

基于双池 DQN 的 HVAC 无模型优化控制方法

马帅^{1,2,3}, 傅启明^{1,2,3}, 陈建平^{1,2,3}, 冯帆⁴, 陆悠^{1,2,3}, 李铮伟^{5,6}, 裘舒年^{5,6}

(1. 苏州科技大学电子与信息工程学院, 江苏 苏州 215009;

2. 江苏省智慧建筑节能重点实验室, 江苏 苏州 215009;

3. 苏州移动网络及应用技术重点实验室, 江苏 苏州 215009;

4. 德州农工大学, 美国 得克萨斯州, TX 77843;

5. 同济大学机械与能源学院, 上海 200092;

6. 同济大学工程结构性能演化与控制教育部重点实验室, 上海 200092)

摘要: 在 HVAC 控制领域, 基于模型的最优控制方法得到了学者的广泛研究与验证, 但是该方法高度依赖模型的准确性、大量历史数据的收集以及传感器的部署。针对上述问题, 结合 EnergyPlus、实际系统参数以及历史数据, 构建 HVAC 优化控制模型, 并提出一种改进的双池 DQN 算法, 最后将其应用于 HVAC 系统中不同型号冷机的负荷分配、冷却塔风机频率以及冷却水泵频率的组合优化控制。基于所构建的问题模型, 针对决策优化过程中存在的样本不平衡的问题, 该算法在 DQN 的基础上, 建立两个独立的经验池, 分别存储负荷分配与非负荷分配样本, 在训练过程中, 遵循一定的比例从经验池中进行采样, 以加快算法收敛。将所提出的方法与基于模型的控制方法及基线方法进行比较, 实验结果表明, 与基线方法相比, 基于模型的 HVAC 控制器能够节能 11.5% (最优节能效率), 而基于双池 DQN 的 HVAC 控制器在第 1 年就能够节能 7.5%, 同时, 随着系统运行, 该控制器在第 8 年左右即可获得接近最优节能效率的结果。此外, 与基于模型的 HVAC 控制器相比, 该控制器不依赖于系统模型, 且在在线控制过程中, 所需的先验知识及传感器较少, 在实际的工程应用中更具价值。

关键词: 深度强化学习; 无模型优化控制; HVAC 系统; 建筑节能

中图分类号: TP273

文献标志码: A

doi: 10.11959/j.issn.2096-6652.202208

HVAC model-free optimal control method based on double-pools DQN

MA Shuai^{1,2,3}, FU Qiming^{1,2,3}, CHEN Jianping^{1,2,3}, FENG Fan⁴, LU You^{1,2,3}, LI Zhengwei^{5,6}, QIU Shunian^{5,6}

1. School of Electronics & Information Engineering, Suzhou University of Science and Technology, Suzhou 215009, China

2. Jiangsu Province Key Laboratory Intelligent Building Energy Efficiency, Suzhou University of Science and Technology, Suzhou 215009, China

3. Suzhou Key Laboratory of Mobile Networking and Applied Technology, Suzhou 215009, China

4. Texas A&M University, College Station TX 77843, USA

5. School of Mechanical Engineering, Tongji University, Shanghai 200092, China

6. Key Laboratory of Performance Evolution and Control for Engineering Structures of Ministry of Education, Tongji University, Shanghai 200092, China

Abstract: In the field of HVAC (heating, ventilation and air conditioning) control, the model-based optimal control method has been extensively studied and verified by scholars, but this method highly depends on the accuracy of the model, the collection of a large amount of historical data, and the deployment of sensors. In response to the above problems,

收稿日期: 2021-06-24; 修回日期: 2021-08-28

通信作者: 傅启明, fqm_1@126.com

基金项目: 国家重点研发计划基金资助项目(No.2020YFC2006602); 国家自然科学基金资助项目(No.62072324, No.61876217, No.61876121, No.61772357); 江苏省重点研发计划基金资助项目 (No.BE2020026)

Foundation Items: The National Key Research and Development Program of China (No.2020YFC2006602), The National Natural Science Foundation of China (No. 62072324, No.61876217, No.61876121, No.61772357), The Research and Development Program of Jiangsu Province (No.BE2020026)

combined with EnergyPlus, actual system parameters and historical data, the HVAC optimized control model was constructed, and an improved double pools-based DQN (DPs-DQN) algorithm was proposed. Finally, it was applied to the load distribution of different types of chillers, the combined optimal control of cooling tower fan frequency and cooling water pump frequency in HVAC system. Based on the constructed problem model, aiming at the problem of sample imbalance in the decision-making optimization process, the algorithm established two independent experience pools on the basis of DQN to store load distribution and non load distribution samples respectively. During the training process, followed a certain ratio to sample from the experience pool to speed up the algorithm convergence. The proposed method was compared with the model-based control method and the baseline method. The experimental results show that compared with the baseline method, the model-based HVAC controller can save 11.5% (optimal energy-saving efficiency), while the DPs-DQN can save energy by 7.5% in the first year. At the same time, as the system runs, the controller can obtain results close to the optimal energy saving efficiency in the eighth year. In addition, compared with the model-based HVAC controller, the controller does not depend on the system model, and requires less prior knowledge and sensors in the online control process, which is more valuable in actual engineering applications.

Key words: deep reinforcement learning, model-free optimal control, HVAC system, building energy saving

0 引言

据统计,在建筑物总能耗中,HVAC (heating, ventilation and air conditioning) 系统能耗占比高达甚至超过 50%^[1-5],其中冷机的能耗又是 HVAC 能耗的重要组成部分,因此,冷机的优化控制对降低整个 HVAC 系统能耗甚至建筑总能耗尤为重要。目前,优化不同冷机负荷量 (optimal chiller loading, OCL) 是对冷机优化控制的主要方法之一。该方法主要针对总系统负荷分配方式进行优化,使得不同冷机的部分负荷率 (partial load ratio, PLR) 更加合理,从而实现节能。此外,冷却水循环系统也是 HVAC 的重要组成部分,主要包含冷却塔、冷却水泵等,其旨在释放冷机所排出的热量,同时该系统运行对冷机的性能系数 (coefficient of performance, COP) 也存在较大影响。因此,针对 HVAC 系统的优化控制方法需要同时考虑冷机及冷却水循环系统,实现两者的组合优化。

在当前针对 HVAC 系统的优化实践中,较优的控制方法主要包含基于规则的控制、基于模型的控制以及无模型的控制等。基于规则的控制往往是静态的,控制规则根据工程师以及设施管理员的经验确定。基于模型的控制需要大量的历史数据和传感器信息,以建立精确的 HVAC 模型,但是该方法通常缺乏较好的鲁棒性,同时并不适用于缺乏历史数据以及传感器的旧建筑群。典型的基于模型的控制方法通常包含以下步骤:步骤一是通过历史数据以及其他先验知识建立准确的系统模型;步骤二是获取系统的实时检测数据;步骤三是使用最优算法

确定最优的控制动作。Ma Z J 等人^[6]提出基于模型的在线故障检测和最优控制方法,该方法对冷凝器和冷却塔的性能进行简化,并将系统的冷负荷以及室外湿球温度作为实时的环境变量,通过一个混合的快速搜索方法寻找最优控制策略,以最大化系统性能系数。Yao Y 等人^[7]通过住宅楼历史数据和经验关系建立系统最佳控制模型,在此基础上,为了达到最大的性能,研究在非控制变量 (如室外湿球温度和冷冻水回水温度) 下,针对系统其他变量 (如冷冻水和冷却水流量以及冷冻水出水温度等) 的优化问题。Huang S 等人^[8]对包含多个冷机、冷却塔以及恒速的冷却水泵的系统建立仿真模型,在不改变结构的基础上,能够对冷却水回水温度 T_{cwr} 进行优化,从而最小化冷机和冷却塔的总能耗。

系统具体参数和监测数据是基于模型预测控制方法的基础,其中数据的准确性会直接影响模型的准确性。因此,基于模型预测控制方法存在以下不足。一是对先验知识的依赖。建立模型需要现场测试数据、设备运行的历史数据或来自制造商的数据等。现场测试数据需要测试设备以及人工参与;历史数据需要目标系统中具有充分的传感器,其中实时检测的数据类型与数量直接决定目标系统中所需传感器的种类和数量,进而直接影响基于模型控制预测方法的实用性。二是模型不确定性的风险。模型的不确定性直接影响基于模型控制器的性能。Zhu N 等人^[9]将模型的不确定性分成两类:模型结构错误和模型参数错误。模型简化是导致模型结构错误的主要原因;历史数据错误会导致模型参数错误,因为模型参数通常是由历史数据回归拟合

而成的。除此之外,即使建立的初始模型相对准确,但是其准确性也会随着设备的老化逐年下降。参考文献[10-11]的研究表明,模型简化所导致的计算误差不可忽视,而且通常会带来比较严重的影响。

为了避免建立准确数学模型的过程,有学者将研究聚焦于无模型的控制方法。Qiu S N 等人^[12]提出用 Q-learning 方法对冷却侧的冷却水泵频率和冷却塔风机频率进行优化。作者将系统冷负荷和室外湿球温度分别以 100 kW 和 1℃ 为单位进行离散化,并组合成环境状态,但是该方法并未考虑系统负荷分配对节能效果的影响。Baghaee S 等人^[13]使用 Q-learning 同时调控室内 CO₂ 浓度以及 HVAC 系统。Faddel S 等人^[14]提出一种数据驱动的 Q-learning 方法,在保证室内人员的舒适性的同时最小化建筑能耗。

传统的 Q-learning 方法需要对状态动作进行离散化,这导致算法泛化能力下降,无法针对复杂问题进行求解^[15-16]。近年来,Mnih V 等人^[17]提出深度 Q 网络(deep Q network, DQN)算法,该算法结合深度神经网络和 Q-learning 方法,能够对高维状态空间进行函数逼近,实现端到端的决策控制,实验证明,在 Atari 游戏控制上,其能够获得接近甚至超过人类专家的决策能力。Ahn K U 等人^[18]利用 EnergyPlus 对 HVAC 系统进行建模,并使用 DQN 算法,以 CO₂ 浓度、室内温度、冷却水温度等 54 种实时数据为状态,对冷冻水温度、冷却水温度以及 AHU (air handle unit) 开放率进行优化控制,进而最小化系统总能耗。该算法经过 15 天的训练,能达到 15.7% 的节能效果(与基线(baseline)方法相比),但其并未考虑大规模动作空间导致算法收敛较慢问题,以及冷却水泵、冷却塔风机频率和不同型号冷机的负荷分配对 HVAC 能耗的影响。

本文提出双池 DQN (double-pools-based DQN, DPs-DQN) 算法,以上海市某办公楼的实际 HVAC 系统为案例,对冷却水循环系统中冷却水泵、冷却塔以及不同型号冷机的负荷分配进行无模型优化控制。整个案例系统包含两种不同型号的冷机、3 组相同型号的冷却水泵和冷却塔。结合 EnergyPlus 中的参数,利用 Python 对该案例进行仿真建模,仿真实验结果表明,双池 DQN 在 6 个月数据基础上能够学习到较优的

节能策略。

1 相关理论

1.1 马尔可夫决策过程

强化学习问题通常可以被建模为马尔可夫决策过程(Markov decision process, MDP),由四元组表示: $M = \langle S, A, R, P \rangle$, 其中 S 代表所有的状态集合, A 代表所有的动作集合, R 代表奖赏函数, P 代表状态迁移函数。如果环境的下一个状态依赖且仅依赖当前状态和当前采取的动作,则这样的过程就具有马尔可夫性质。对于一个 MDP 问题来说,在每一个时间步 t , 智能体(agent) 根据环境状态 s_t , 采取一个动作 a_t 并且获得立即奖赏 r_t , 然后获取下一个状态 s_{t+1} 。强化学习算法的目的是求解一个最优策略,并利用这个最优策略进行决策。在强化学习中, π 表示策略, $\pi(s, a)$ 表示在状态 s 下选择动作 a 的概率。如果 π 是一个确定的策略,在任意状态 $s \in S$, $\pi(s, a)$ 表示状态 s 下所选择的动作 a 。强化学习利用值函数来评估策略 π 的优劣,值函数包含状态值函数 V^π 和动作值函数 Q^π , 其中 $V^\pi(s)$ 表示在状态 s 下由策略 π 得到的累计期望奖赏, $Q^\pi(s, a)$ 表示状态动作对 (s, a) 由策略 π 得到的累计期望奖赏。通常采用动作值函数来评估策略 π 的优劣,即:

$$Q^\pi(s, a) = r(s, a) + \gamma \sum_{s' \in S} p(s'|s, a) \sum_{a' \in A} \pi(s', a') Q^\pi(s', a') \quad (1)$$

其中, $r(s, a)$ 表示在状态 s 下选取动作 a 获得的立即奖赏, $\gamma \in (0, 1)$ 表示折扣因子, $p(s'|s, a)$ 表示在状态 s 下选取动作 a 转移到下一个状态 s' 的概率,式(1)也被称为贝尔曼(Bellman)方程。强化学习的目的是获得最大化累计奖赏的最优策略 π^* , 对应的 $Q^*(s, a)$ 可以表示为:

$$Q^*(s, a) = r(s, a) + \gamma \sum_{s' \in S} p(s'|s, a) \left\{ \max_{a' \in A} Q^*(s', a') \right\} \quad (2)$$

式(2)也被称作最优 Bellman 方程。

1.2 深度强化学习

因为内存和表格计算的开销, Q-learning 方法不适用于具有大规模状态空间以及连续状态空间的问题。针对上述问题,深度强化学习方法(如 DQN)使用函数近似器(深度神经网络)来代替 Q 值函数表格: $Q(s, a) = Q(s, a; \theta)$, 其中 θ 是神经网络的参数,通过优化 θ 来近似 Bellman 方程(式(1))。DQN 可以使用少量 θ 来拟合 $Q(s, a)$, 这不仅可以降低内存和表格

计算开销,同时具有强大的泛化能力。

为了解决基于离策略(off-policy)的强化学习方法收敛性弱的问题,DQN 采用经验回放机制以及独立目标网络的方式。DQN 通过建立经验池来存储训练样本 (s_t, a_t, r_t, s_{t+1}) ,同时采用随机采样方式进行网络训练,从而打破强化学习中样本的关联性。在此基础上,为了使收敛更加稳定,DQN 使用目标网络 $\hat{Q}(s, a; \theta^-)$ 生成目标值来更新 $Q(s, a; \theta)$ 中的 θ ,而 $Q(s, a; \theta)$ 在用于实时更新的同时提供最优动作。基于所构建的两个网络,DQN 通过最小化目标函数 $L_t(\theta_t)$ 进行训练,在一定的时间步数 C 之后更新 $\theta^- = \theta$,如式(3)所示:

$$L_t(\theta_t) = E \left[\left(r + \gamma \max_{a'} \hat{Q}(s', a'; \theta^-) - Q(s, a; \theta) \right)^2 \right] \quad (3)$$

2 问题建模

2.1 系统工作流程

系统的工作流程如图1所示。为了在保护冷机的同时降低不必要的制冷量,系统采用基于实时系统冷负荷值确定各台冷机开关状态的序列控制方法,因此,系统运行主过程主要包含冷机运行序列控制以及双池 DQN 优化控制两方面,具体步骤如下。

第一步:在当前时间步 t ,根据实时系统冷负荷进行冷机启停控制,见图1右侧。在满足系统制

冷需求的前提下,为了避免系统供冷能力的浪费,本文始终采用较小的总制冷能力来启动机组。当一个时间步开始时,若系统的冷负荷低于小冷机额定制冷量的40%,则关闭整个系统冷机以保护冷机^[19];若系统冷负荷超过小冷机额定制冷量的40%且不高于一台小冷机的额定制冷量,则仅开启一台小冷机;若系统冷负荷低于一台大冷机额定制冷量且高于一台小冷机的额定制冷量,则只需要开启其中一台大冷机;若系统冷负荷高于一台大冷机且低于一台大冷机和小冷机的额定制冷量之和,则开启一台大冷机和一台小冷机;若系统冷负荷高于两台大冷机的额定制冷量,就开启系统中的所有冷机,否则仅开启两台大冷机^[20]。冷却水泵、冷却塔以及冷冻水泵的开关状态与对应冷机的开关状态相同。

第二步:观测环境状态 s_t 。记录系统实时冷负荷、室外湿球温度等数据,并将这些数据传送给无模型的深度强化学习算法。将 (s_t, a_t, r_t, s_{t+1}) 存储至经验池,并从经验池中随机采样,执行算法训练更新网络参数。

第三步:无模型方法给定控制动作 a_t 。为了保证大小冷机的PLR值在 $[0.4, 1]$ ^[21-23],需要根据系统冷负荷的值确定合理的负荷分配比值(ratio)区间,并在该区间中选取优化控制动作 a_t 。

第四步:系统执行控制动作。根据ratio计算冷

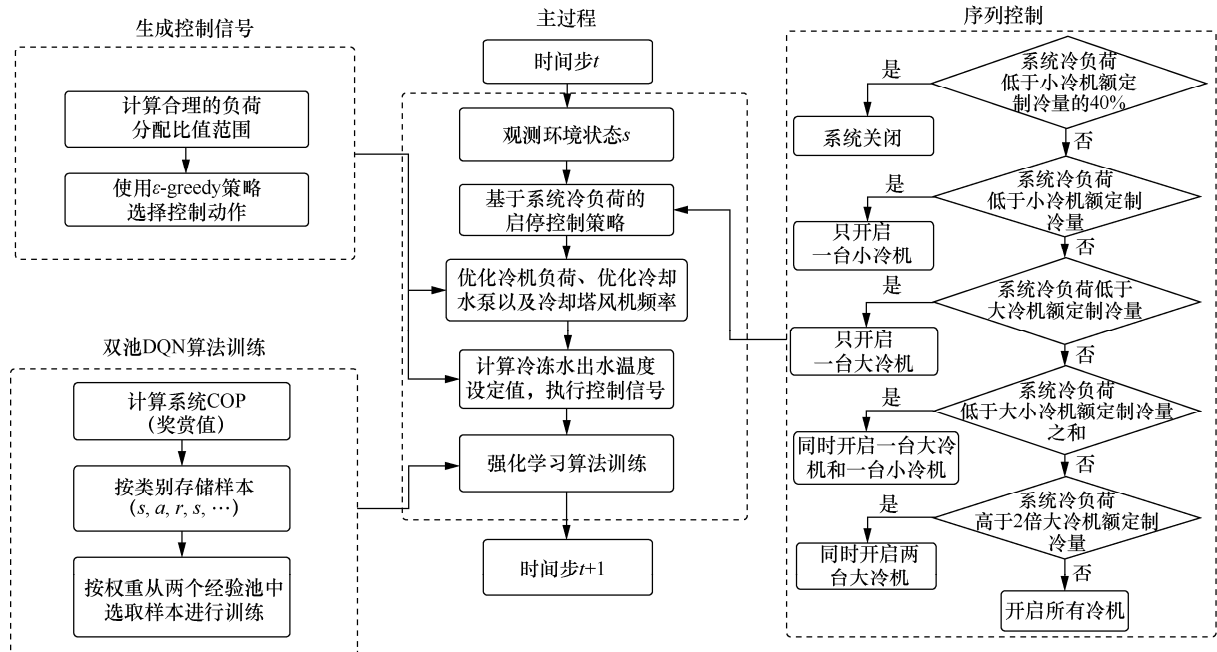


图1 系统的工作流程

冻水出水温度的设定值 $T_{\text{chws,set}}$ ，执行 3 个控制信号： $T_{\text{chws,set}}$ 、 f_{pump} 、 f_{tower} ，并获得下一个环境状态 s_{t+1} 。计算当前动作下系统的 COP 值，并将其作为强化学习算法中的奖赏值 r_t 。

第五步：算法训练，执行参数更新。将样本 (s_t, a_t, r_t, s_{t+1}) 存储至经验池，并从经验池中随机采样，执行算法训练，更新网络参数。

第六步：结束当前时间步 t 。开始下一个时间步 $t+1$ 。

2.2 模型特征与约束

(1) 优化目标

系统性能参数 COP 直接反映系统的节能性能，因此提高系统 COP 是该优化问题的目标，其计算方式如式 (4) 所示，其中系统的冷负荷的计算方式如式 (5) 所示：

$$\text{COP} = \frac{\text{CL}_s}{\sum P_{\text{chiller},i} + P_{\text{tower},i} + P_{\text{pump},i}} \quad (4)$$

$$\text{CL}_s = C_p \times \rho \times F_{\text{chw}} \times (T_{\text{chwr}} - T_{\text{chws}}) \div 3\,600 \quad (5)$$

其中， CL_s 是系统的冷负荷 (kW)， $P_{\text{chiller},i}$ 、 $P_{\text{tower},i}$ 和 $P_{\text{pump},i}$ 分别是第 i 台冷机、冷却塔和冷却水泵的功率 (kW)， C_p 是水的比热容 (4.2 kJ/(kg·K))， F_{chw} 是系统冷冻水流量 (m^3/h)， ρ 是水的密度 ($1\,000 \text{ kg}/\text{m}^3$)， T_{chwr} 、 T_{chws} 分别是冷冻水回水和出水温度 ($^{\circ}\text{C}$)。

(2) 先验知识和实时数据

优化控制前，问题模型提供系统的布局以及设备的铭牌 (例如大小冷机数量)。优化控制时，问题模型提供实时系统冷负荷、室外湿球温度。与基于模型的方法相比，无模型的控制方法并不需要系统组件 (冷机、冷却塔以及冷却水泵等) 的性能模型。

(3) 约束条件

系统包含 2 种规格的冷机、3 台冷却塔和 3 台冷却水泵，其中所有冷却塔和冷却水泵规格相同 (见表 1)；每台冷机与 1 台冷却水泵和冷却塔对应，当该冷机开启时，其对应的冷却水泵和冷却塔开启；冷冻水泵以额定功率运行，每台冷机的冷冻水流量恒定；相同规格冷机的性能相同，即相同冷机分得的冷负荷值相同，同时本方法依旧适用于包含 $n(n \geq 3)$ 个相同规格的小冷机和 $m(m \geq 3)$ 个相同规格的大冷机的系统，虽然相同型号冷机的负荷分配也应该优化控制，但 Chang Y C^[19] 的计算结果表明，最优的分配方式就是相同规格冷机分配相同的负荷；系统负荷分配方式无法直接对冷机进行控制，因此将负荷分配方式转换成冷机的冷冻水出水温度 ($T_{\text{chws},j,\text{set}}$)，从而实现对冷机的控制^[24]，如式 (6) 所示，其中 $F_{\text{chw},j}$ 是第 j 台冷机的冷冻水流量 (m^3/h)， ρ 是水的密度 ($1\,000 \text{ kg}/\text{m}^3$)， $T_{\text{chwr},j}$ 是第 j 台冷机额定冷冻水出水温度， CL_j 是分配给第 j 台冷机的冷负荷量。

$$T_{\text{chws},j,\text{set}} = T_{\text{chwr},j,\text{ref}} - \frac{\text{CL}_j}{C_p \times \rho \times F_{\text{chw},j} \div 3\,600} \quad (6)$$

3 基于双池 DQN 的 HVAC 控制方法

本节将阐述双池 DQN 算法的详细应用方法，包含状态空间、动作空间和奖赏的设置方法，以及对于制冷季节 6 个月中负荷分配样本数据不均衡导致算法收敛较慢问题的解决方法。

3.1 MDP 建模

(1) 动作空间设置

本文以大冷机的 $\text{ratio}_{\text{bigchillers}}$ 、 f_{tower} 和 f_{pump} 为控制变量 $\left(\text{ratio}_{\text{bigchillers}} = \frac{\text{CL}_{\text{bigchillers}}}{\text{CL}_{\text{system}}} \right)$ 。根据系统冷负荷分配给所有大冷机的 ratio 值以及系统冷机开合状

表 1 制造商提供的设备特性

设备	数量	名义参数
冷机	2 台大冷机	大冷机：功率 504 kW，制冷能力 2 810 kW，冷冻水出水温度 6 $^{\circ}\text{C}$ ，冷冻水回水温度 12 $^{\circ}\text{C}$
	1 台小冷机	小冷机：功率 314 kW，制冷能力 1 760 kW，冷冻水出水温度 6 $^{\circ}\text{C}$ ，冷冻水回水温度 12 $^{\circ}\text{C}$
冷冻水泵	2 台大型冷冻水泵	大型冷冻水泵：功率 30 kW，水流量 403 m^3/h ，扬程 19 m，恒速
	1 台小型冷冻水泵	小型冷冻水泵：功率 15 kW，水流量 252 m^3/h ，扬程 15 m，恒速
冷却水泵	3 台大冷却水泵	大冷却水泵：功率 75 kW，水流量 581 m^3/h
冷却塔	3 台大冷却塔	大冷却塔：功率 30 kW，水流量 569 m^3/h

态（由序列控制产生的启停控制信号），能够分别计算出每台大冷机以及小冷机的负荷量，因此以大冷机的 $\text{ratio}_{\text{bigchillers}}$ 为控制变量足以进行负荷分配计算。为避免 agent 因探索动作而损坏设备，算法控制 f_{tower}^k 和 f_{pump}^k 的取值区间为 $[25, 50]$ Hz，步长为 1 Hz， $\text{ratio}_{\text{bigchillers}}$ 的取值区间为 $[0.1, 1]$ ，步长为 0.1，同时为保护冷机，需要控制各台冷机的 PLR 值在 $[0.4, 1]$ ，因此，在算法做决策时，需要分别计算大小冷机的 PLR 值，以确定合理的 $\text{ratio}_{\text{bigchillers}}$ 控制区间。在实际控制中，按照 $[\text{ratio}_{\text{bigchillers}}, f_{\text{tower}}, f_{\text{pump}}]$ 的顺序对每一个设定点空间中所有值进行排列，组合成动作空间（如 $[[\text{ratio}_{\text{bigchillers}} = 0.9, f_{\text{tower}} = 50, f_{\text{pump}} = 50], [\text{ratio}_{\text{bigchillers}} = 0.9, f_{\text{tower}} = 50, f_{\text{pump}} = 49], \dots]$ ）。虽然按该方式处理时动作数量达到 6 760 个，但当 agent 做决策时，较小的 PLR 取值范围能够减小 $\text{ratio}_{\text{bigchillers}}$ 的合理控制区间（例如本文根据中大小冷机制冷性能， $\text{ratio}_{\text{bigchillers}}$ 的合理取值始终大于 0.2），同时，在 6 个月的制冷季节中，较多时间无须进行负荷分配控制，为了进一步减小动作空间，此时只需要在 $\text{ratio}_{\text{bigchillers}} = 0.1$ 对应的动作空间里面选择 f_{tower}^k 和 f_{pump}^k 控制动作，以此合理调整动作选择空间，从而提高算法收敛速度。

（2）状态空间设置

考虑到室外湿球温度对冷却水循环系统的散热性能影响较大和不同的系统冷负荷下负荷分配策略不同，将系统的冷负荷 CL_s 和 T_{wet} 组合作为算法的状态输入，如式（7）所示：

$$S_t = [\text{CL}_s^t, T_{\text{wet}}^t] \quad (7)$$

其中， CL_s^t 是 t 时刻的系统冷负荷， T_{wet}^t 是 t 时刻的室外湿球温度。

（3）奖赏设置

为了在满足制冷量的同时最小化每个设备的功耗总和，直接将系统 COP 作为奖赏函数，其计算式如下：

$$r_{t+1} = \text{COP} = \frac{\text{CL}_s}{\sum (P_{\text{chiller}}^k + P_{\text{tower}}^k + P_{\text{pump}}^k)} \quad (8)$$

其中， CL_s 代表系统冷负荷， P_{chiller}^k 、 P_{tower}^k 、 P_{pump}^k 分别代表第 k 台冷机、冷却塔和冷却水泵的功率。

3.2 双池 DQN 算法

只有当至少有一台大冷机和小冷机同时开启

时才需要进行负荷分配，但只要有至少一台冷机开启就会有一台冷却塔和冷却水泵同时开启。在 6 个月的制冷季节的大部分时间中，系统冷负荷的值超过一台大冷机和小冷机的制冷量之和或者超过两台大冷机制冷量之和的时间段较少，即较多时刻并不需要同时开启大小冷机。

这导致进行负荷分配优化的时间段远少于冷却塔和冷却水泵的优化时间段，因此在优化过程中，关于负荷分配优化的样本量将远少于关于冷却塔和冷却水泵优化的样本量，即学习过程中的样本不均衡。样本不均衡问题会导致传统的 DQN^[17] 算法收敛较慢，针对该问题，本文提出双池 DQN 算法。该算法采用两个经验池分别存储需要负荷分配的样本和不需要负荷分配的样本，同时在算法训练阶段的每个时间步中，分别从两个经验池选取固定批次（batch）个样本进行组合，进而对算法进行训练，见算法 1。

算法 1 双池 DQN 算法

参数：

- 1: N_1 、 N_2 // 分别存储负荷分配样本以及非负荷分配样本的经验池容量
- 2: $\varepsilon_{\min}$ 、 $\Delta\varepsilon$ // $\varepsilon_{\min}$ 为 ε 的最小值， $\Delta\varepsilon$ 为 ε 的线性衰减量
- 3: batch // 随机采样的批次大小
- 4: γ 、 C_{step} // γ 为学习衰减因子， C_{step} 为强化学习更新参数的间隔步数
- 5: 初始化容量为 N_1 的经验池 D_1 以及容量为 N_2 的经验池 D_2
- 6: 初始化动作值函数 $Q(s, a; \theta)$ ，随机化生成参数 θ
- 7: 初始化目标值函数 $\hat{Q}(s, a; \theta^-)$ ， $\theta^- = \theta$
- 8: **for** episode=1, M **do**
- 9: 观测状态 s_1
- 10: **for** $t=1, T$ **do**
- 11: **if** operate load sharing:
- 12: 根据 PLR 范围计算合理的 $\text{ratio}_{\text{bigchillers}}$ 区间： $[\text{ratio}_{\text{start}}, \text{ratio}_{\text{end}}]$
- 13: 使用 ε -greedy 选取合理的动作 a_t ，保证 a_t 中 $\text{ratio}_{\text{bigchillers}} \in [\text{ratio}_{\text{start}}, \text{ratio}_{\text{end}}]$
- 14: **else**: 直接使用 ε -greedy 选取 a_t

15: $\varepsilon = \max(\varepsilon - \Delta\varepsilon, \varepsilon_{\min})$ //线性 ε 衰减
 16: 执行动作 a_t 并观测奖赏 r_t 、状态 s_{t+1}
 17: **if** operate load sharing:
 18: 存储样本 (s_t, a_t, r_t, s_{t+1}) 至经验池 D_1
 19: **else:**
 20: 存储样本 (s_t, a_t, r_t, s_{t+1}) 至经验池 D_2
 21: 分别从 D_1 和 D_2 中随机选取 minibatch
 个样本并组合成一个样本批次
 (s_j, a_j, r_j, s_{j+1})
 22: **et** $y_j = \begin{cases} r_j, \text{情节结束} \\ r_j + \gamma \max_{a'} \hat{Q}(s_{j+1}, a'; \theta^-), \text{其他} \end{cases}$
 23: 计算目标函数 $(y_j - Q(s_j, a_j; \theta))^2$, 并
 进行梯度下降更新 θ
 24: 每隔 C_{step} 个时间步之后, 设置 $\theta^- = \theta$

25: 设置 $s_{t+1} = s_t$
 26: **end for**
 27: **end for**

4 仿真环境搭建

4.1 案例系统

案例以上海市某办公楼的实际 HVAC 系统为基础。制造商提供的设备特性见表 1。整个案例系统的工作流程如图 1 所示, 其中冷机的开启数量按照当前系统冷负荷确定。案例系统的简化结构如图 2 所示, 其中冷却水循环系统是一个解耦系统。

案例系统的冷负荷数据由 EnergyPlus^[25-27] 仿真生成, 如图 3 所示, 横坐标代表制冷季节 6 个月的

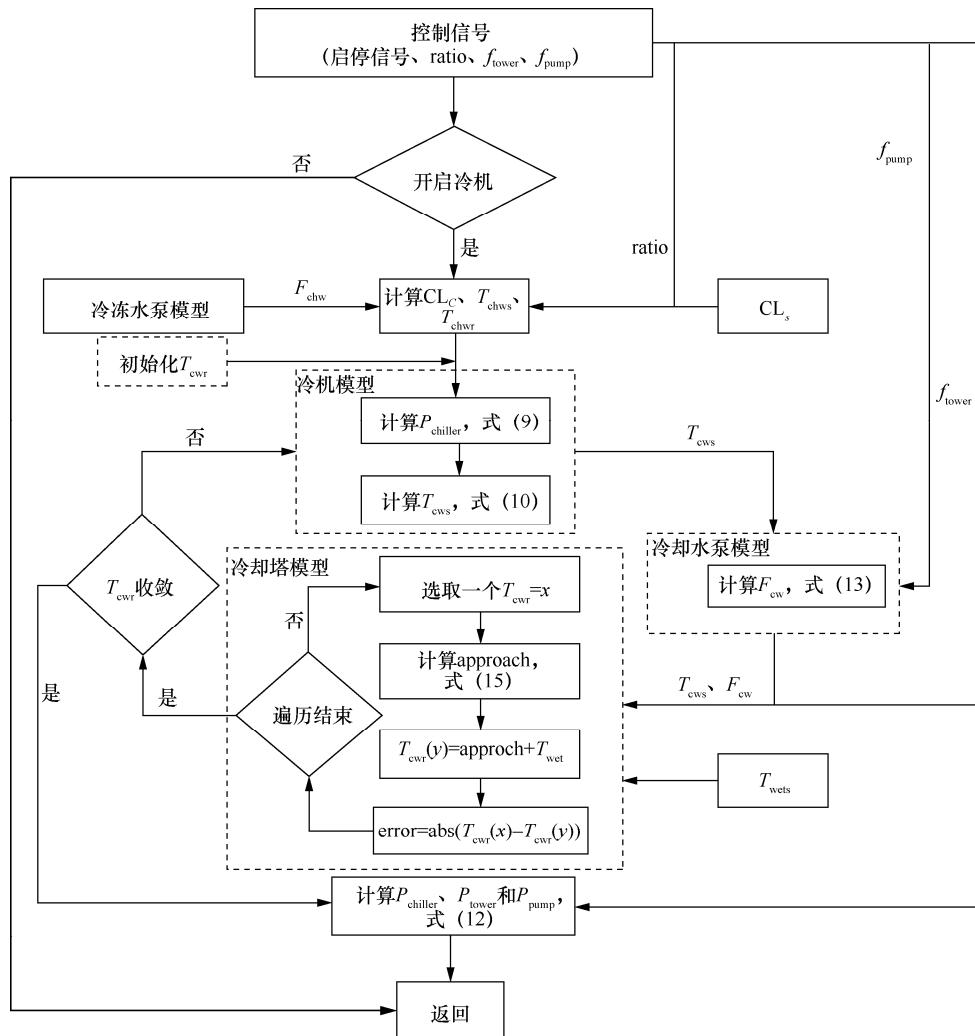


图2 案例系统的简化结构

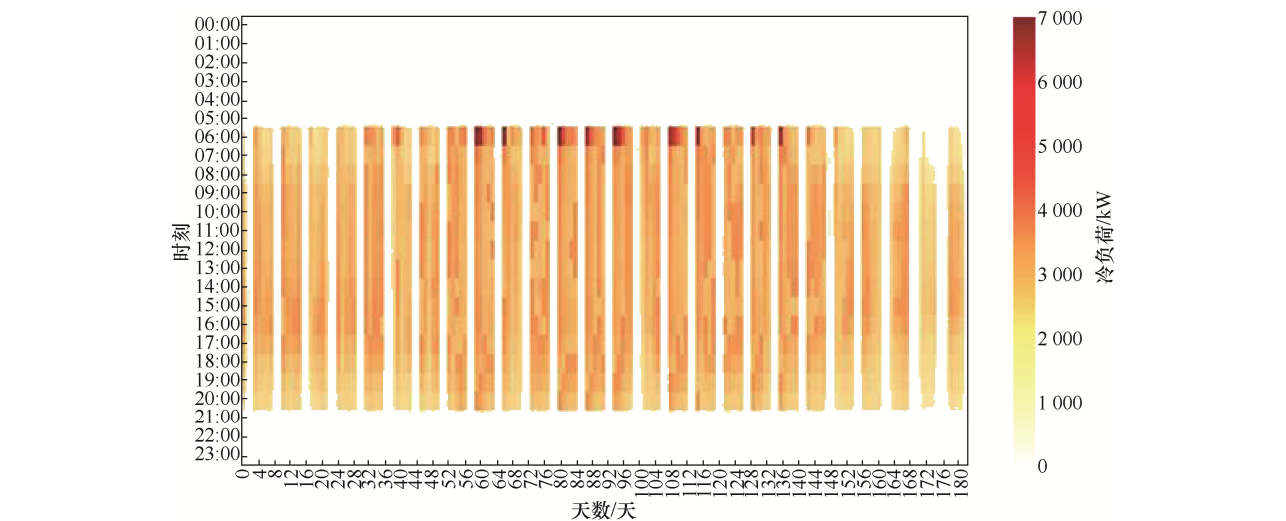


图3 案例系统冷负荷数据

天数，纵坐标表示每天 24 个时刻，图中的颜色代表不同时间段的系统冷负荷大小，其中空白的区域代表周末的能耗为 0。将实测天气数据作为案例系统的输入，室外湿球温度如图 4 所示，其中横坐标为采集的次数，采集间隔为 1 h。

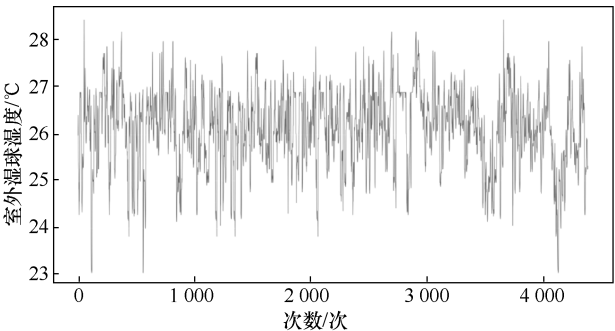


图4 室外湿球温度

4.2 仿真平台搭建

式 (9) ~ 式 (11) 代表冷机模型，其中式 (9) 采用 EnergyPlus 以及 DOE-2.1 中的标准冷机模型。式 (12) ~ 式 (13) 代表冷却水泵模型，其中 a 、 b 、 c 是设备特定系数，该系数由具体项目中的监测数据回归生成。式 (15) 代表变速冷却塔模型，该模型由 CoolTool 开发并被 EnergyPlus 采用，同时其遵循一定约束条件，具体见表 2。考虑到冷却水系统中的管道阻力不会显著变化，为了简化模型，式 (15) 中 FR_{water} 设置为 $F_{nominal}/f_{pump}$ ，其中， T_{cwr} 是冷却水回水温度 (°C)， T_{cws} 是冷却水出水温度 (°C)， P 是功率， f 是频率设定值， P_{ref} 是额定功率， F_{chw} 是冷冻水流量 (m³/h)， CL_c 是冷机的冷负荷 (kW)， F_{cw} 是冷却水流量 (m³/h)，

$F_{nominal,pump}$ 是冷却水泵额定水流量 (m³/h)， T_{wet} 是室外湿球温度 (°C)， $approach$ 是 T_{cwr} 和 T_{wet} 之间的临近温度， FR_{air} 是空气流量比 (实际的空气流量与额定的空气流量的比)， FR_{water} 是水流量比 (实际的水流量比额定的水流量)， Tr 是冷却塔温度变化的范围 (°C) (即 $T_{cws} - T_{cwr}$)， T_{wb} 是输入的空气湿球温度 (°C) (即 T_{wet})， b_i 、 d_i 、 g_i 以及 $Coeff(i)$ 都是冷机模型固定的系数，该系数来源于 EnergyPlus V9.2、Engineering reference 文件。冷机模型系数见表 3， $Coeff$ 系数见表 4。

表 2 临近温度相关系数的最大最小值限制

独立变量限制	值
输入最小室外湿球温度/℃	-1
输入最大室外湿球温度/℃	26.7
冷却塔温度范围最小值/℃	1.1
冷却塔温度范围最大值/℃	11.1
冷却塔临近温度最小值/℃	1.1
冷却塔临近温度最大值/℃	11.1
最小冷却水流量比	0.75
最大冷却水流量比	1.25

图 2 给出了冷机、冷却塔与冷却水泵三者仿真结构。 T_{cws} 和 T_{cwr} 是两个耦合变量， T_{cws} 影响冷却水回水温度 T_{cwr} ，同时 T_{cwr} 影响冷机性能，进而影响 T_{cws} ，因此， T_{cws} 和 T_{cwr} 值需要通过在 3 个设备模型大循环中不断迭代直到 T_{cwr} 收敛才能确定。收敛的判断条件是 T_{cwr} 值误差不超过 0.1，如若超过 50 个

表3

冷机模型系数

系数	大冷机 1# 和大冷机 2#			小冷机 3#		
	DOE-2.1 generic centrifugal chiller (行号: 11 170)			York VSD centrifugal chiller (行号: 5 684)		
	ChillerCapFTemp	ChillerEIRFTemp	ChillerEIRFPLR	ChillerCapFTemp	ChillerEIRFTemp	ChillerEIRFPLR
b1、d1、g1	0.257896	0.933884	0.222903	0.4575085	0.6794525	0.07859908
b2、d2、g2	0.0389016	-0.058212	0.313387	0.1313508	0.06694756	0.1950291
b3、d3、g3	-0.00021708	0.00450036	0.463710	-0.004408831	-0.003625396	0.7241581
b4、d4	0.0468684	0.00243	—	0.01930354	-0.01018762	—
b5、d5	-0.00094284	0.000486	—	-0.0005479641	0.001066394	—
b6、d6	-0.00034344	-0.001215	—	-0.001376580	-0.00211342	—

注: 根据 EnergyPlus V9.2 Engineering reference.pdf 和 Datasets/Chillers.idf

表4

Coeff 系数

系数序号	系数值	系数序号	系数值	系数序号	系数值
Coeff(1)	0.52049709836241	Coeff(2)	-10.617046395344	Coeff(3)	10.7292974722538
Coeff(4)	-2.74988377158227	Coeff(5)	4.73629943913743	Coeff(6)	-8.25759700874711
Coeff(7)	1.57640938114136	Coeff(8)	6.51119643791324	Coeff(9)	1.50433525206692
Coeff(10)	-3.2888529287801	Coeff(11)	0.0257786145353773	Coeff(12)	0.182464289315254
Coeff(13)	-0.0818947291400898	Coeff(14)	-0.215010003996285	Coeff(15)	0.0186741309635284
Coeff(16)	0.0536824177590012	Coeff(17)	-0.00270968955115031	Coeff(18)	0.00112277498589279
Coeff(19)	-0.00127758497497718	Coeff(20)	0.0000760420796601607	Coeff(21)	1.43600088336017
Coeff(22)	-0.5198695909109	Coeff(23)	0.117339576910507	Coeff(24)	1.50492810819924
Coeff(25)	-0.135898905926974	Coeff(26)	-0.152577581866506	Coeff(27)	-0.0533843828114562
Coeff(28)	0.00493294869565511	Coeff(29)	-0.00796260394174197	Coeff(30)	0.000222619828621544
Coeff(31)	-0.0543952001568055	Coeff(32)	0.00474266879161693	Coeff(33)	-0.0185854671815598
Coeff(34)	0.00115667701293848	Coeff(35)	0.000807370664460284		

循环都未收敛 (极少数情况), 则选取最后一个值作为 T_{cwr} 。整个系统的 T_{chws} 、 T_{chwr} 以及 F_{chw} 不会在模型中进行迭代计算。

$$\begin{cases} \text{ChillerCapFTemp} = b_1 + b_2 T_{\text{chws}} + b_3 T_{\text{chws}}^2 + \\ b_4 T_{\text{cwr}} + b_5 T_{\text{cwr}}^2 + b_6 T_{\text{cwr}} T_{\text{chws}} \\ \text{ChillerEIRFTemp} = d_1 + d_2 T_{\text{chws}} + d_3 T_{\text{chws}}^2 + \\ d_4 T_{\text{cwr}} + d_5 T_{\text{cwr}}^2 + d_6 T_{\text{cwr}} T_{\text{chws}} \\ \text{ChillerEIRFPLR} = g_1 + g_2 \text{PLR} + g_3 \text{PLR}^2 \\ P_{\text{chiller}} = P_{\text{ref}} \times \text{ChillerCapFTemp} \times \\ \text{ChillerEIRFTemp} \times \text{ChillerEIRFPLR} \end{cases} \quad (9)$$

$$T_{\text{cws}} = T_{\text{cwr}} + (P_{\text{chiller}} + \text{CL}_c) \div \left(\frac{c_p \times F_{\text{cw}} \times \rho}{3\ 600} \right) \quad (10)$$

$$T_{\text{chwr}} = T_{\text{chws}} + \text{CL}_c \div \left(\frac{c_p \times F_{\text{cw}} \times \rho}{3\ 600} \right) \quad (11)$$

$$P = a f^2 + b f + c \quad (12)$$

$$F_{\text{cw}} = \frac{f_{\text{pump}}}{f_{\text{nominal}}} \times F_{\text{nominal}} \quad (13)$$

$$\text{approach} = T_{\text{cwr}} - T_{\text{wet}} \quad (14)$$

$$\begin{aligned} \text{approach} = & \text{Coeff}(1) + \\ & \text{Coeff}(2) \cdot \text{FRair} + \\ & \text{Coeff}(3) \cdot (\text{FRair})^2 + \\ & \text{Coeff}(4) \cdot (\text{FRair})^3 + \\ & \text{Coeff}(5) \cdot \text{FRwater} + \\ & \text{Coeff}(6) \cdot \text{FRair} \cdot \text{FRwater} + \\ & \text{Coeff}(7) \cdot (\text{FRair})^2 \cdot \text{FRwater} + \\ & \text{Coeff}(8) \cdot (\text{FRwater})^2 + \\ & \text{Coeff}(9) \cdot \text{FRair} \cdot (\text{FRwater})^2 + \end{aligned}$$

$$\begin{aligned}
& \text{Coeff}(10) \cdot (\text{FRwater})^3 + \\
& \text{Coeff}(11) \cdot T_{wb} + \\
& \text{Coeff}(12) \cdot \text{FRair} \cdot T_{wb} + \\
& \text{Coeff}(13) \cdot (\text{FRair})^2 \cdot T_{wb} + \\
& \text{Coeff}(14) \cdot \text{FRwater} \cdot T_{wb} + \\
& \text{Coeff}(15) \cdot \text{FRair} \cdot \text{FRwater} \cdot T_{wb} + \\
& \text{Coeff}(16) \cdot (\text{FRwater})^2 \cdot T_{wb} + \\
& \text{Coeff}(17) \cdot (T_{wb})^2 + \\
& \text{Coff}(18) \cdot \text{FRair} \cdot (T_{wb})^2 + \\
& \text{Coff}(19) \cdot \text{FRwater} \cdot (T_{wb})^2 + \\
& \text{Coff}(20) \cdot (T_{wb})^3 + \\
& \text{Coff}(21) \cdot \text{Tr} + \\
& \text{Coeff}(22) \cdot \text{FRair} \cdot \text{Tr} + \\
& \text{Coeff}(23) \cdot (\text{FRair})^2 \cdot \text{Tr} + \\
& \text{Coeff}(24) \cdot \text{FRwater} \cdot \text{Tr} + \\
& \text{Coeff}(25) \cdot \text{FRair} \cdot \text{FRwater} \cdot \text{Tr} + \\
& \text{Coeff}(26) \cdot (\text{FRwater})^2 \cdot \text{Tr} + \\
& \text{Coeff}(27) \cdot T_{wb} \cdot \text{Tr} + \\
& \text{Coeff}(28) \cdot \text{FRair} \cdot T_{wb} \cdot \text{Tr} + \\
& \text{Coeff}(29) \cdot \text{FRwater} \cdot T_{wb} \cdot \text{Tr} + \\
& \text{Coeff}(30) \cdot (T_{wb})^2 \cdot \text{Tr} + \\
& \text{Coeff}(31) \cdot (\text{Tr})^2 + \\
& \text{Coeff}(32) \cdot \text{FRair} \cdot (\text{Tr})^2 + \\
& \text{Coeff}(33) \cdot \text{FRwater} \cdot (\text{Tr})^2 + \\
& \text{Coeff}(34) \cdot T_{wb} \cdot (\text{Tr})^2 + \\
& \text{Coeff}(35) \cdot \text{Tr}
\end{aligned} \tag{15}$$

冷却塔模型考虑自由对流,通过冷却塔出水温度确定风机控制频率,其中 approach 由 T_{cwr} 决定(式(15)),同时 $\text{approach} = T_{\text{cwr}} - T_{\text{wet}}$,为了在已知风机的频率情况下求解出冷却塔出水温度,仿真平台采用遍历求解方式,具体过程如图2所示。预先按照一定的步长(本文为0.1)选取 T_{cwr} 为 x ,利用 x 值计算 approach (式(15)),同时利用式(14)计算出该情况下的 $T_{\text{cwr}}(y)$,然后记录 $|x - T_{\text{cwr}}(y)|$ 值,直到将所有 T_{cwr} 值遍历完,最终将最小的 $|x - T_{\text{cwr}}(y)|$ 对应的 T_{cwr} 值作为实际的冷却水出水温度。仿真系统由多台冷机组组成,存在多处水管混水,按照惯例,以各个分水管中水温的加权平均为总管处的水温,权重按照各个分水管的水流量比计算。

5 结果分析

5.1 实验设置

(1) 算法参数设置

为了尽可能地提高第1年的能源节约率,同时保证设备性能老化时依旧能有一定的探索率学习到全局最优的解,使用线性 ε 衰减, $\Delta\varepsilon$ 设置为0.0005, $\varepsilon_{\min}$ 设置为0.01;同时,结合样本数量关系,设置基于冷却水泵和冷却塔的存储样本池规模 $N_1 = 1000$,基于负荷分配的经验池规模 $N_2 = 500$;两个经验池的 minibatch 都设置为32;为了加大当前时刻下的系统 COP 的影响,设置 $\gamma = 0.01$;为了使算法学习更加稳定,设置 $C_{\text{step}} = 200$,学习率 $\text{learning-rate} = 0.01$ 。

(2) 对比方法设置

本文采用两种方法进行对比实验: baseline 方法以及基于模型的控制方法。为了在保护设备的同时维持室内的舒适程度, $T_{\text{chws},j,\text{set}}$ 维持在5~9℃,3台冷机组的控制方式按照图1进行。

baseline 方法:按照工程惯例,大冷机组和小冷机组的冷却水泵和冷却塔风机的频率都设置为50 Hz,同时每台冷机的 $T_{\text{chws},j,\text{set}}$ 都设置为额定值6℃。

基于模型的控制方法:大小冷机冷负荷分配使用基于约束优化和拉格朗日乘数法的 OCL 方法^[16],在此基础上,冷却塔和冷却水泵的实时优化值采用遍历求解的计算方式,从而寻找出最优控制动作。该 OCL 方法的优化目标是最小化所有冷机的功率和 $\sum_{j=1}^m (\text{ChillerEIRFPLR}_j \times P_{\text{ref},j})$,其中 $\lambda[\text{CL} - \sum_{j=1}^m (\text{PLR}_j \times \text{CC}_j)]$ 是约束项。式(16)表示冷机的准确模型(参数见表3),其中 CC_j 表示第 j 台冷机的制冷量。

$$\text{ChillerEIRFPLR}_j = g_{1,j} + g_{2,j} \times \text{PLR}_j + g_{3,j} \times \text{PLR}_j^2 \tag{16}$$

$$L = \sum_{j=1}^m (\text{ChillerEIRFPLR}_j \times P_{\text{ref},j}) + \lambda \left[\text{CL} - \sum_{j=1}^m (\text{PLR}_j \times \text{CC}_j) \right] \tag{17}$$

$$\lambda = \frac{2\text{CL} + \sum_{j=1}^m (\text{CC}_j \times g_{2,j} + g_{3,j})}{\sum_{j=1}^m \frac{\text{CC}_j^2}{g_{3,j} \times P_{\text{ref},j}}} \tag{18}$$

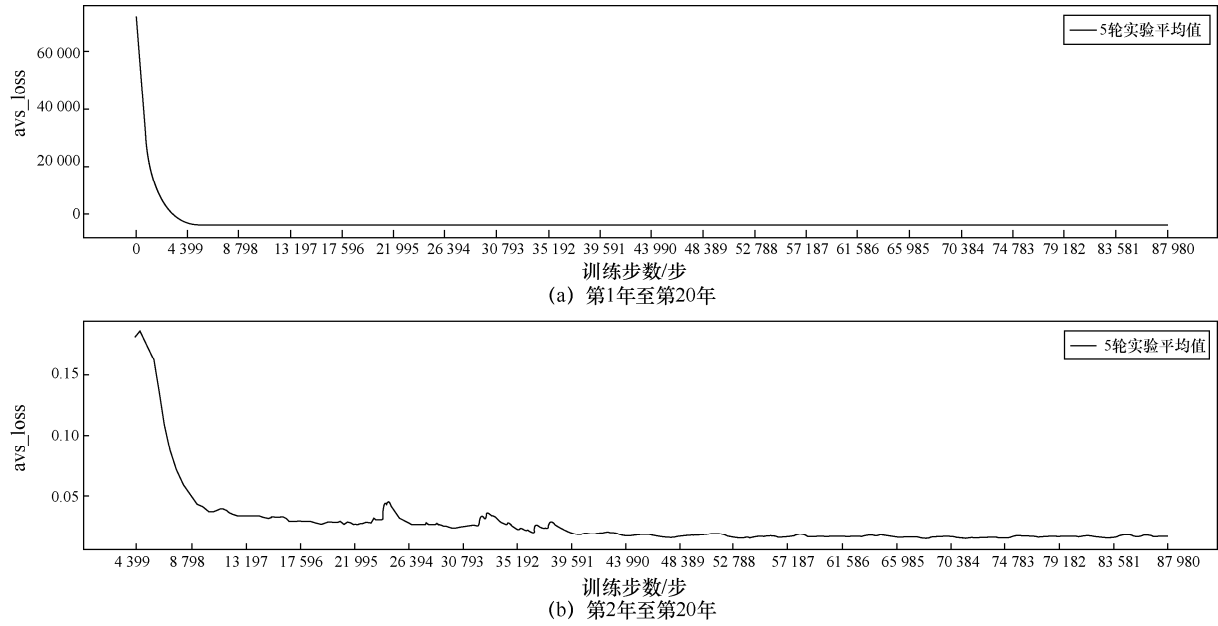


图5 经过平滑处理的 loss 曲线

$$PLR_j = \frac{\lambda \times CC_j + P_{ref,j} - g_{2,j}}{2g_{3,j}} \quad (19)$$

5.2 实验分析

(1) 算法收敛性分析

经过平滑处理的损失 (loss) 值曲线如图 5 所示, 曲线根据 5 轮独立重复实验进行平滑处理, 处理方式如式 (20) 所示, 其中 $w = 0.999$ 。图 5 中横坐标代表训练步数 (每个间隔为 1 年), 纵坐标代表平滑处理后的 loss 值 (avs_loss)。在实验过程中, 由于第 1 年的 avs_loss 值变化较大, 而第 2 年之后值变化较小, 因此为了方便观察, 图 5 (a) 和图 5 (b) 分别为第 1 年至第 20 年的 avs_loss 曲线、第 2 年至第 20 年的 avs_loss 曲线。

$$avs_loss_{t+1} = w_t \times avs_loss + (1 - w) \times loss_t \quad (20)$$

从图 5(a) 可以看出, 算法从第 2 年 (约 4 399 步) 开始逐步收敛。进一步从图 5 (b) 可以看出, 第 5 年 (约 17 596 步) 已经基本收敛至 0.027, 同时 loss 值仍有较小的波动变化, 而从第 10 年 (约 35 591 步) 之后, loss 值处于稳定, 约为 0.014。前 20 年 5 轮实验逐年奖赏值见表 5, 由表 5 可知, 前 5 轮实验结果相似, 可以看出算法相对稳定。此外, 前 10 年中 5 轮实验的平均奖赏值逐年升高, 且在第 10 年之后, 平均奖赏趋于稳定, 与图 5 (b) 中第 10 年之后的 loss 值变化相符, 因此, 可以进一步说明算法的有效性与稳定性。

(2) 系统运行效用分析

① 冷机序列控制分析

系统的序列控制通过对实时系统冷负荷进行逻辑判断, 进而达到对大小冷机的开关控制, 如图 1 所示。图 6 给出系统运行的序列控制结果, 其中横坐标表示 6 个月的制冷天数, 纵坐标表示每天的 24 个时刻段, 颜色的深浅代表大小冷机的开合状态: 0 代表所有冷机都关闭, 1 代表只开一台小冷机, 2 代表只开一台大冷机, 3 代表同时开一台大冷机和一台小冷机, 4 代表同时开两台大冷机, 5 代表所有冷机同时开放。

图 3 表明系统冷负荷在 5 月、6 月、9 月和 10 月相对较小, 在 7 月、8 月时系统冷负荷较大。在冷负荷相对较小的时间段内, 只需要开启 1 台大冷机或者 1 台小冷机; 在冷负荷相对较大的时间段内, 需要同时开启大冷机和小冷机甚至同时开启所有冷机。图 6 序列控制的结果与图 3 冷负荷数据相匹配, 这说明控制的有效性, 同时也说明数据的合理性。

② 冷负荷分配分析

大小冷机的 COP 与 PLR 的关系如图 7 所示, 其中横坐标表示冷机的 PLR 值, 纵坐标表示该 PLR 下的冷机 COP。在进行负荷分配时, 大冷机 1#和小冷机 3#第 1、第 3、第 10 年的 PLR 密度分布分别如图 8、图 9 所示, 其中横坐标表示冷机的 PLR 值, 纵坐标表示冷机 PLR 的密度分布值。图 10 给出双池 DQN 算法第 1、第 3、第 10 年选取的比率值 (ratio_{bigchiller}) 在该年份动作值集合中的密度分布。

表 5 前 20 年 5 轮实验逐年奖赏值

年	第 1 轮	第 2 轮	第 3 轮	第 4 轮	第 5 轮	平均奖赏
第 1 年	10 349.89	10 357.77	10 387.58	10 363.52	10 366.72	10 365.1
第 2 年	10 535.55	10 509.10	10 513.19	10 500.12	10 496.02	105 10.80
第 3 年	10 597.68	10 559.71	10 582.77	10 570.67	10 577.30	10 577.63
第 4 年	10 601.53	10 613.87	10 613.42	10 608.57	10 616.41	10 610.76
第 5 年	10 600.53	10 606.80	10 589.93	10 625.97	10 610.18	10 606.68
第 6 年	10 652.15	10 603.44	10 628.59	10 625.11	10 622.89	10 626.44
第 7 年	10 747.57	10 638.37	10 599.96	10 625.98	10 638.49	10 650.07
第 8 年	10 745.00	10 700.56	10 605.54	10 695.93	10 619.20	10 673.25
第 9 年	10 757.94	10 744.98	10 718.23	10 736.72	10 680.65	10 727.70
第 10 年	10 752.87	10 734.93	10 742.69	10 746.35	10 752.39	10 745.85
第 11 年	10 762.14	10 738.32	10 744.64	10 744.21	10 754.14	10 748.69
第 12 年	10 769.02	10 744.01	10 762.12	10 733.01	10 750.21	10 751.67
第 13 年	10 756.92	10 744.24	10 747.26	10 748.52	10 756.01	10 750.59
第 14 年	10 772.67	10 743.76	10 764.79	10 760.99	10 760.07	10 760.46
第 15 年	10 769.84	10 750.26	10 759.73	10 752.60	10 762.29	10 758.94
第 16 年	10 762.88	10 744.44	10 765.18	10 743.96	10 756.43	10 754.58
第 17 年	10 761.88	10 750.15	10 754.62	10 745.17	10 762.10	10 754.78
第 18 年	10 761.18	10 750.95	10 758.79	10 756.46	10 757.46	10 756.97
第 19 年	10 758.67	10 759.30	10 756.47	10 756.38	10 770.90	10 760.34
第 20 年	10 762.38	10 752.29	10 762.39	10 761.24	10 763.33	10 760.33

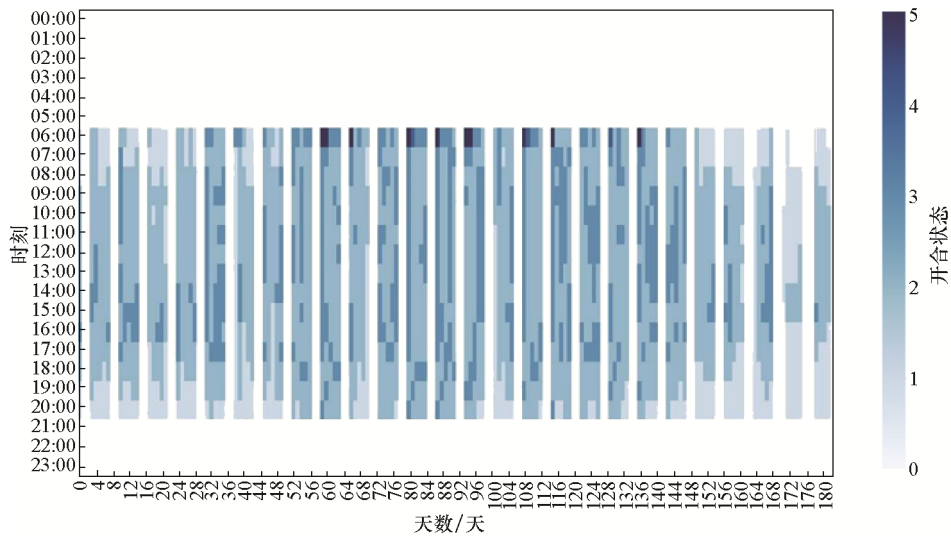


图 6 大小冷机的开合状态

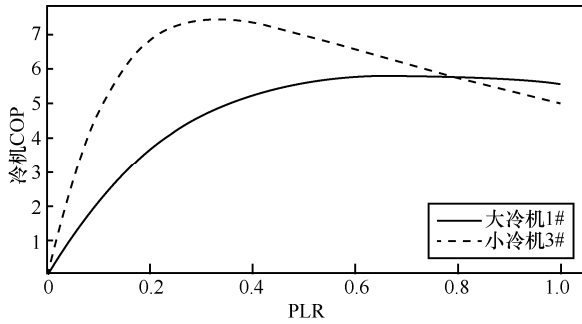


图7 大小冷机的 COP 与 PLR 的关系

基于模型的控制方法 (model-based) 的节能策略由大小冷机的特性决定。图 7 表明小冷机的 COP 在 PLR 取 30% 时最大, 之后随着 PLR 的增大逐渐变小; 大冷机的 COP 随着 PLR 的增加而增加, 在 PLR 达到 60% 之后基本维持不变。基于以上的特性, 最优的策略应该是分配较大的冷负荷给大冷机, 较少的负荷给小冷机, 使得大小冷机总的 COP 最大。

图 8、图 9 表明, 基于模型的控制方法维持

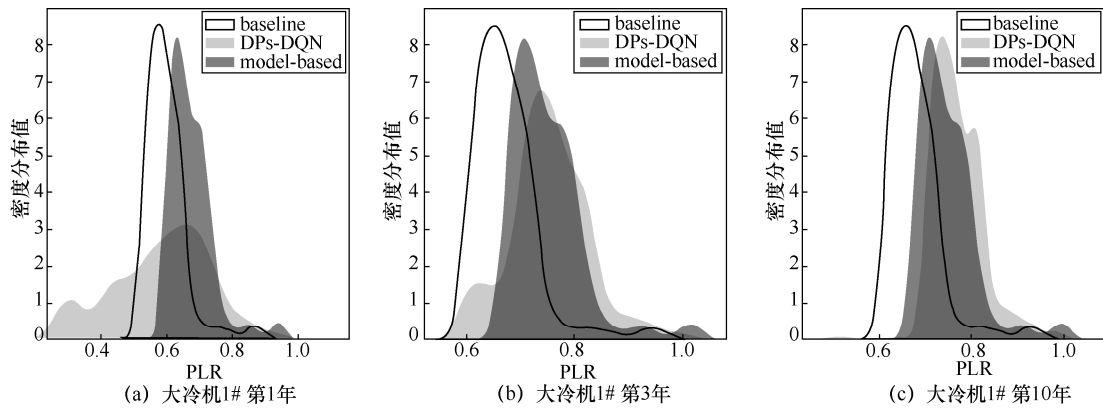


图8 大冷机年度 PLR 分布密度

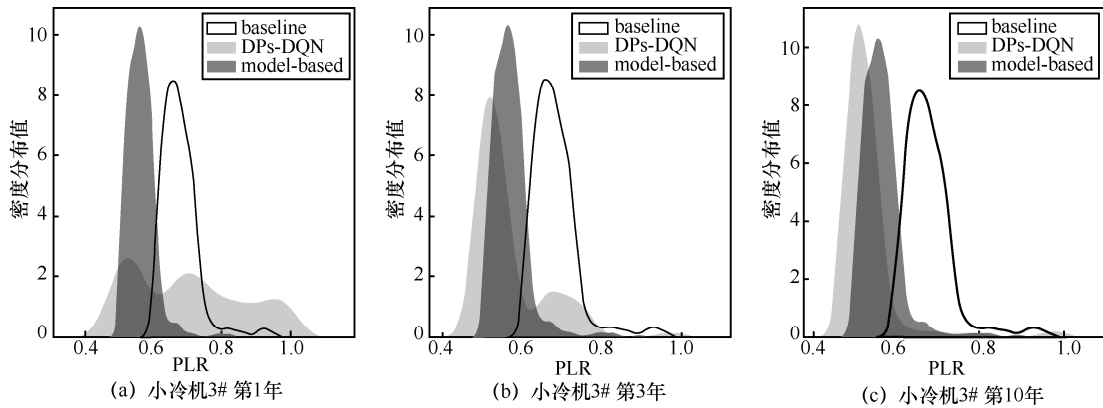


图9 小冷机年度 PLR 分布密度

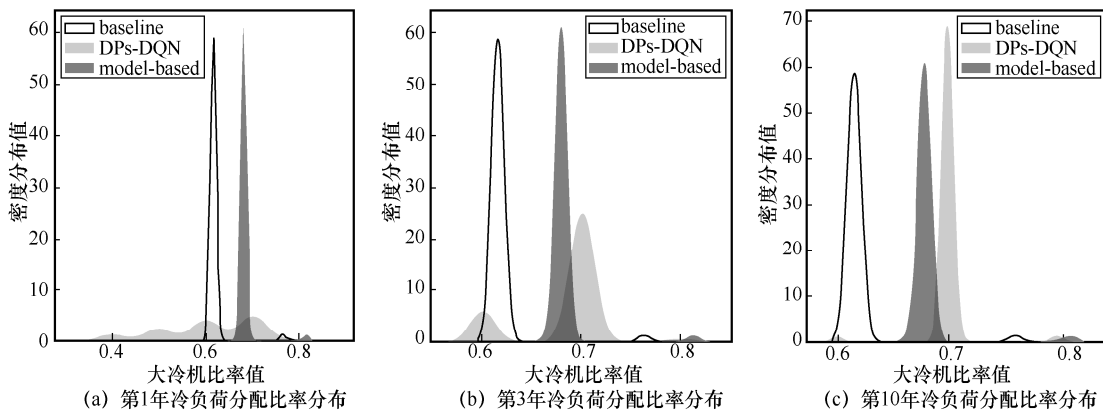


图10 分配大冷机的 ratio 分布密度

大冷机的 PLR 在 70% 左右, 小冷机的 PLR 在 55% 左右; baseline 方法根据大小冷机的冷冻水流量比值来分配冷负荷, 这使得大冷机的 PLR 维持在 65% 左右, 小冷机的 PLR 维持在 70% 左右; 基于双池 DQN 的方法在第 1 年存在大量的探索, 导致 PLR 的选取相对分散, 当学习到第 3 年时, 大冷机的 PLR 更多集中在 70% 左右, 到第 10 年时已经稳定在 75%, 基本接近基于模型的控制方法。

相比基于模型的控制方法, 双池 DQN 算法在分配更多的冷负荷给大冷机的同时, 分配更少的冷负荷给小冷机, 该结果由 agent 的动作空间决定。控制负荷分配的 agent 的动作空间的步长是 0.1, 基于模型的控制方法选取的最优 $\text{ratio}_{\text{bigchiller}}$ 集中分布在 0.67 左右, 而双池 DQN 学到的最优策略是分配 0.7 给大冷机 (如图 10 所示), 这也导致大冷机的 PLR 会比基于模型的控制方法的 PLR 稍大, 同时这也是基于双池 DQN 方法无法达到基于模型的控制方法的节能效果的主要原因之一。

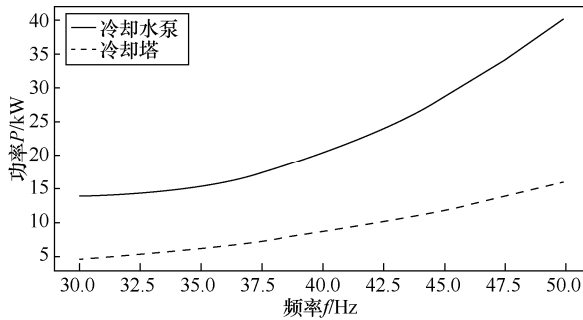


图 11 冷却水泵和冷却塔风机的功率和频率关系

③冷却塔和冷却水泵性能分析

冷却水泵和冷却塔风机的功率和频率关系如图 11 所示, 其中横坐标是频率值, 纵坐标是频率值对应的功率 (即式 (12), 其中 a 、 b 、 c 这 3 个系数由具体项目的数据通过回归方法得出)。冷却水泵和冷却塔年度频率设定点值分布密度如图 12 所示。

图 11 表明冷却水泵和冷却塔的能耗随着频率增加变化的幅度不一样, 相同频率下冷却水泵的功率值更大, 同时冷却水泵频率增加得更快, 这表明冷冻水泵总能耗高于冷却塔总能耗, 同时冷冻水泵节能幅度大于冷却塔, 年度能耗见表 6。其中冷却塔 (大) 和冷却水泵 (大) 为大冷机对应的冷却水泵和冷却塔, 大冷机开启时其开始工作, 冷却塔 (小) 和冷却水泵 (小) 为小冷机对应的冷却水泵和冷却塔, 小冷机开启时其开始工作。

根据表 1 给出的设备选型以及图 11 中冷却塔和冷却水泵的特性, 基于模型的控制方法给出的最优冷却塔风机频率和冷冻水泵频率主要分布在 (30, 30) 附近。从图 12 可以看出, 对比 baseline 方法 ($f_{\text{tower}} = 50 \text{ Hz}$, $f_{\text{pump}} = 50 \text{ Hz}$), 双池 DQN 算法在第 1 年选取的设定点分布范围比较大, 该现象主要是由强化学习方法在训练前期存在较多的探索动作导致, 当双池 DQN 算法学习到第 3 年时, 设定点的选择已经能很好地逼近最优动作, 到第 10 年时, 已经和基于模型的控制方法所给出的最优的动作基本一致, 因此, 该结果也可以验证双池 DQN 算法的有效性。

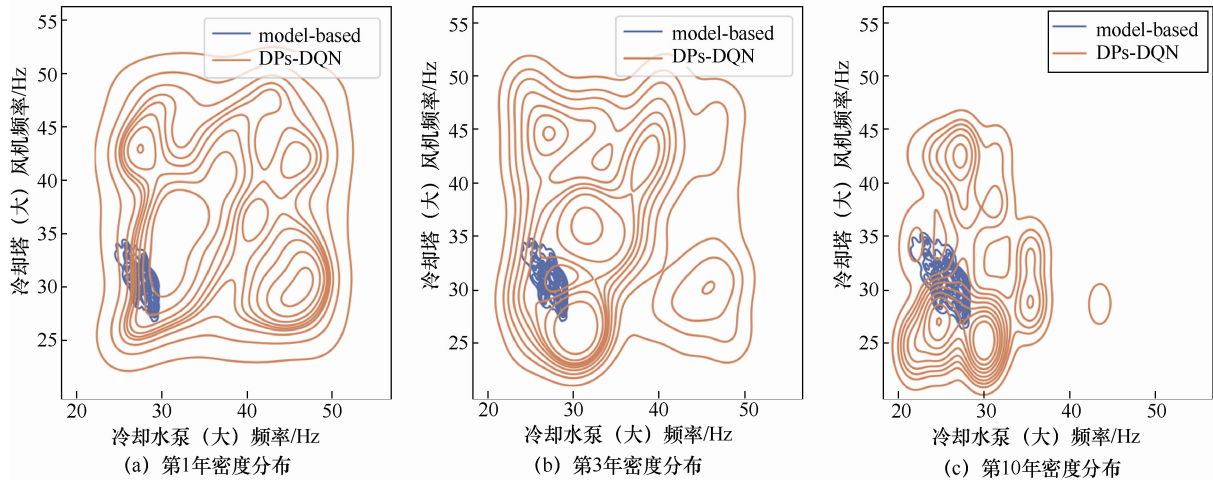


图 12 冷却水泵和冷却塔年度频率设定点值分布密度

表 6 年度能耗

方法	冷却水泵 (大) /kWh	冷却塔 (大) /kWh	冷却水泵 (小) /kWh	冷却塔 (小) /kWh	大冷机/kWh	小冷机/kWh	冷机能耗 和/kWh	总能耗/kWh
baseline	64 908.01647	25 587.13833	31 078.15252	12 251.1984	611 127.2573	167 894.455	779 021.7	912 846.2
model_based	23 377.99124	7 907.06462	10 733.4084	3 138.543602	616 325.6191	145 865.508	762 191.1	807 348.1
DPs-DQN_第 1 年	33655.12816	13 027.18418	15 703.50809	6 114.719079	603 065.8731	172 638.5861	775 704.5	844 205.0
DPs-DQN_第 2 年	30 354.49095	12 298.02102	13 562.82658	5 219.839276	614 923.3224	153 671.347	768 594.7	830 029.8
DPs-DQN_第 3 年	28 842.43923	12 093.44981	12 860.63961	5 191.032207	618 803.9614	146 572.312	765 376.3	824 363.8
DPs-DQN_第 4 年	27 193.99196	11 220.04437	12 259.24104	5 017.730849	620 979.4015	143 765.5538	764 745.0	820 436.0
DPs-DQN_第 5 年	26 677.9163	11 320.42568	12 335.2977	5 140.941364	619 248.5502	145 726.6786	764 975.2	820 449.8
DPs-DQN_第 6 年	27 151.13763	11 732.73569	12 718.44811	5 239.95047	619 721.4394	144 886.1188	764 607.6	821 449.8
DPs-DQN_第 7 年	26 080.31963	10 901.35292	11 992.71584	4 506.80003	619 883.6275	144 882.4102	764 766.0	818 247.2
DPs-DQN_第 8 年	24 783.19592	9 404.312392	11 764.99045	4 000.551805	621 846.8614	142 298.1062	764 145.0	814 098.0
DPs-DQN_第 9 年	23 424.53443	8 122.959886	10 991.93368	3 640.873975	622 796.631	141 457.2419	764 253.9	810 434.2
DPs-DQN_第 10 年	23 527.48478	8 472.471621	11 063.05897	3 926.412888	622 494.1131	141 424.745	763 918.9	810 908.3
DPs-DQN_第 11 年	23 232.457	9 605.28206	11 015.67236	3 905.091689	621 612.5477	141 355.1308	762 967.7	810 726.2
DPs-DQN_第 12 年	24 146.31587	8 757.547623	11 160.67447	3 754.003801	621 371.5256	141 156.9539	762 528.5	810 347.0
DPs-DQN_第 13 年	23 514.53632	7 960.616987	11 126.74509	3 585.772903	622 803.198	141 377.8823	764 181.1	810 368.8
DPs-DQN_第 14 年	23 880.84119	8 901.725799	11 246.87424	3 786.493418	621 172.3233	141 200.1382	762 372.5	810 188.4
DPs-DQN_第 15 年	23 561.9639	8 645.72904	11 052.77735	3 844.50329	621 676.7996	141 117.5224	762 794.3	809 899.3
DPs-DQN_第 16 年	23 269.04265	8 487.371801	11 047.49848	3 798.053333	622 130.8019	141 782.3786	763 913.2	810 515.1
DPs-DQN_第 17 年	24 108.61137	8 110.85739	11 207.49851	3 636.430704	621 415.1327	141 520.7743	762 935.9	809 999.3
DPs-DQN_第 18 年	23 045.32234	7 956.261497	10 913.77962	3 612.96288	623 052.1172	141 519.1315	764 571.2	810 099.6
DPs-DQN_第 19 年	23 208.17606	8 183.664655	10 971.49878	3 539.736593	622 153.3032	141 358.4312	763 511.7	809 414.8
DPs-DQN_第 20 年	24 227.18333	7 564.407838	11 196.99755	3 473.707856	622 159.0155	141 306.4567	763 465.5	809 927.8

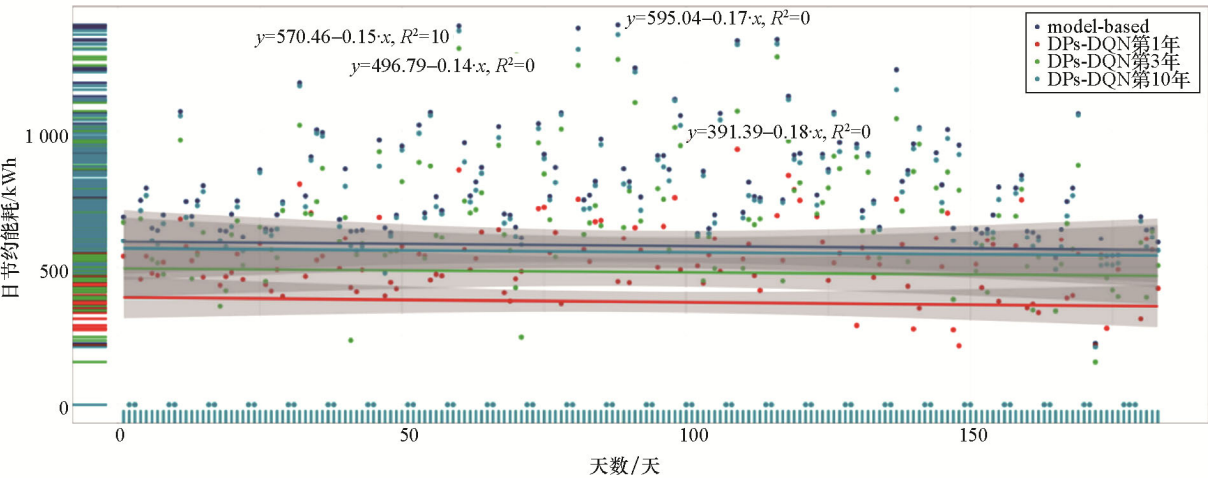


图 13 日节约能耗值

(3) 能耗分析

图 13 给出双池 DQN 算法在第 1、第 3、第 10 年和基于模型的控制方法每天相对于 baseline 方法所节约的能耗值，图 13 中的直线是根据节约能耗

值回归的一次函数。图 14 给出了 5 轮独立实验下 20 年间基于模型的控制方法相较于 baseline 方法的年度能耗节约率。

图 13 表明，在制冷需求最多的 7 月、8 月，无论

是基于模型还是无模型的控制方法, 每天节能都相对较多, 大约在 9 月之后, 由于商业楼制冷需求的下降, 节约的能耗也有所下降, 这也导致拟合出来的曲线稍微降低, 同时, 由第 1、第 3、第 10 年的回归线逐年增高可以看出, 无模型方法随着部署时间的增多节能效果越来越好, 这也验证了该双池 DQN 算法的有效性。

图 14 表明, 双池 DQN 算法的年度能耗节约率 (相较于 baseline 方法) 在 10 年间能逐步增加, 在第 3 年已经有 10% 的节能率, 到第 10 年就已经接近基于模型的控制方法的节能效果。即使是在第 1 年, 算法在存在探索的情况下, 依旧能有近 7.5% 的节能量, 从一定程度上讲, 该算法已经具备一定的用价值。为了证该算法的有效性, 对该算法进行 5 轮独立实验, 5 轮独立实验结果也验证了算法节能效果的稳定性。

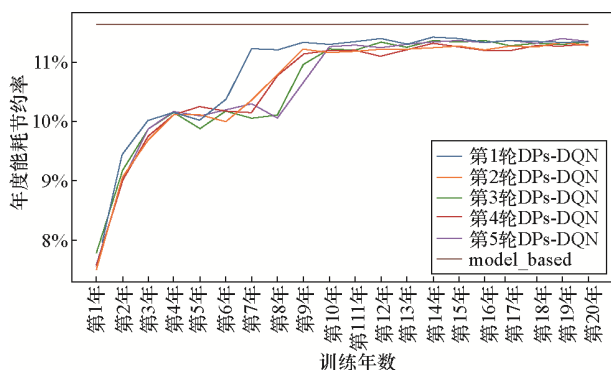


图 14 年度能耗节约率

基于双池 DQN 方法的 agent 在负荷分配、冷却塔风机频率以及冷却水泵频率三者之间学习最优的组合解。年度设备能耗变化如图 15 所示, 表 6 以及图 15 表明, 基于双池 DQN 方法的总能耗整体趋势在 10 年间不断降低, 大冷机的能耗由第 1 年的 603 065.8731 kWh 逐渐升高并稳定在 622 130 kWh 附近, 小冷机的能耗由 172 638.5861 kWh 逐渐降低并稳定在 141 500 kWh 附近, 冷却塔以及冷却水泵的功耗不断降低直至第 10 年时基本稳定。其中, 基于双池 DQN 算法的大冷机功耗不断上升直至高于基于模型的控制大冷机功耗, 小冷机功耗不断下降直至低于基于模型的控制方法的小冷机功耗的现象, 主要是由双池 DQN 算法的负荷分配方式不断优化, 同时相比基于模型的控制方法, 分配更多的冷负荷给大冷机, 更少的冷负荷给小冷机 (图 8 和图 9) 导致的。无论是基于模型的控制方法还是基于双池 DQN 算法,

冷却塔和冷却水泵能耗都低于 baseline 方法, 这导致在冷却水散热方面, 基于模型的控制方法以及基于双池 DQN 算法都不如 baseline 方法, 因此, 此时这两种方法中的冷机需要承担更大的功耗, 这也是表 6 中, 在负荷分配方式较优的情况下, 大小冷机的总能耗节约量反而较小的原因。基于双池 DQN 方法的系统总能耗逐年降低, 直至第 11 年基本稳定在 810 000 kWh, 略高于基于模型的控制方法, 这也体现出双池 DQN 算法优异的优化控制性能。

6 结束语

本文提出一种基于双池 DQN 的无模型优化控制方法, 对商业建筑中的 HVAC 系统进行优化控制。该方法在实际部署时无须建立准确的系统模型, 以逐时的系统冷负荷以及室外湿球温度为状态, 以负荷分配率、冷却塔风机频率以及冷却水泵频率为控制动作, 以系统的 COP 为奖赏。本文建立基于 EnergyPlus 标准设备模型的仿真案例系统, 并且对比 baseline 方法基于模型的控制方法, 以验证所提方法的性能。仿真结果表明, 无模型的双池 DQN 控制器能够在系统运行期间进行有效的节能控制, 同时能够随着时间的推移不断优化, 在第 10 年左右便能拥有与基于模型控制器相当的节能效用。

本文提出的基于双池 DQN 的无模型控制方法并不旨在完全取代基于模型的控制方法, 而是在缺少数据或无法部署大量传感器的情况下, 在无法建立准确系统性能模型时, 提供替代的优化控制方案。后续工作可以从以下 3 个方面展开。

(1) 由实验结果可知, 负荷比动作空间的划分方式影响冷机节能效果, 因此, 可以对动作空间进行细化, 或者选择处理连续型动作空间问题的强化学习算法。

(2) 由实验结果可知, 深度强化学习算法在第 1 年的节能效果还存在进一步提升的空间, 后续工作也可以对经验池采样方式进行优化处理, 提高算法收敛速度, 从而提升深度强化学习算法在运行第 1 年的性能表现。

(3) 本文提出的方法只验证相同型号的冷却水泵和冷却塔的节能效果, 而不同型号的冷却塔和冷却水泵的节能性能值得进一步探索。同时, 当中央冷水机组比较复杂时, 多智能体强化学习方法可能拥有更大的节能潜力。

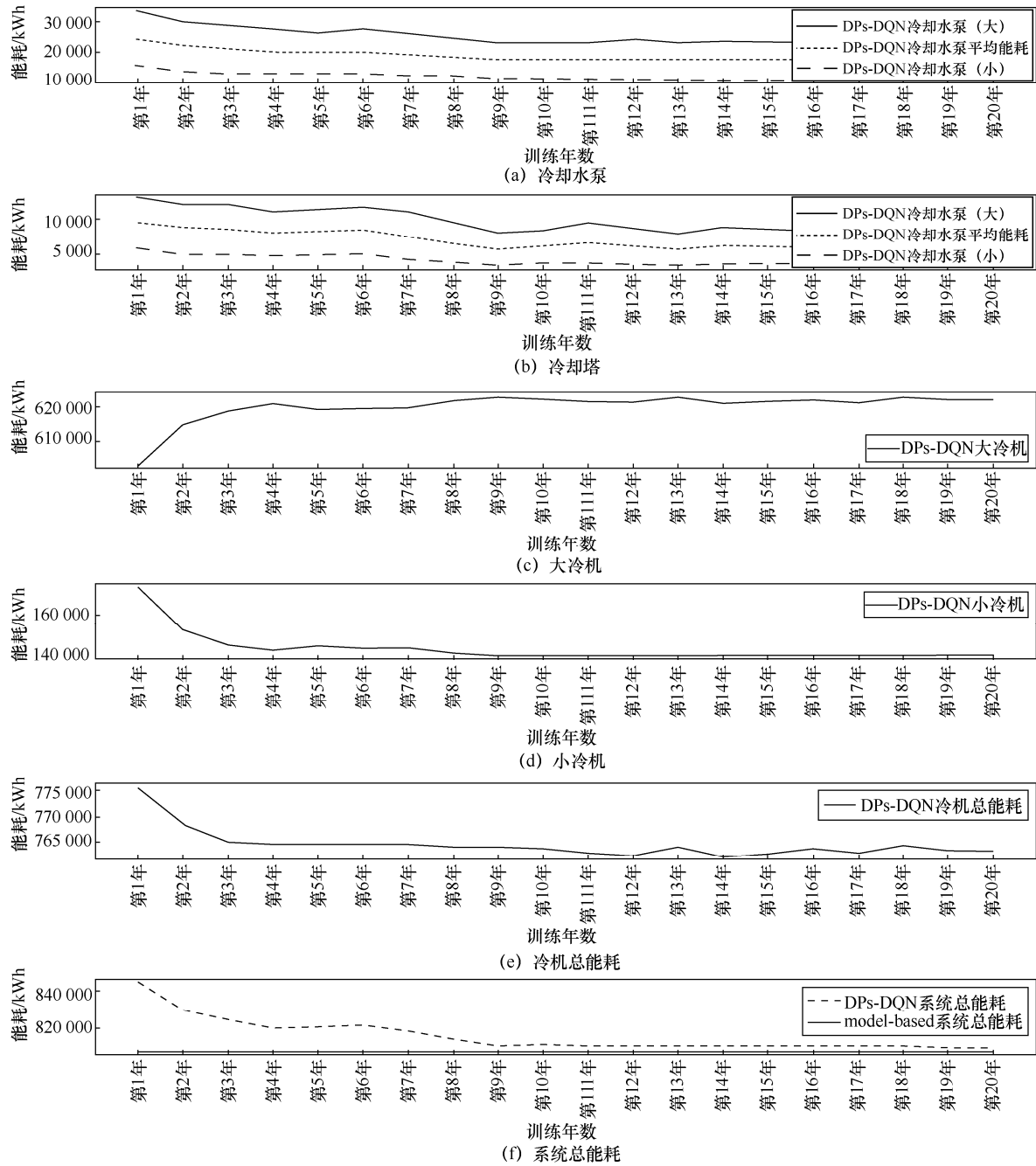


图 15 年度设备能耗变化

参考文献:

- [1] PÉREZ-LOMBARD L, ORTIZ J, POUT C. A review on buildings energy consumption information[J]. Energy and Buildings, 2008, 40(3): 394-398.
- [2] WANG J Q, HUANG J J, FENG Z B, et al. Occupant-density-detection based energy efficient ventilation system: prevention of infection transmission[J]. Energy and Buildings, 2021, 240: 110883.
- [3] HOU J, XU P, LU X, et al. Implementation of expansion planning in existing district energy system: a case study in China[J]. Applied Energy, 2018, 211: 269-281.
- [4] PANG Z H, O'NEILL Z, LI Y F, et al. The role of sensitivity analysis in the building performance analysis: a critical review[J]. Energy and Buildings, 2020, 209: 109659.
- [5] WANG J Q, HOU J, CHEN J P, et al. Data mining approach for improving the optimal control of HVAC systems: an event-driven strategy[J]. Journal of Building Engineering, 2021, 39: 102246.
- [6] MA Z J, WANG S W. Online fault detection and robust control of

- condenser cooling water systems in building central chiller plants[J]. Energy and Buildings, 2011, 43(1): 153-165.
- [7] YAO Y, LIAN Z W, HOU Z J, et al. Optimal operation of a large cooling system based on an empirical model[J]. Applied Thermal Engineering, 2004, 24(16): 2303-2321.
- [8] HUANG S, ZUO W D, SOHN M D. Improved cooling tower control of legacy chiller plants by optimizing the condenser water set point[J]. Building and Environment, 2017, 111: 33-46.
- [9] ZHU N, SHAN K, WANG S W, et al. An optimal control strategy with enhanced robustness for air-conditioning systems considering model and measurement uncertainties[J]. Energy and Buildings, 2013, 67: 540-550.
- [10] SUN Y J, WANG S W, HUANG G S. Chiller sequencing control with enhanced robustness for energy efficient operation[J]. Energy and Buildings, 2009, 41(11): 1246-1255.
- [11] LI Z W, HUANG G S, SUN Y J. Stochastic chiller sequencing control[J]. Energy and Buildings, 2014, 84: 203-213.
- [12] QIU S N, LI Z H, LI Z W, et al. Model-free control method based on reinforcement learning for building cooling water systems: validation by measured data-based simulation[J]. Energy and Buildings, 2020, 218: 110055.
- [13] BAGHAEI S, ULUSOY I. User comfort and energy efficiency in HVAC systems by Q-learning[C]//Proceedings of 2018 26th Signal Processing and Communications Applications Conference. Piscataway: IEEE Press, 2018: 1-4.
- [14] FADDEL S, TIAN G Y, ZHOU Q, et al. Data driven Q-learning for commercial HVAC control[C]//Proceedings of IEEE Southeast Con 2020. Piscataway: IEEE Press, 2020: 1-6.
- [15] 刘朝阳, 穆朝絮, 孙长银. 深度强化学习算法与应用研究现状综述[J]. 智能科学与技术学报, 2020, 2(4): 314-326.
- LIU Z Y, MU C X, SUN C Y. An overview on algorithms and applications of deep reinforcement learning[J]. Chinese Journal of Intelligent Science and Technology, 2020, 2(4): 314-326.
- [16] 刘莹莹, 王占山. 异构多智能体系统的输出同步: 一个基于数据的强化学习方法[J]. 智能科学与技术学报, 2020, 2(4): 394-400.
- LIU Y Y, WANG Z S. Output synchronization of heterogeneous multi-agent system: a reinforcement learning approach based on data[J]. Chinese Journal of Intelligent Science and Technology, 2020, 2(4): 394-400.
- [17] MNIH V, KAVUKCUOGLU K, SILVER D, et al. Human-level control through deep reinforcement learning[J]. Nature, 2015, 518(7540): 529-533.
- [18] AHN K U, PARK C S. Application of deep Q-networks for model-free optimal control balancing between different HVAC systems[J]. Science and Technology for the Built Environment, 2020, 26(1): 61-74.
- [19] CHANG Y C. A novel energy conservation method—optimal chiller loading[J]. Electric Power Systems Research, 2004, 69(2/3): 221-226.
- [20] LIAO Y D, SUN Y J, HUANG G S. Robustness analysis of chiller sequencing control[J]. Energy Conversion and Management, 2015, 103: 180-190.
- [21] ARDAKANI A J, ARDAKANI F F, HOSSEINIAN S H. A novel approach for optimal chiller loading using particle swarm optimization[J]. Energy and Buildings, 2008, 40(12): 2177-2187.
- [22] CHANG Y C, LIN J K, CHUANG M H. Optimal chiller loading by genetic algorithm for reducing energy consumption[J]. Energy and Buildings, 2005, 37(2): 147-155.
- [23] LEE W S, LIN L C. Optimal chiller loading by particle swarm algorithm for reducing energy consumption[J]. Applied Thermal Engineering, 2009, 29(8/9): 1730-1734.
- [24] QIU S N, LI Z H, LI Z W, et al. Model-free optimal chiller loading method based on Q-learning[J]. Science and Technology for the Built Environment, 2020, 26(8): 1100-1116.
- [25] CRAWLEY D B, LAWRIE L K, WINKELMANN F C, et al. EnergyPlus: creating a new-generation building energy simulation program[J]. Energy and Buildings, 2001, 33(4): 319-331.
- [26] ZHU D D, HONG T Z, YAN D, et al. A detailed loads comparison of three building energy modeling programs: EnergyPlus, DeST and DOE-2.1E[J]. Building Simulation, 2013, 6(3): 323-335.
- [27] BENTON D J, BOWMAN C F, HYDEMAN M, et al. An improved cooling tower algorithm for the CoolTools™ simulation model[C]//Proceedings of ASHRAE Annual Meeting. [S.l.:s.n.], 1999.

[作者简介]



马帅(1997—), 男, 苏州科技大学电子与信息工程学院硕士生, 主要研究方向为强化学习在智慧建筑中的应用。



傅启明(1985—), 男, 博士, 苏州科技大学电子与信息工程学院副教授, 主要研究方向为强化学习、深度学习和建筑节能。



陈建平(1963—), 男, 博士, 苏州科技大学电子与信息工程学院教授, 研究方向为大数据分析、建筑节能和智能信息。



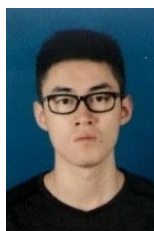
冯帆 (1993–), 男, 德州农工大学博士生, 主要研究方向为 HVAC 优化控制。



李铮伟 (1981–), 男, 博士, 同济大学机械与能源学院教授, 主要研究方向为建筑能源系统的优化运行及故障诊断等。



陆悠 (1977–), 男, 博士, 苏州科技大学电子与信息工程学院副教授, 主要研究方向为下一代网络、网路管理和机器学习。



裘舒年 (1995–), 男, 同济大学机械与能源学院博士生, 主要研究方向为强化学习在 HVAC 中的应用。