的除霜效果。

1 强化学习算法

首先，定义热泵除霜过程的状态和可能采取的行动，然后通过模拟除霜过程进行交互学习，逐步优化 Q−learning 算法。采用 epsilon−greedy 策略在探索和利用之间进行权衡，以提高学习的效率和结果的稳定性。该策略需在一个迷宫中找到目标点。初始阶段，epsilon−greedy 策略的探索率较高，智能体会选择随机动作进行探索，以便更好地了解环境。随着学习的进行，随机程度逐渐减弱，智能体更多地利用当前最优动作，以获得更高的累积奖励。这样，智能体就能够在探索和利用之间找到平衡，实现最优策略的学习。

1.1 Q−learning 算法原理阐述

Q−learning 算法是基于马尔可夫决策过程(MDP)的框架。MDP 由五元组(S, A, P, R, γ)定义，其中：S 表示状态空间，即智能体可能处于的不同状态集合；A 表示动作空间，即智能体可能采取的行动集合；P 表示状态转移概率，即描述在状态 s 下采取行动 a 后转移到下一个状态 s'的概率；R 表示奖励函数，即给出在状态 s 下采取行动 a 后获得的即时奖励；γ 表示折扣因子，用于衡量未来奖励的重要性。Q−learning 算法旨在寻求一个最优的动作值函数 $Q(s, a)$，用于表示在状态 s 下采取行动 a 所能获得的累积奖励。该算法通过不断更新 Q 值表来逼近最优策略。

1.2 Q−learning 算法实现步骤

(1)初始化 Q 值表：将所有 Q 值初始化为 0 或随机值。

(2)选择动作：根据当前状态 s 和 Q 值表选

择一个动作 a。

(3)执行动作并观察环境：智能体执行选择动作 a，并观察环境返回的下一个状态 s'和即时奖励 r。

(4)更新 Q 值：根据 Bellman 方程更新 Q 值表的相应条目，即

$$Q_{t+1}(s,a)=Q_t(s,a)+\alpha\left[r+\gamma\max Q_t(s',a')-Q_t(s,a)\right] \tag{1}$$

式中：$Q_t(s,a)$、$Q_{t+1}(s,a)$分别为迭代步数为 t、t+1 时的 Q 值；α 为学习率，其控制 Q 值更新的速度；a'为下一个状态 s'下采取的行动。

(5)转移到下一个状态：将下一个状态 s'设置为当前状态 s，并返回步骤(2)。

(6)重复执行步骤(2)～(5)，直到达到停止条件(如达到最大迭代次数或收敛)。

1.3　epsilon−greedy 策略简介

epsilon−greedy 策略是一种常用的强化学习策略，用于在探索和利用之间进行权衡。该策略通过在一部分时间选择最优动作，而在另一部分时间进行随机探索，从而实现智能体在学习过程中的平衡。该策略通过设定一个探索率ε来决定智能体在选择动作时是进行探索还是进行利用。具体而言，当随机数小于ε时，智能体选择一个随机动作进行探索；当随机数大于等于ε时，智能体选择当前最优的动作进行利用。

1.4　epsilon−greedy 策略实现步骤

(1)初始化参数：设置ε值和动作值函数 Q。

(2)选择动作：生成一个随机数r_0，如果$r_0<\varepsilon$，则选择一个随机动作作为当前动作；否则，仍选择当前最优动作。

(3)执行动作并观察环境：智能体执行选择的动作，并观察环境返回的下一个状态和即时奖励。

(4)更新动作值函数：根据得到的奖励更新动作值函数 Q。

(5)转移到下一个状态：将下一个状态设置为当前状态，并返回步骤(2)。

(6)重复执行步骤(2)～(5)，直到达到停止条件(如达到最大迭代次数或收敛)。

1.5　模型代码

建立一个基础的强化学习环境来模拟真实环境下的热泵除霜过程。图 1 为通过编程实现的一个模型代码简图。该环境有一个状态变量 ice_state，它代表冰的累积量。每一步可以选择两种动作：“heat”和“defrost”。这两种动作会分别减少或增加 ice_state，以便找出在给定状态下的最优动作。图 2 展示了强化学习算法的基本循环，包括初始化环境、选择动作、执行动作并获取反馈、更新智能体状态或策略，以及检查是否终

```
# Simplified Environment class showcasing core logic
class Environment:
    def __init__(self):
        self.ice_state = 0  # Initialize ice state

    def step(self, action):
        reward_value = self.reward(action)  # Compute reward
        # Update ice state based on action
        self.ice_state = max(0, self.ice_state - 1) if action == 'heat' else min(1(
        # Check if episode is done
        done = self.ice_state == 10
        return self.ice_state, reward_value, done

    def reward(self, action):
        return -100 if self.ice_state >= 10 else (1 if action == 'heat' else -10)

# Q-Learning loop (simplified)
for epoch in range(num_epochs):
    env = Environment()
    state = env.ice_state
    while not done:
        action = choose_action(state, epsilon)  # Choose action
        next_state, reward, done = env.step(action)  # Perform action
        # Update Q-table (details omitted)
        # ...
```

图 1　模型代码简图

Fig. 1　Code overview of the model

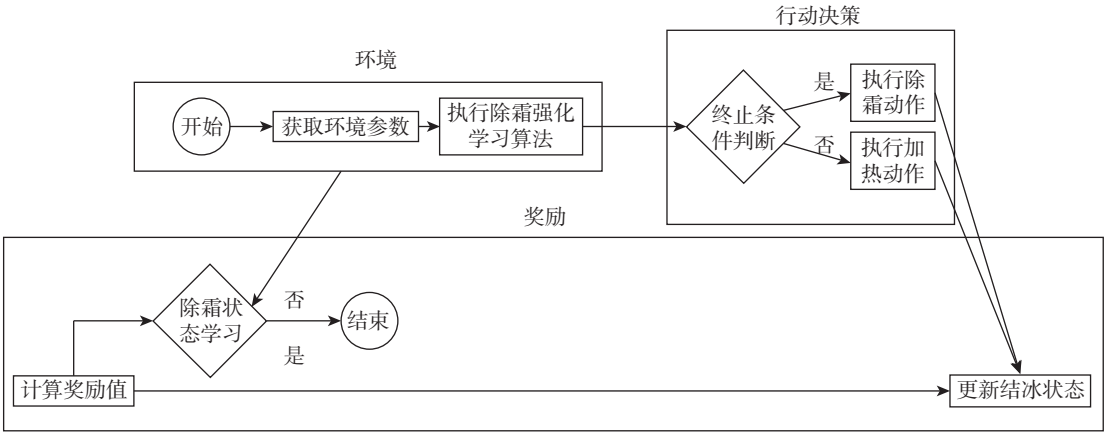

图 2 算法流程图

Fig. 2 Flowchart of the algorithm

止。如果回合未终止，则继续选择动作；否则，结束当前回合并可能开始新的回合。

需要注意的是强化学习算法不仅仅注重以往传统的物理因果关系，更能够体现出在实时反馈条件下智能体做出准确决策的相关性。

2 数据处理

本文除霜模型数据源自北京首都机场连续 10 年的实际环境参数，数据来源由站点“Reliable Prognosis”提供。具体数据是基于北京 (机场) 气象站，气象参数 METAR 代码为 ZBAA。为了便于更直观地理解数据集的结构和内容，表 1 给出了北京首都机场部分环境参数数据集。有关气象参数的详细代码和说明，可参考网址 (http://rp5.ru/metar.php?metar=ZBAA&lang=cn) 上的内容。

从环境参数数据集中筛选出以往学者的研究成果中对热泵结霜具有主要影响的因素，主要变量包括室内温度 t_{in}、室外温度 t_{out}、大气压、湿度、风向和风速，且均可直接从实地测量得到。为了保证数据的一致性和准确性，需对数据进行必要的整理(室内温度始终维持在 20 ℃)。图 3 为环境参数变化。

3 数据分析

利用 Q−learning 模型可以直观地观察和分

表 1 北京首都机场部分环境参数数据集

Tab. 1 Environmental parameter dataset of Beijing Capital International Airport

时间戳	室内温度/℃	室外温度/℃	大气压/kPa	湿度/%	风向	贝福特风力等级
31.12.2 022 23:00	20	−4.8	102.91	43	东风	1
01.01.2023 23:00	20	−5.3	103.32	62	北风	1
01.01.2023 20:00	20	−4.3	103.28	55	北风	0
01.01.2023 17:00	20	1.2	103.13	35	西南风	1
01.01.2023 14:00	20	3.4	103.08	31	东南风	3
01.01.2023 11:00	20	1.2	103.21	39	东风	3
01.01.2023 08:00	20	−5.9	103.17	52	东南风	1
01.01.2023 05:00	20	−7.5	103.08	59	东南风	1
01.01.2023 02:00	20	−6.9	102.99	52	东风	1

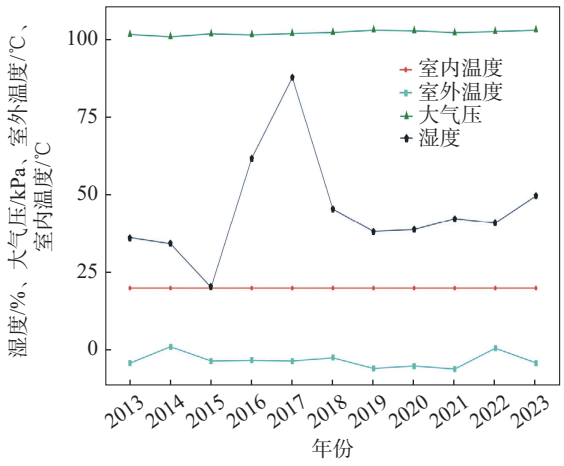

图 3　环境参数变化

Fig. 3　Variation of environmental parameters

析热泵除霜过程中各参数的变化趋势和相互作用。对不同工况下的除霜过程进行模拟，并与实际实验数据进行对比和验证。

Q 值图提供了一个可量化的度量，可直观体现算法在执行特定行动(加热或除霜)预期能够带来的效益或回报。图 4 为训练后局部细分温度区间 Q 值分布图，其中行代表状态(结冰状态)，列代表行动(加热和除霜)。颜色深浅代表 Q 值的大小，颜色越深代表该状态下采取该行动的 Q 值越大。从图 4 中可以看出，对于大多数的结冰状态(即行)，模型更倾向于执行“defrost”(除霜)操作，因为“defrost”列的 Q 值普遍高于“heat”列。

图 4 的结果是符合预期的。因为在环境设置

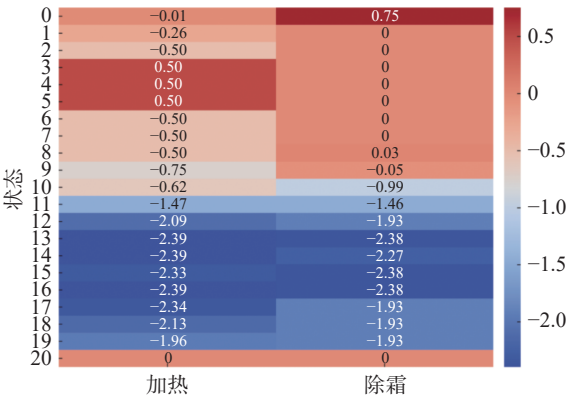

图 4　训练后局部细分温度区间 Q 值分布

Fig. 4　Q-value distribution of local subdivision temperature range after training

和奖励函数中，除霜操作会大幅减少结冰状态，因此通常会得到更高的奖励。同时可以看出，Q 值在不同的状态间有变化，这说明模型在学习过程中已学会区分不同状态，并对不同状态采取不同行动策略。

图 5 显示了在每个训练轮次中模型获得的总奖励。总奖励为两种状态的叠加态。由图 5 中可直观地看出波动趋势与以下因素有关：①环境状态的变化：环境状态是从真实的天气数据中抽取的，因此它们会随着时间的推移而变化。这意味着在某些时间步时环境状态可能更倾向于触发除霜，而在其他时间步时则不是。②模型的学习过程：训练初期模型可能会经常选择错误的行动，导致奖励较低。随着训练的进行，模型应能够学习到更好的策略，从而获得更高的奖励。

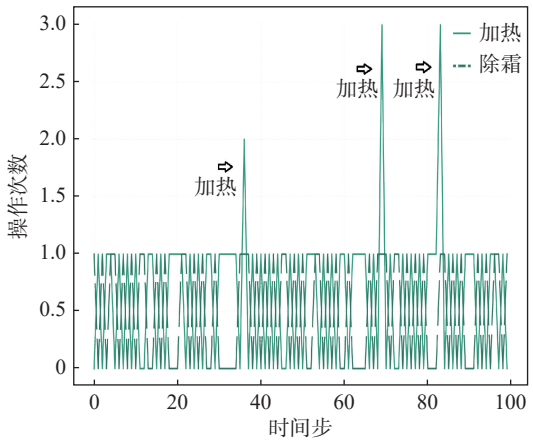

图 5　每个训练轮次的总奖励

Fig. 5　Total reward of each training round

图 6 显示了在每个训练轮次中模型选择加热操作和除霜操作次数对比。可以看到，随着训练的进行，无效除霜操作次数逐渐减少。

图 7 显示了在每个训练轮次中模型选择除霜操作的占比。可以看到，随着训练的进行，除霜操作的占比逐渐减少。

图 8 显示了每个训练轮次的能量消耗。可以看到，尽管能量消耗在整个训练过程中有所波动，但总体上呈现下降趋势。这表明随着训练的进行，该模型可更有效地管理能量消耗，从而减少不必要的除霜操作(为了突出效果，模型中假设除霜操作的能量消耗是加热操作的 10 倍)。

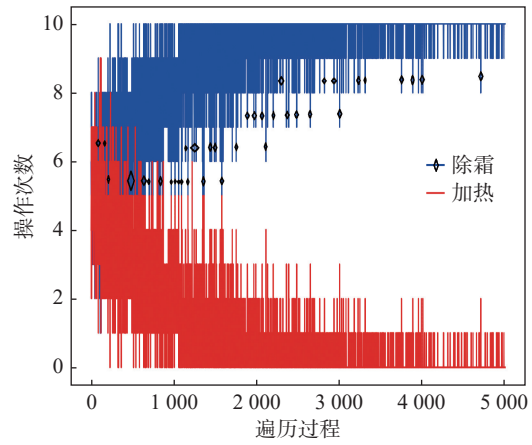

图 6 每个训练轮次中加热操作和除霜操作次数对比

Fig. 6 Comparison of heating and defrosting operations in each training round

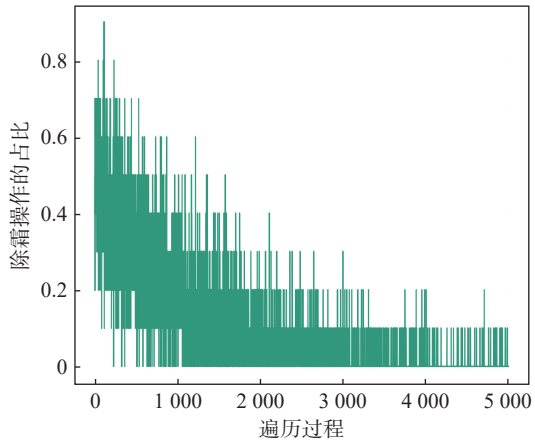

图 7 除霜操作的占比

Fig. 7 Defrosting operation percentages

4 结论

通过建立基于 Q−learning 算法的模型并结合 epsilon−greedy 策略优化技术策略证实了热泵除霜技术演变的可行性，并通过可视化方式展示了测试结果。结果表明：该模型能够准确地预测热泵除霜过程中的参数变化，并可实现热泵系统更高的能源利用效率，可以为热泵除霜技术的优化提供理论基础和实际指导，促进热泵系统的发展和应用。

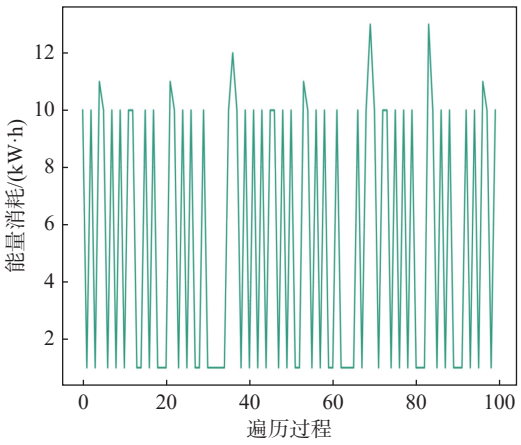

图 8 能量消耗随时间变化

Fig. 8 Evolution of energy consumption

参考文献：

[1] 王沣浩, 王志华, 郑煜鑫, 等. 低温环境下空气源热泵的研究现状及展望 [J]. 制冷学报, 2013, 34(5)：47 − 54.

[2] 董建锴. 空气源热泵延缓结霜及除霜方法研究 [D]. 哈尔滨: 哈尔滨工业大学, 2012.

[3] 李刚, 田小亮. 空气源热泵系统结霜及除霜实验研究 [J]. 科学技术创新, 2020(12)：7 − 9.

[4] 刘启芬, 黄虎. 空气源热泵模糊除霜控制的实验研究 [J]. 南京理工大学学报, 2003, 27(6)：763 − 765.

[5] 王世权, 王伟, 孙育英, 等. 基于光照自适应的空气源热泵图像识别测霜技术研究 [J]. 暖通空调, 2022, 52(7)：113 − 117, 68.

[6] 钱付平, 范树华, 张吉光. 空气源热泵空调系统节能分析 [J]. 能源研究与信息, 2004, 20(1)：23 − 28.

[7] EOM Y H, CHUNG Y, PARK M, et al. Deep learning-based prediction method on performance change of air source heat pump system under frosting conditions[J]. Energy, 2021, 228：120542.

[8] SHIN J H, CHO Y H. Machine-learning-based coefficient of performance prediction model for heat pump systems[J]. Applied Sciences, 2021, 12(1)：362.

[9] WANG W Y, JIANG J T, HU B, et al. Performance improvement of air-source heat pump heating system with variable water temperature difference[J]. Applied Thermal Engineering, 2022, 210：118366.