计算机系统应用 ISSN 1003-3254, CODEN CSAOBN
Computer Systems & Applications,2021,30(10):271−279 [doi: 10.15888/j.cnki.csa.008121]
©中国科学院软件研究所版权所有.
E-mail: csa@iscas.ac.cn
http://www.c-s-a.org.cn
Tel: +86-10-62661041

Dueling-DQN 在空调节能控制中的应用①

李骏翔[1], 李兆丰[1], 杨赛赛[1], 陶洪峰[1], 姚　辉[1], 吴　超[2]

[1](浙江省邮电工程建设有限公司 大数据研究院, 杭州 310052)
[2](浙江大学 公共管理学院, 杭州 310058)
通讯作者: 吴　超, E-mail: chao.wu@zju.edu.cn

摘　要: 针对电信机房空调运行耗电量大, 空调自动控制系统设计困难的问题, 提出了一种规则约束和 Dueling-DQN 算法相结合的空调节能控制方法. 该方法能根据不同机房环境自适应学习建模, 在保证机房室内温度在规定范围的前提下, 节省空调耗电量. 同时针对实际机房应用场景, 设计节能控制算法中的状态, 动作和奖励函数, 并采用深度强化学习算法 Dueling-DQN 提高模型表达能力和学习效率. 在电信机房实际验证结果表明: 该控制方法与空调默认设定参数运行相比节能 18.3%, 并可以很方便推广到不同环境场景的机房环境中, 为电信机房节能减排提供解决方案.

关键词: 节能控制; Dueling-DQN; 强化学习; 机房空调调控

引用格式: 李骏翔,李兆丰,杨赛赛,陶洪峰,姚辉,吴超.Dueling-DQN 在空调节能控制中的应用.计算机系统应用,2021,30(10):271–279. http://www.c-s-a.org.cn/1003-3254/8121.html

Application of Dueling-DQN in Air Conditioning Control for Energy Saving

LEE Chun-Hsiang[1], LI Zhao-Feng[1], YANG Sai-Sai[1], TAO Hong-Feng[1], YAO Hui[1], WU Chao[2]

[1](Research Institute of Big Data, Zhejiang Post and Telecom Engineering Construction Co. Ltd., Hangzhou 310052, China)
[2](School of Public Affairs, Zhejiang University, Hangzhou 310058, China)

Abstract: To tackle the problems of large power consumption and intricate design of the automatic control system for air conditioning in a telecommunication room, this study proposes an energy-saving control method based on the Dueling-DQN algorithm and rule constraint for mechanical control system design. With the ability to learn modeling adaptively according to the environments of different computer rooms, this method can save the power consumption of air conditioning while ensuring the indoor temperature in the specified range. Moreover, according to the actual application scenarios of computer rooms, the states, actions and reward functions of the energy-saving control algorithm are designed. Besides, a deep reinforcement learning algorithm Dueling-DQN is used to improve the model expression ability and learning efficiency. The results of actual verification in telecommunication rooms show that the control method can save energy by 18.3% compared with the air conditioning at default parameters. It can be easily extended to machines in different environmental scenarios to provide solutions for energy conservation and emission reduction of telecommunication rooms.

Key words: energy saving control; Dueling-DQN; reinforcement learning; air conditioning control in telecommunication rooms

① 收稿时间: 2021-01-04; 修改时间: 2021-01-29; 采用时间: 2021-02-23

(C)1994-2023 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

随着通讯和数字化产业的发展, 机房电量消耗越来越大. 统计发现, 机房中空调的耗电量占机房总耗电量 50% 左右[1], 怎样节省空调设备的耗电量成为亟待解决的问题. 另一方面, 电信机房很多是无人值守的, 空调的设置参数通常是固定不变或者维护人员定期去调整, 但机房的热负荷是在动态变化的 (机房设备工作负载发热, 室外温湿度情况, 机房门禁开关等都是影响热负荷的因素), 传统的空调设定参数不变或很长时间人为去调整一次的方式, 由于没有自动化调节, 通常会把空调制冷量输出设定的较大以应对可能的机房高负荷情况 (突然的高温天气或机房设备负载升高), 但多数情况下机房实际热负荷并没有那么大, 这种过度设置造成会不必要的电能浪费, 所以设计一种自适应调节的空调节能控制系统成为解决问题的关键.

郭晓岩提出了一种变风量空调系统的模糊神经网络预测控制系统[2], 动态调节出风量达到节能目的, 虽然取得一定的效果, 但针对非线性, 时序性, 状态空间较大的机房空调控制系统, 这种控制方式很难取得更好的效果, 而且空调的可控参数不止有出风量, 还有设定温度, 开关机等. Congradac 和 Kulic 采用遗传算法优化空调控制系统达到节能的目的[3], 这种控制优化方法需要专家设计建模, 人为调整优化参数, 实际应用中比较复杂, 需要大量分析调整才能找到合适参数设定.

强化学习是一种用于控制决策的人工智能算法, 它能够从与环境交互中学习控制策略, 是目前比较热门的一种机器学习算法, 广泛应用在机器人控制, 自动驾驶, 能源调度等领域. 在空调节能控制方面, 文献 [4,5] 提出了基于表格存储参数的 Q-learning 强化学习空调节能控制算法, 这种方法对状态空间较小的情况具有好的效果, 但对于输入变量较多且输入变量是连续的应用场景, 由于状态空间较大, 构建的参数表格巨大, 无法存储或检索, 导致维数灾难问题. 文献 [6–8] 提出了使用深度神经网络作为记忆体的深度强化学习, 解决空调控制中状态空间较大的问题, 取得不错的节能控制效果, 但以上文献提出的方案是在模拟环境中构建模型和实验分析, 真实的机房环境场景对室内温度和湿度有严格的要求, 不允许出现温度超出规定范围的情况, 所以很难实际推广应用.

基于以上分析, 本论文提出了一种规则约束和深度强化学习相结合的控制算法, 该算法以保证机房室内温度在要求的范围之内和节省空调耗电量为优化目标. 在保证不产生高温告警, 同时符合机房温度要求的条件下, 利用强化学习算法学习在不同环境场景下的最节能的控制策略. 同时设计了模型自动更新机制, 具备持续学习能力, 能够随着数据的积累, 不断改善空调节能控制策略, 越用越好. 该算法已成功应用到实际机房环境, 在保证机房室内温度符合要求条件下取得很好的节能效果.

1 强化学习算法介绍

1.1 强化学习算法简介

强化学习是一种需要不断与环境交互, 根据奖励反馈调整模型, 最终得最佳策略模型的机器学习方法. 强化学习是一种探索试错过程. 其在机房空调节能控制的工作原理如图 1 所示, 强化学习模型从机房及空调系统环境中观测到一个状态数据 *St*, 根据模型参数推理计算得到执行动作 *At*, 把动作 *At* 作用到机房及空调系统环境中, 环境会发生变化, 达到新的状态 *St*+1, 同时得到环境的反馈奖励 *Rt*+1, 强化学习模型根据反馈奖励优化自身参数: 某个动作导致机房内温湿度超出规定或导致空调耗电量增加, 就通过调整参数, 减弱输出该动作的趋势, 反之则会加强输出该动作的趋势. 通过强化学习模型和机房及空调系统环境的不断交互, 最终得节能收益最大化的策略, 即最优策略.

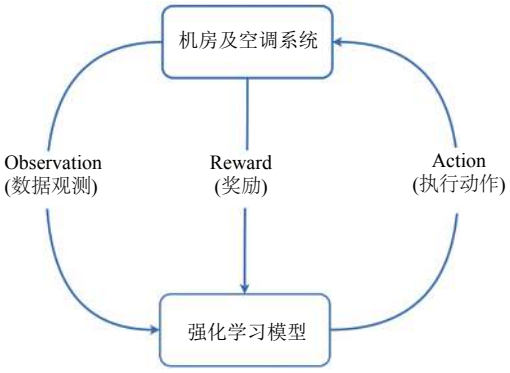

图 1 强化学习在机房空调节能控制中的原理

在机房空调节能控制场景中, 强化学习中的状态, 动作, 奖励函数的设定如下:

状态 (state, *S*): 状态是从机房及空调系统环境中观测到的信息, 需要从机房中安装的传感器和设备工作状态数据中, 选取能够反映机房环境变化的特征作为状态. 例如: 室外温度, 室内温度, 机房负载等都可以加入状态空间中.

(C)1994-2023 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

动作 (action, A): 动作是强化学习模型输出可控制空调设置的集合. 输出动作需要根据空调可控参数和空调数量来制定. 通常选取空调开关机, 设定温度作为控制动作.

奖励函数 (reward, R): 奖励函数是反馈动作作用到机房环境后造成的影响好坏的评价指标, 需要根据优化目标确定. 例如控制目标是节省空调耗电量, 则奖励函数可以根据空调耗电量值来设计.

1.2 Q-learning 算法介绍

Q-learning 是由 Watkins 提出的最简单的值迭代 (value iteration) 强化学习算法[9,10]. 与策略迭代 (policy iteration) 算法[11] 不同, 值迭代算法会计算每个“状态-动作”的价值 (value) 或是效用 (utility), 把这个值叫 Q 值, 其计算如式 (1) 所示.

$$Q(s_t,a)=R+\gamma\times\max_a Q(s_{t+1},a) \tag{1}$$

其中, s_t 为在 t 时刻的环境状态; R_{t+1} 为在当前状态 s_t 执行动作 a 之后的即时奖励; γ 为折扣因子, 表示对长期奖励回报的权重; s_{t+1} 为执行动作 a 到达的新的状态.

Q-learning 算法将环境状态 (state) 和执行动作 (action) 构建成一张 Q-table 表格来存储 Q 值参数, 表的行代表环境状态 (state), 列代表执行动作 (action) 形式如表 1 所示. 对每个状态和动作对应的 Q 值的准确估计, 是值迭代算法的核心.

表 1 Q-table 表格形式

Q-table	a_1	a_2	a_3
s_1	$Q(s_1, a_1)$	$Q(s_1, a_2)$	$Q(s_1, a_3)$
s_2	$Q(s_2, a_1)$	$Q(s_2, a_2)$	$Q(s_2, a_3)$
s_3	$Q(s_3, a_1)$	$Q(s_3, a_2)$	$Q(s_3, a_3)$

Q-learning 算法的执行流程如图 2 所示.

首先初始化 Q-table 表, 一般把 Q 值全部初始化为 0 或随机初始化, 然后算法模型与环境交互, 计算新的 Q 值, 计算方法如式 (1) 所示. 利用新计算的 Q 值更新原表格中的 Q 值, 为了使 Q-table 表的迭代更新更加平缓, 一般更保守的更新 Q-table 表, 即引入松弛因子变量 α, 保留一定比例的原 Q 值, 取一定比例新计算的 Q 值, 按式 (2) 所示更新.

$$Q(s_t,a)\leftarrow(1-\alpha)\times Q(s_t,a)+\alpha\times(R_{t+1}+\gamma\times\max_a Q(s_{t+1},a)) \tag{2}$$

在强化学习更新迭代过程中, 使用 epsilon-greedy 算法[12] 选择输出动作, 如图 3 所示, 以一部分概率随机选择动作执行, 以一部分概率按照贪婪策略, 取最大 Q 值对应动作执行. 在开始阶段由于采用随机初始化或全部初始化为 0 的方式, 所以 Q 值估计不准, 而且为了更好的探索环境, 随机选取动作的概率较大, 随着学习迭代次数的增加, 随机选取动作的概率逐渐减少, 直到降低为 0. 最终收敛到最优策略, 在执行时选择某状态下最大的 Q 值对应的动作执行.

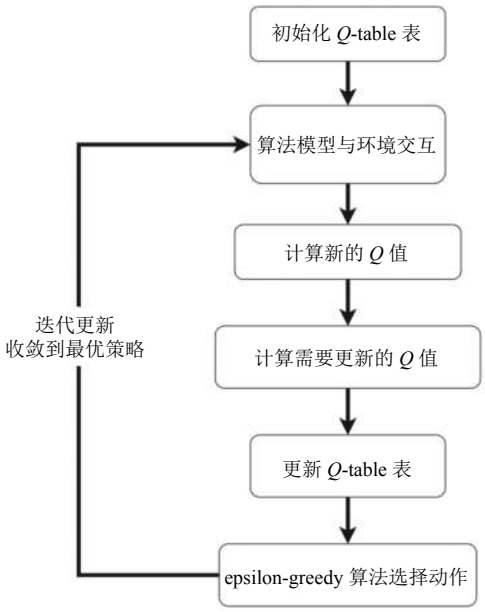

图 2 Q-learning 算法执行流程

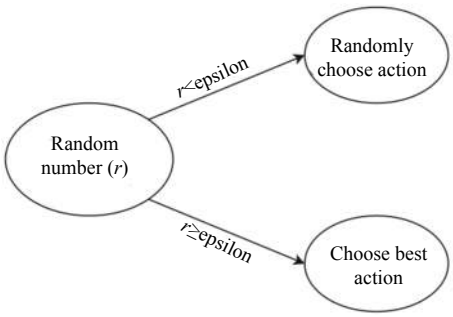

图 3 Epsilon-greedy 算法选择动作

1.3 深度强化学习算法

在电信机房环境中, 传感器和设备比较多, 状态维数空间非常大, 无法利用 Q-learning 算法建立有效的 Q-table 表格, 针对这个问题, Mnih 等人在 Q-learning 算法基础上提出利用神经网络取代 Q-table 表, 利用神经网络对值函数进行估计, 神经网络的输入是状态, 输

(C)1994-2023 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

出是各个动作的 Q 值, 即 DQN (Deep Q Network) 算法[13,14]. 全连接神经网络具有拟合泛化能力强, 结构简单, 容易训练等优点, 所以采用全连接网络作为算法存储单元. 假如输入状态有 p 个, 输出动作空间为 n, 网络结构如图 4 所示.

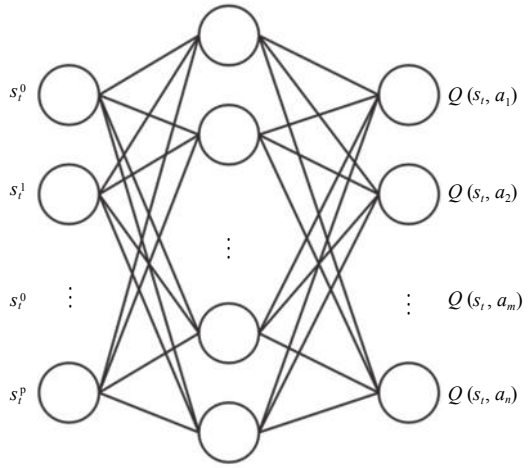

图 4 全连接 DQN 网络结构

1.4 Dueling-DQN 算法

在机房热负荷较小时 (机房设备工作负载较低, 并且机房外界温度较低), 无论空调采用什么设定动作, 机房空调低负荷运行, 耗电量基本相同. 或者在机房热负荷较大 (机房内设备工作负载高, 外界温度高) 时, 无论空调采用什么设定动作, 空调总是满负荷运行, 耗电量基本相同. 针对以上两种情况, 传统的 DQN 算法只能学习在同一状态下哪种动作设置是有价值的, 无法判定状态的价值[15], 所以无法对以上两种情况的 Q 值进行准确的估计.

Dueling-DQN 算法[15] 在不改变 DQN 输入输出的情况下, 在输出层之前将网络结构拆分为两个分支: 价值估计分支和优势估计分支, 分别对状态的价值和在该状态下不同动作的价值进行估计, 然后再把两个网络分支线性合并为输出层, 如式 (3) 所示.

$$Q(s,a;\theta,\alpha,\beta) = V(s;\theta,\alpha) + A(s,a;\theta,\alpha) \tag{3}$$

其中, V 为价值函数估计, 参数为 α. A 为优势函数估计, 参数为 β, 公共隐含层的参数为 θ, 网络结构如图 5 所示.

Dueling-DQN 算法采用把网络拆分成价值函数和优势函数单独估计, 适用于实际的机房空调控制场景, 能够在热负荷较低或较高时也能准确的对 Q 值进行估计, 并且具有很好的泛化能力[15], 所以本论文的控制决策部分采用 Dueling-DQN 算法.

1.5 算法对比

针对机房空调节能控制场景, 对比几种常用的强化学习算法 (DQN, Double-DQN, Dueling-DQN) 在机房空调节能数据集上损失函数的表现如图 6 所示.

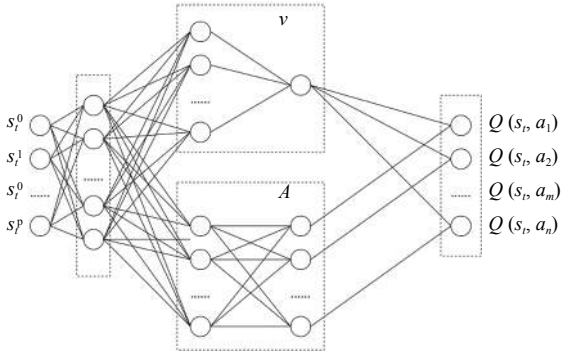

图 5 Dueling-DQN 网络结构

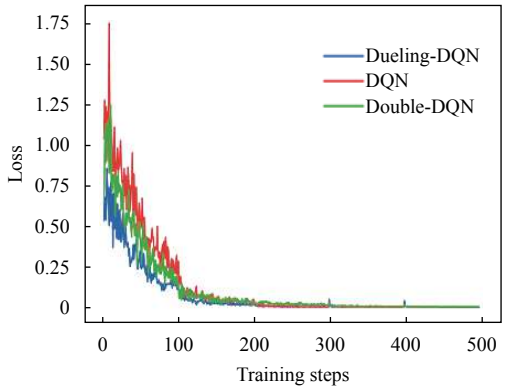

图 6 不同深度强化学习算法的损失函数变化

可以看到与 DQN, Double-DQN 相比, Dueling-DQN 算法损失函数有更快的收敛速度, 同时损失函数的波动较小, 表明在机房空调节能控制场景中, 采用 Dueling-DQN 算法具有更好的算法稳定性和训练性能表现.

2 空调节能控制系统设计

2.1 系统总体框架

Dueling-DQN 算法由于需要不断的试错才能学习到最优的控制策略, 但实际的机房环境不允许调控过程使机房室内温度超出规定的范围, 否则可能会导致机房设备工作不稳定. 另外, 强化学习算法实际应用时需要考虑机房环境发生变化时, 算法模型也要相应更新. 所以需要结合实际应用场景设计整个控制系统, 具

(C)1994-2023 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

体架构如图 7 所示, 首先控制系统会定时采集机房中所有传感器数据并存储, 然后基于采集的传感器数据通过规则约束分析和 Dueling-DQN 推理相结合得到最终的输出动作, 并下发到实际机房中执行. 执行后可以得到空调耗电量的奖励反馈及执行动作前后的传感器状态, 把这些数据存储到存储库 (memory) 中用于 Dueling-DQN 模型的训练更新. 控制系统每天固定时间从存储库中提取数据进行模型重新训练并替换旧的模型.

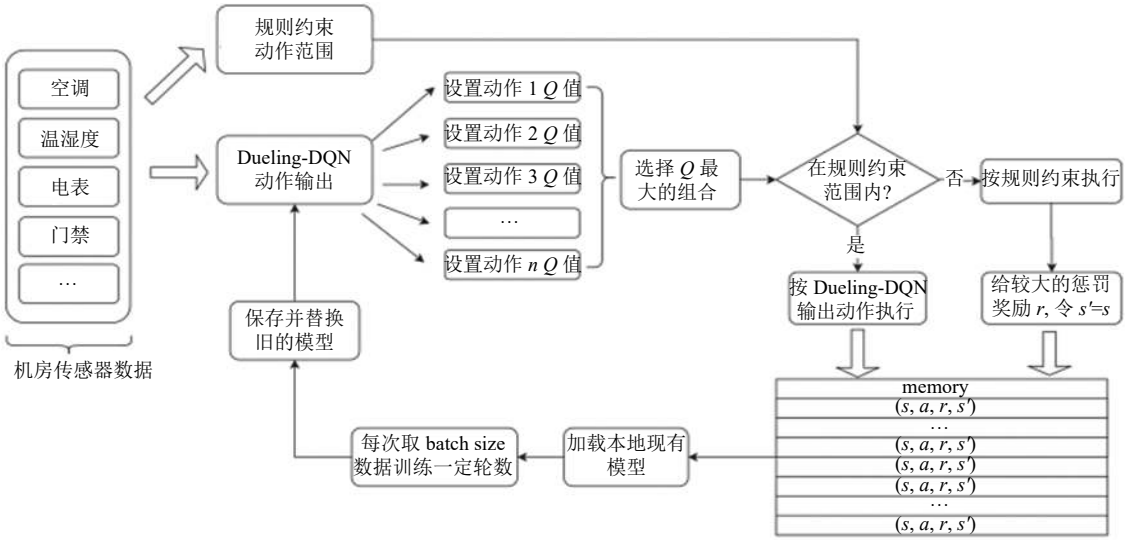

图 7 系统总体框架图

2.2 决策系统

节能控制决策部分使用 Dueling-DQN 算法和基于专家经验的规则约束相结合的方法输出最终的执行动作. 在探索学习阶段, Dueling-DQN 神经网络模型对 Q 值的估计还不够准确, 有可能输出的控制动作会导致机房室内温度超出规定的范围, 存在安全隐患, 真实的机房环境不允许出现超出温度范围要求 (一般 24–27℃, 具体根据机房管理要求确定) 的情况. 所以在 Dueling-DQN 算法基础上增加安全规则约束, 让 Dueling-DQN 算法在安全范围内运行, 同时通过奖励函数的设定, 不断优化 Dueling-DQN 神经网络参数, 使算法决策输出朝安全节能的方向进化.

机房空调不同的设定动作, 输出的制冷量不同, 耗电量也不同. 传统的空调控制系统根据室内温度和室外温度通过规则约束直接得到了确定的空调设置动作输出, 虽然也能满足机房制冷需求, 但不同机房的具体环境场景, 规则得到的动作不一定是最节能的. 约束规则如表 2 所示, 其中, a_1, a_2, a_3 为在不同室内外温度范围下空调可以执行的控制动作.

本决策系统具体执行流程如图 8 所示, 首先根据室内温度和室外温度确定多个可以设置的安全动作范围, 而不是具体一个设置动作, 这些动作具体执行时能满足机房制冷需求, 但是具体哪个动作执行后最节能还需要后续判断. 具体约束规则如表 3 所示.

表 2 传统规则约束动作输出

室内温度 (℃)	室外温度 (℃)	输出动作
…	…	…
[20, 26]	[0, 20]	a_1
[20, 26]	[20, 25]	a_2
[20, 26]	[25, 30]	a_3
…	…	…

然后决策系统推理计算 Dueling-DQN 算法的设定温度动作输出, 再判定 Dueling-DQN 算法设定温度输出是否在上述安全的设定温度范围内, 如果在安全动作范围之内, 则按 Dueling-DQN 算法输出动作执行, 如果不在安全范围之内则按传统规则约束输出动作执行, 同时给予该步动作执行一个惩罚奖励.

2.3 模型更新机制

控制算法在运行时从机房中获取各种传感器数据作为当前状态 s, 经过控制决策部分输出执行动作 a 作用到真实的机房环境中执行, 获取即时奖励 r, 机房环境发生变化达到新的状态 st+1, 在这个过程中积累一个训练样本数据 (s, a, r, s'). 系统设计了用于存储训练样本的存储库, 并引入经验回放机制[12] 便于模型的离

(C)1994-2023 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

线训练. 存储库具有固定容量大小, 当存储库存储满了之后, 再存入新的数据时, 最先存入到存储库中的数据会被丢弃, 这种机制保证了训练样本的时效性, 保证模型能够自适应机房环境的变化.

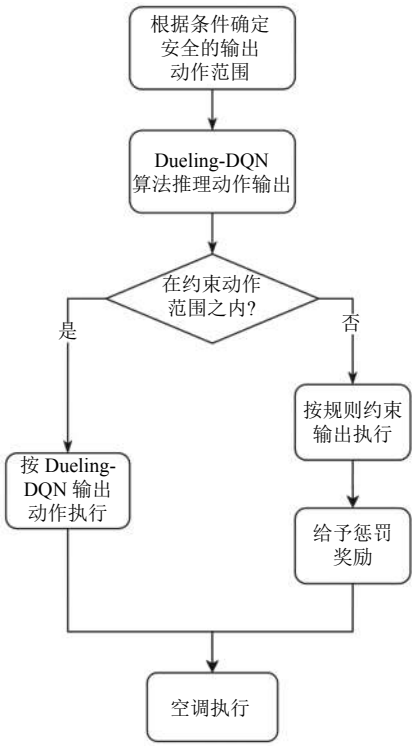

图 8 规则约束 Dueling-DQN 算法执行流程

表 3 根据约束得到安全动作范围

室内温度	室外温度	安全动作范围
…	…	…
[20, 26]	[0, 20]	a_1, a_2, a_3, a_4, a_5
[20, 26]	[20, 25]	a_3, a_4, a_5, a_6
[20, 26]	[25, 30]	a_5, a_6, a_7
…	…	…

Dueling-DQN 算法模型每天固定时间训练更新: 首先加载当前模型参数, 然后从存储库中每次随机提取 batch size 的数据样本进行训练, 训练一定轮数收敛到更好的模型参数时, 用新模型替换旧模型. 这种模型更新机制既能使模型继承原来的部分策略又能根据机房环境的变化更新策略, 能够自适应处理随季节变化, 机房设备性能衰减导致的控制策略变化.

3 实验验证

3.1 实验机房准备

选择电信某机房作为实验环境, 该机房有一台定频空调设备, 送风方式为上通风, 机房建筑面积 33.86 m^2, 机房设备工作总负载为 6.6 kW 左右, 热量来源包括 3 个方面: 太阳辐射, 室内外温度差, 设备工作散热. 该机房维护管理要求室内温度保持在 20–25℃ 之间, 机房示意图如图 9 所示.

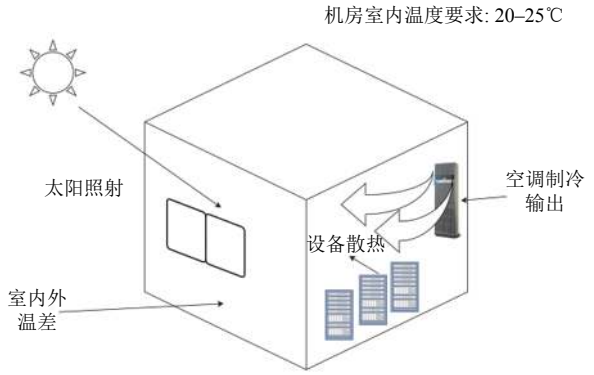

图 9 机房热平衡示意图

实验前在空调和其他工作设备上安装电表, 统计空调耗电量和机房设备总工作负载; 并在机房外面安装一个温湿度传感器; 室内安装两个温湿度传感器; 通过监控系统可以获取室内外温湿度数据, 设备耗电量和功率数据, 空调运行状态数据.

3.2 算法参数设定

在实际应用中, 需要根据机房实际情况确定控制系统中的各个参数, 才能使算法正常运行, 主要包括以下参数设定: Dueling-DQN 模型输入参数选择, 控制输出动作范围设定, 算法奖励函数设定, 模型训练和更新参数设定.

(1) Dueling-DQN 模型输入参数选择

根据机房热量来源和与节能控制相关性分析, 选择室外温湿度, 室内平均温度, 室内平均湿度, 天气状况, 门禁状态, 空调工作状态, 机房设备总负载功率作为输入特征, 这里设定数据采样频率为 1 次/min (当然也可以提高采集频率, 要综合考虑数据采集设备的承受能力), 选取最近 60 min 以上特征的历史数据作为模型输入. 具体特征介绍如下:

机房室外温湿度: 通过机房外墙温湿度传感器采集得到, 包括湿度和温度数据.

室内平均温湿度: 室内两个温湿度点位的平均值, 包括温度和湿度数据.

天气状况: 通过气象接口获取, 包括晴, 阴, 多云, 雨等不同状态.

(C)1994-2023 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

门禁状态: 通过监控系统中门禁设备采集到的工作状态, 判定机房门是打开还是关闭.

空调工作状态: 空调运行状态数据, 包括送风温度, 回风温度, 设定温度, 压缩机工作状态.

机房设备总负载功率: 通过机房中所有工作设备的电表采集到的功率值相加得到.

(2) 控制输出动作范围设定

控制动作是作用在机房空调上的, 不同品牌的空调, 可以进行控制的控制操作不同, 这里我们空调品牌是爱默生, 可进行的控制动作包括: 空调开关机和设定温度调节 (调节范围为 20–26℃), 通过监控系统可以把控制动作下发到空调执行.

(3) Dueling-DQN 算法奖励函数设定

为了确保算法能朝着减少空调耗电量方向优化, 并且在规定的安全约束内运行, 设计的奖励函数分为两部分, 公式为 $R=-P+S$, $-P$ 是空调在一个交互周期中的耗电量, 优化目标是减少空调耗电量所以取负值. S 是安全约束奖励, 在安全约束范围之内取 0, 在安全约束之外取 $-|T_{上限}-T_{下限}|$, 其中, $T_{上限}$和 $T_{下限}$分别是安全约束要求的上限和下限温度.

(4) 算法模型训练和更新参数设定

Dueling-DQN 算法模型神经网络部分使用的是全连接神经网络, 包括两个公共隐含层, 神经元节点数都为 20. 价值函数分支包括一个隐含层, 节点数为 10, 优势函数包括一个隐含层, 节点数为 20, 具体结构如图 5 所示. 实验验证过程中, Dueling-DQN 算法其他超参数设定如表 4 所示.

表 4 其他模型和训练超参数设定值

超参数	参数值
学习率	0.01
初识探索比率	0.5
贪婪增量	0.01
存储库 (memory)大小	150
batch size	30
折扣因子	0.9
训练优化算法	RMSPropOptimizer

3.3 模型更新验证

本文所述空调节能控制系统的一个创新点就是能够在原有模型基础上继续训练优化策略得到新的模型, 新的模型即保留了旧模型部分策略, 又根据机房环境的变化添加或替换部分策略, 更适应当前机房环境, 实现模型的自适应更新, 同时也能减少重新训练所消耗的计算资源. 为了验证模型根据新数据更新策略的同时保留旧的策略, 进行以下实验, 通过观测模型损失函数的变化验证以上特性.

首先进行第 1 个实验, 选取 100 个样本数据, 训练 200 轮, 得到损失函数随着训练轮数的变化趋势, 如图 10 所示, 可以看到模型损失函数随着训练轮数的增加逐渐收敛到一个固定的值, 说明模型得到了一个稳定的控制策略. 然后进行第 2 个实验, 先选取 100 个样本数据, 训练 200 轮; 然后再加载训练好的模型, 选取 100 个新的样本数据, 再训练 200 轮; 然后再加载训练好的模型, 再选取 100 个新的样本数据, 再训练 200 轮; 总共训练 600 轮, 损失函数随训练轮数的变化趋势如图 11 所示, 可以看到, 在 200 轮, 400 轮加入新的数据继续训练时, 模型损失函数虽然略有上升, 但整体峰值逐渐下降, 模型能够根据新数据学习到新的策略, 同时保留历史训练数据的得到的策略.

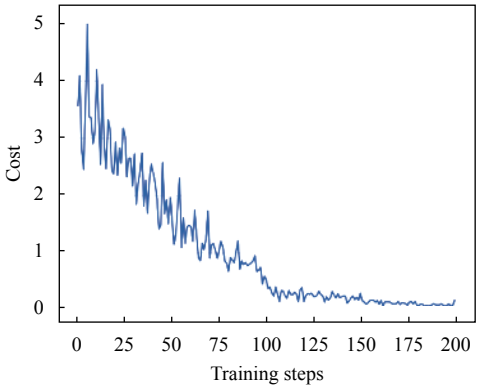

图 10 初始训练 200 轮损失函数变化

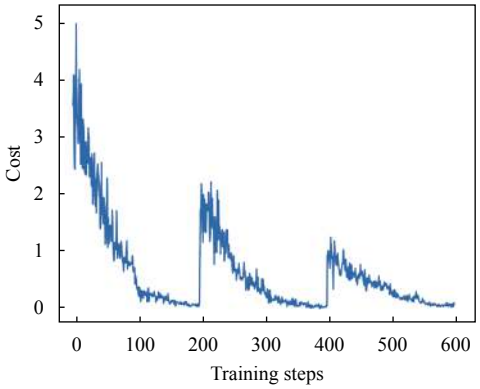

图 11 加载模型继续训练损失函数变化

假设每增加 100 个样本需要训练 200 轮, 每一轮训练占用的计算资源和消耗时间相同, 随着新数据的

(C)1994-2023 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

不断增加, 如果每次都采用全部数据重新训练, 那么模型训练时间会成倍增加, 当模型训练时间大于控制间隔时间, 新的控制策略无法及时部署, 会对下一次的控制策略产生影响, 所以模型根据新的数据继续训练更新很有必要.

3.4 节能效果验证

为了证明本文所述 Dueling-DQN 控制算法具有节能效果, 我们对比采用 Dueling-DQN 控制算法和不采用该控制算法机房空调耗电量, 这里把采用 Dueling-DQN 算法的运行方式称为节能控制模式, 把不采用 Dueling-DQN 算法的运行方式称为传统运行模式, 也即是机房原来的默认设置 (空调设定温度为固定值, 例如空调开机设定 20℃). 机房按照以下方式运行: 奇数天按 Dueling-DQN 算法调控运行, 偶数天按传统默认设置运行, 运行一个月时间, 统计耗电量数据如表 5 所示, 经过一个月数据实验对比, 传统运行模式平均每天耗电量 37.61 kWh, 节能控制模式平均每天耗电量 30.72 kWh, 平均节能百分比为 18.3%. 受季节因素影响, 在冬天室外环境较冷的情况下, 节能效果会更好, 夏天节能效果会稍差.

表 5 耗电量数据对比

日期	运行模式	空调耗电量 (kWh)
2020/02/11	节能控制模式	26.3
2020/02/12	传统运行模式	34.5
2020/02/13	节能控制模式	30.6
2020/02/14	传统运行模式	38.2
2020/02/15	节能控制模式	29.1
2020/02/16	传统运行模式	35.9
2020/02/17	节能控制模式	28.8
2020/02/18	传统运行模式	40.2
…	…	…
2020/03/09	节能控制模式	26.9
2020/03/10	传统运行模式	35.8
2020/03/11	节能控制模式	30.6

从表 3 中取室外温度接近, 机房负载接近的两种运行模式的数据进行对比, 如图 12 所示, 在室外温度基本相同, 节能控制模式总负载功率与传统运行模式接近 (部分时间段负载更高) 的情况下, 空调耗电量反而更低, 并且能够保证室内温度在安全约束范围之内 (20–25℃), 说明节能控制算法能够学习到较好的节能控制策略.

把两种运行模式的数据分开来看, 图 13 为采用 Dueling-DQN 算法的节能控制模式, 可以看到空调设定温度动作是在动态变化, 算法模型会根据外部数据的变化动态调整空调设定温度 (例如机房室外温度升高, 机房负载功率升高时, 设定温度降低, 提供更多的制冷量). 而传统运行模式, 如图 14 所示, 并没有结合其他传感器数据的动态调节, 空调设定温度一般是人为设定的一个固定值, 并且为了应对可能的高温天气和设备负载突然升高, 一般会设定到一个较低的值, 提供有冗余的制冷量输出. 本实验验证作为对比的传统运行模式的空调设定温度为 20℃, 当然也可以把对比的传统运行模式空调设定温度调整到 21℃ 或更高, 对应的节能百分比也会降低, 但风险也会增大, 而且这种改变是在我们经过数据分析这个先决条件之后得到的, 所以本论文采用空调设定温度为 20℃ 作为传统运行模式.

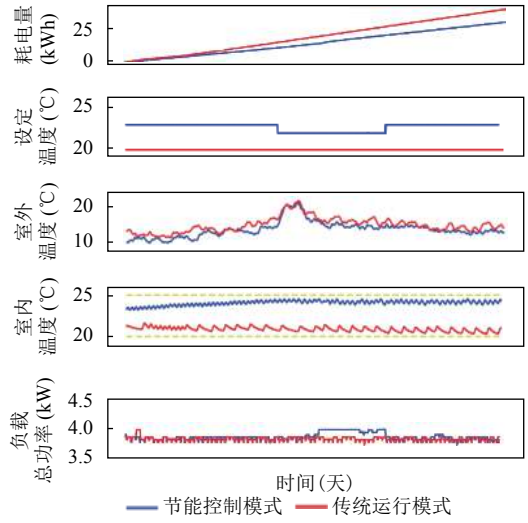

图 12 不同模式运行对比图

4 结论

本论文提出了基于规则约束的强化学习节能控制算法, 设计基于规则约束的强化学习探索和执行方法, 保证模型运行过程中输出动作的安全可靠. 结合电信机房应用场景, 采用 Dueling-DQN 算法, 提高在机房负荷较低或较高的情况下模型的训练的速度和表达能力. 设计模型自动更新机制, 根据新数据持续优化网络模型, 使节能效果越来越好. 本论文提出的机房空调控制方法解决了强化学习在实际应用中的问题: 学习过程中安全性问题, 模型自动更新问题. 控制算法具有很好的推广应用能力, 通过对控制参数的简单修改可以推广到铁塔机房, 基站机房, 大型 IDC 机房等场景.

(C)1994-2023 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

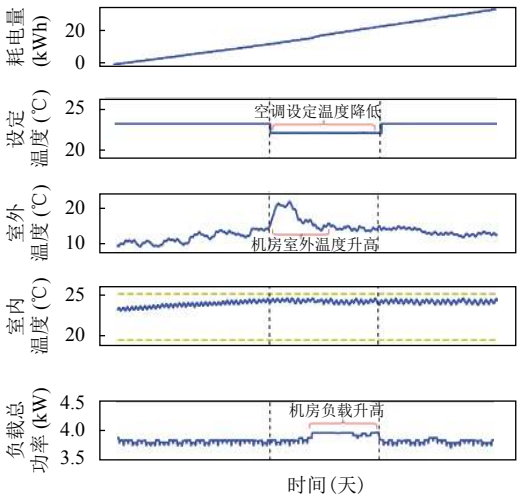

图 13 节能控制模式运行数据图

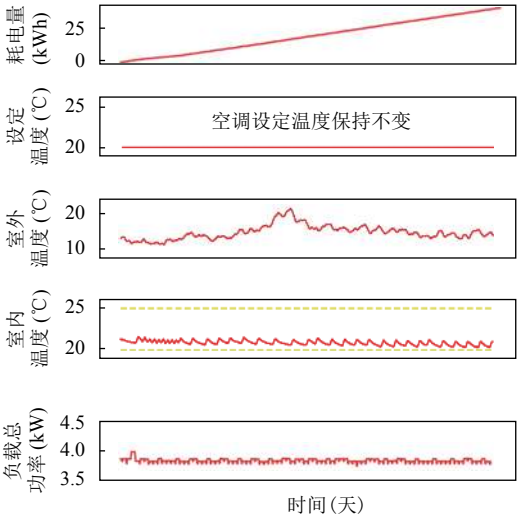

图 14 传统运行模式运行数据图

参考文献

1 张会平, 王小召. 建筑节能及建筑节能措施. 四川建筑科学研究, 2006, 32(4): 178–180. [doi: 10.3969/j.issn.1008-1933.2006.04.051]

2 郭晓岩. 变风量空调系统的模糊神经网络预测控制. 沈阳工业大学学报, 2013, 35(1): 99–103. [doi: 10.7688/j.issn.1000-1646.2013.01.17]

3 Congradac V, Kulic F. HVAC system optimization with CO_2 concentration control using genetic algorithms. Energy and Buildings, 2009, 41(5): 571–577. [doi: 10.1016/j.enbuild.2008.12.004]

4 Chen YJ, Norford LK, Samuelson HW, *et al*. Optimal control of HVAC and window systems for natural ventilation through reinforcement learning. Energy and Buildings, 2018, 169: 195–205. [doi: 10.1016/j.enbuild.2018.03.051]

5 胡龄爻, 陈建平, 傅启明, 等. 一种面向建筑节能的强化学习自适应控制方法. 计算机工程与应用, 2017, 53(21): 239–246. [doi: 10.3778/j.issn.1002-8331.1702-0217]

6 Wei TS, Wang YZ, Zhu Q. Deep reinforcement learning for building HVAC control. Proceedings of the 54th ACM/EDAC/IEEE Design Automation Conference. Austin: ACM, 2017. 1–6.

7 Wang Y, Velswamy K, Huang B. A long-short term memory recurrent neural network based reinforcement learning controller for office heating ventilation and air conditioning systems. Processes, 2017, 5(3): 46.

8 闫军威, 黄琪, 周璇. 基于 Double-DQN 的中央空调系统节能优化运行. 华南理工大学学报 (自然科学版), 2019, 47(1): 135–144.

9 Watkins CJCH, Dayan P. Q-learning. Machine Learning, 1992, 8(3): 279–292.

10 Tsitsiklis JN. Asynchronous stochastic approximation and Q-learning. Machine Learning, 1994, 16(3): 185–202.

11 Sutton RS, Barto AG. Reinforcement Learning: An Introduction. 2nd ed. Cambridge: MIT Press, 2018.

12 Tokic M, Palm G. Value-difference based exploration: Adaptive control between epsilon-greedy and Softmax. Proceedings of the 34th Annual Conference on Artificial Intelligence. Berlin Heidelberg: Springer, 2011. 335–346.

13 Mnih V, Kavukcuoglu K, Silver D, *et al*. Playing atari with deep reinforcement learning. arXiv: 1312.5602, 2013.

14 Mnih V, Kavukcuoglu K, Silver D, *et al*. Human-level control through deep reinforcement learning. Nature, 2015, 518(7540): 529–533. [doi: 10.1038/nature14236]

15 Wang ZY, Schaul T, Hessel M, *et al*. Dueling network architectures for deep reinforcement learning. arXiv: 1511.06581, 2015.

(C)1994-2023 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net