DOI:10.19651/j.cnki.emt.2210117

基于强化学习的艾灸机器人温度控制策略研究*

张 博[1] 黄 山[1] 张洽芮[2] 李应昆[3] 涂海燕[1]
(1.四川大学电气工程学院 成都 610065; 2.四川省成都市第五人民医院康复医学科 成都 611130;
3.四川省成都中医药大学附属医院针灸康复科 成都 610072)

摘 要: 针对传统 PID 控制算法在艾灸机器人温度控制中存在参数辨识复杂、适应性差等问题,将强化学习引入到艾灸机器人温控领域中,提出了一种改进强化学习算法。首先,通过多物理场仿真软件和神经网络联合搭建智能体离线训练仿真环境,以解决智能体在线训练效率低下的问题;然后,提出一种结合奖励引导和余弦退火策略的改进强化学习算法,提高算法的收敛性和成功率;最后,将仿真环境训练后的模型迁移到真实环境进行实验验证。实验结果表明,温度超调量为 0.2 ℃,稳态温度保持在 43.1 ℃±0.4 ℃内,改进后的强化学习算法相比于传统 PID 控制算法的温度控制能力更好。

关键词: 艾灸机器人;温度控制;强化学习;奖励引导;余弦退火
中图分类号: TP399 **文献标识码**: A **国家标准学科分类代码**: 510.8050

Study on temperature control strategy of moxibustion robot based on reinforcement learning

Zhang Bo[1] Huang Shan[1] Zhang Hanrui[2] Li Yingkun[3] Tu Haiyan[1]
(1. School of Electrical Engineering, Sichuan University, Chengdu 610065, China; 2. Department of Rehabilitation Medicine, Chengdu Fifth People's Hospital, Chengdu 611130, China; 3. Department of Acupuncture and Rehabilitation, Affiliated Hospital of Chengdu University of Traditional Chinese Medicine, Chengdu 610072, China)

Abstract: Aiming at the problems of complex parameter identification and poor adaptability of traditional PID control algorithm in temperature control of moxibustion robot, reinforcement learning is introduced into the field of temperature control of moxibustion robot, and an improved reinforcement learning algorithm is proposed. First, the offline training simulation environment of the agent is jointly built by multi-physics simulation software and neural network to solve the problem of low efficiency of online training of the agent; then, an improved reinforcement learning algorithm combining reward guidance and cosine annealing strategy is proposed to improve the convergence and success rate of the algorithm; finally, the model trained in the simulation environment is transferred to the real environment for experimental verification. The experimental results show that the temperature overshoot is 0.2 ℃, and the steady-state temperature is kept within 43.1±0.4 ℃. The improved reinforcement learning algorithm has better temperature control ability than the traditional PID control algorithm.

Keywords: moxibustion robot; temperature control; reinforcement learning; reward guidance; cosine annealing

0 引 言

艾灸是传统医学重要组成部分,在医疗保健领域有广泛应用[1]。艾灸通过艾条燃烧产生的温热刺激作用于人体穴位或病灶,从而达到预防保健和治疗的目的[2-3]。传统人工施灸效率低,利用可控器械模拟艾灸手法,可在一定程度上提高施灸效率。赵国友等[4]设计了基于 STM32 的艾灸机械臂,可进行模拟施灸操作。姚勇等[5]设计了艾灸手法模拟系统,能模拟雀啄灸等施灸动作。张艳[6]设计了一种智能艾灸仪,可通过控制机械器件模拟艾灸手法进行施灸。

温度控制是施灸过程的重要环节,施灸时穴位皮肤温度过高可能导致烫伤,温度过低将影响疗效[7]。赵鑫等[8]在自制的迷你红外灸疗仪中,通过 Bang Bang 控制方法启

收稿日期:2022-05-25
* 基金项目:四川省重大科技专项(2019ZDZX0019)、四川省中医药管理局项目(2018KF013)资助

停电阻发热片来防止温度超过阈值。但艾灸温度变化存在大时滞大惯性的特点,Bang Bang 控制容易使温度产生震荡,导致控温效果不佳。袁浩等[9]设计了一种共享艾灸理疗系统,使用 PID 算法进行温度调节。但 PID 算法在温度控制中存在参数设定和调试困难等问题,不能满足在不同施灸对象条件下对参数自整定的需求,自适应性较差[10]。徐新等[11]利用自整定模糊 PID 控制算法对目标点温度进行控制,减小温度超调量,一定程度上提高了自适应性。但模糊控制方法通常需要大量先验知识设计模糊规则和隶属函数。

强化学习是通过智能体与环境的持续交互,不断获取环境状态信息来更新控制策略的一种智能控制算法。强化学习在控制过程中不依赖于被控对象模型和过多先验知识,具有自适应、自学习的优点,并且强化学习算法能够根据被控对象状态信息对策略进行动态调整,实现对复杂系统的智能控制[12-13]。基于此,本文将双重深度 Q 网络强化学习算法(double DQN,DDQN)引入到艾灸机器人温度控制领域中,但传统 DDQN 奖励函数由于缺乏引导和惩罚等约束项,训练时会出现稀疏奖励问题,智能体学习速度慢甚至难以收敛[14]。此外,传统的动作选择策略中如果探索率选择不合适,容易导致智能体不能学习到有效的控制策略或者陷入局部最优解。

针对以上问题,本文对传统 DDQN 算法进行改进,提出了一种结合奖励引导和余弦退火策略的强化学习算法(reward guidance and cosine annealing DDQN, RGCA-DDQN),在奖励函数和动作选择策略方面对传统 DDQN 算法进行了改进,并对改进后的强化学习算法温度控制能力进行了验证。首先,通过 COMSOL 多物理场仿真软件和长短时记忆完全卷积神经网络(long short term memory-fully convolutional networks, LSTM-FCN)联合搭建艾灸机器人温度仿真环境。其次,以控制施灸点皮肤温度达到期望温度范围为目标,将改进后的强化学习算法 RGCA-DDQN 在仿真环境进行学习训练。最后,将训练好的强化学习模型迁移到艾灸机器人硬件平台,对算法的有效性和优越性进行验证,实验结果表明本文算法具备良好的温度控制能力。

1　仿真环境模型搭建

由于进行强化学习训练时智能体需要与环境频繁交互,而对于在现实环境进行艾灸的艾灸机器人(智能体)来说,如果让其直接与被控对象(被灸穴位)进行交互试错来获取状态数据更新控制策略,不仅时间效率低,而且危险性高。因此智能体通常提前在仿真环境进行学习训练,再进行策略迁移,不仅可以使智能体有较高的训练效率,而且不存在安全风险。

COMSOL 是一款基于有限元的多物理场仿真软件,能够对热学等多种物理场进行建模。利用 COMSOL 仿真软件,可以模拟被灸穴位过热等极端情况,不会受到安全距离限制、温度测量等因素影响,可以得到更全面的作用时间-施灸距离-温度序列数据让智能体进行学习训练。但 COMSOL 求解时需要对热场特征方程进行大量耗时的数值计算,若智能体直接和 COMSOL 交互会导致训练效率极低,不利于算法研究,本文借鉴文献[15]思想,利用 COMSOL 和 LSTM-FCN 神经网络联合搭建强化学习智能体离线训练仿真环境,通过数据驱动的方式对艾灸温度进行建模。

首先在 COMSOL 艾灸热场仿真模型中预设艾灸不同的作用时间和施灸距离的组合,得到时间-距离-温度序列数据集,然后使用 LSTM-FCN 神经网络对数据集进行学习,通过神经网络学习时间-距离-温度序列数据集的内在关系,得到温度预测模型,最后以该模型作为强化学习仿真训练环境。温度预测模型以距离信息作为输入以表征强化学习仿真训练环境的动作输入,以温度信息作为输出以表征强化学习仿真训练环境的状态输出。

1.1　艾灸热场仿真模型

通过 COMSOL 搭建的艾灸热场仿真模型如图 1 所示,模型包括艾条燃烧末端,空气介质和生物组织 3 个部分,生物组织部分由表皮层、脂肪层、肌肉层组成。r 为径向水平轴,z 为纵向垂直轴,D 为施灸距离,表示艾条燃烧端到皮肤表面的直线距离,温度监测点为艾条棒末端正下方皮肤表面处的位置,与艾灸机器人实际温度监测位置相同。通过在热场仿真模型预设不同作用时间和施灸距离的组合,经求解共产生 45 000 组时间-距离-温度样本数据,将其随机分成两个数据集(训练集和测试集),用于神经网络模型的训练和测试,数据占比分别为 70% 和 30%。

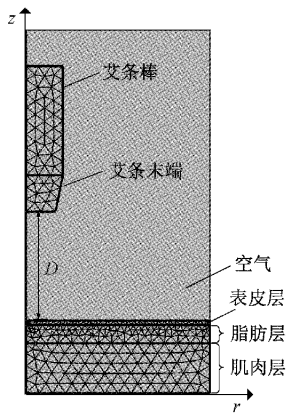

图 1　艾灸热场仿真传热模型

1.2　神经网络结构

本文所搭建的 LSTM-FCN 网络结构如图 2 所示[16],LSTM 模块由 LSTM 层和 Dropout 层组成,FCN 模块由 3 个时序卷积块组成,每个卷积块包含一个卷积层、一个批归

一化层(batch normalization，BN)层和一个 ReLU 激活函数层组成，将 LSTM 模块和 FCN 模块经过 Concat 层拼接，通过 Dense 全连接层输出状态值信息。

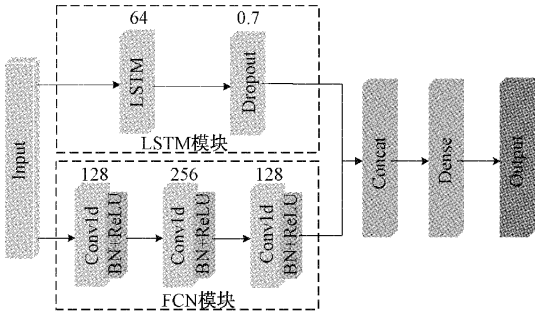

图 2　LSTM-FCN 网络结构

1.3　模型性能分析

使用训练集分别对 LSTM、FCN 和 LSTM-FCN 三种神经网络进行训练得到不同的仿真环境模型，并用测试集对 3 种模型进行性能评估。表 1 为性能评估结果对比，MAE 和 MAPE 分别表示平均误差和平均百分比误差。可以看出，LSTM-FCN 模型 MAE 值为 0.054，MAPE 值为 0.12%，相比其它两种模型误差更小，精度更高，说明本文所搭建的仿真环境模型能够很好地反映输入输出样本数据的非线性映射关系，从而使智能体具有较好的学习训练效率。

表 1　三种网络模型结果

模型名称	MAE	MAPE/%
LSTM	0.107	0.21
FCN	0.188	0.30
LSTM-FCN	**0.054**	**0.12**

2　基于 DDQN 算法的优化设计

2.1　状态信息-动作空间

本文将施灸穴位处皮肤温度 T、当前时间步 t_{step} 和前后时间步的温度差 ΔT 共 3 个状态量定义为状态信息 S，如式(1)所示。

$$S=\{T,t_{step},\Delta T\} \tag{1}$$

以艾条燃烧末端到皮肤表面的距离表示施灸动作，根据施灸距离对温度的影响特点[17-18]，将施灸动作 a 离散化为 3～6 cm 的动作集，如式(2)所示。

$$a\in\{3\text{ cm},3.5\text{ cm},4\text{ cm},4.5\text{ cm},5\text{ cm},5.5\text{ cm},6\text{ cm}\} \tag{2}$$

2.2　设计奖励引导函数

强化学习中智能体在探索学习时，若在施灸处温度达到设定范围时才给予奖励，其他情况下奖励为 0，将会导致所获得的奖励过于稀疏，即智能体不知道某个状态下所输出的动作对最终奖励贡献程度，学习导向性差，学习效率低下[19]。

本文设计了一种多目标奖励引导(reward guidance)函数，使智能体能提高学习效率，能较快达到期望的目标状态。如式(3)所示，将目标 1 奖励引导函数 $F(S)$ 和目标 2 奖励引导函数 $V(S)$ 叠加重构后得到新的奖励函数 $R'(S)$。

$$R'(S)=F(S)+V(S) \tag{3}$$

目标 1：引导智能体使被灸穴位处皮肤温度达到期望温度范围。在艾灸过程中确保施灸安全是非常重要的，本文将 T_{max} 作为期望温度范围的上限，若温度超过 T_{max} 就给予智能体较大的惩罚；将 T_{min} 作为期望温度范围的下限，对越靠近 T_{min} 的温度状态就给予智能体越大的奖励，使智能体学习具有倾向性。基于目标 1 的奖励塑形函数可被描述为如式(4)所示。

$$F(S)=\begin{cases}T-T_{min}+5, & T<T_{min}\\ 5, & T_{min}\leqslant T<T_{max}\\ -10, & T\geqslant T_{max}\text{ 或 }\Delta T>1\end{cases} \tag{4}$$

式中：$F(S)$ 为目标 1 奖励函数，S 为状态信息，T 为当前穴位温度。$T-T_{min}+5$ 为温度值 $T<T_{min}$ 时的奖励值，越接近 T_{min} 所获得奖励值越大；温度值在 T_{min} 和 T_{max} 之间奖励值为 5；温度超过 T_{max} 或者相邻时间步温度差 ΔT 超过 1 ℃智能体就会受到惩罚，奖励值为－10。以 T_d 表示期望目标温度，期望温度范围取 $T_d\pm1$ ℃，期望温度下限 T_{min} 为 T_d-1 ℃，期望温度上限 T_{max} 为 T_d+1 ℃。

目标 2：引导智能体将温度快速上升到期望目标温度。艾灸过程中，如果穴位皮肤温度长时间未达到期望温度值，虽然保证了施灸安全，但长时间温度过低难以使艾灸治疗效果最大化，因此需要引导智能体将温度较快地上升到期望温度值。故而奖励引导函数对于能较快到达期望温度的动作就给予奖励，对于长时间未到达期望温度的动作进行惩罚，基于目标 2 的奖励引导函数可被描述为式(5)所示。

$$V(S)=\begin{cases}10, & t_{step}\leqslant t_{min}\quad T\geqslant T_d\\ -10, & t_{step}>t_{min}\quad T<T_d\end{cases} \tag{5}$$

式中：$V(S)$ 为目标 2 奖励函数，t_{step} 为当前时间步，若在时间步 t_{min} 内达到期望温度则奖励值为 10；如果超过了时间步 t_{min} 还未达到期望温度将给予惩罚，奖励值为－10。

通过选取不同期望目标温度值对智能体进行学习训练，得到不同目标温度所对应的控制策略，从而智能体可根据不同目标温度的控制策略输出动作(施灸距离)来调节施灸处皮肤温度。

2.3　动作选择策略优化

在传统 DDQN 算法中，智能体通常是采用 ε-*greedy* 贪婪策略进行动作选择。智能体做动作决策时既需要尽可能多的尝试不同的动作，也要考虑利用经验选择能获得更高奖励的动作。一方面，若探索率 ε 设置过大会使得智能体在学习过程中过多地进行动作探索，导致收敛速度慢甚

至难以收敛,另一方面,如果 ε 的值设置过小,智能体会过度使用经验作出动作决策,容易陷入局部最优解[20]。

智能体在学习训练初期,应更多的偏向于动作探索以获取环境信息,而当所获得的环境信息较为丰富后,智能体动作探索应当趋于保守,以选择最大价值动作为主,使算法能够稳定收敛,从而得到最大的累计奖励。本文通过余弦退火(cosine annealing)策略对探索率 ε 进行优化,以解决智能体在动作探索和经验利用之间的不平衡问题,使智能体能够跳出局部最优解,快速完成寻优任务。

余弦退火策略公式如式(6)所示。

$$\varepsilon=\begin{cases}\lambda(1+\cos(\pi\cdot m/M))/2, & \varepsilon>10^{-8}\\ 10^{-8}, & \text{其他}\end{cases}\tag{6}$$

式中:ε 为探索率,λ 为最大探索率,满足 $0<\lambda<1$,m 为当前迭代回合,M 为总迭代回合数。动作选择策略原理如下:

$$a=\begin{cases}a_p, & rand()>\varepsilon\\ a_r, & \text{其他}\end{cases}\tag{7}$$

式中:a_p 为根据状态 S 得到的最优策略动作,a_r 为动作集中随机选择的动作,$rand()$ 为 0～1 的随机数。若随机数 $rand()<\varepsilon$ 则采取随机动作 $a=a_r$,否则采取最优策略动作 $a=a_p$。

2.4　算法结构图及步骤

RGCA-DDQN 算法结构如图 3 所示,RGCA-DDQN 算法包含估计网络和目标网络,估计网络用来计算当前 Q 值,并通过余弦退火策略选择动作;目标网络用于计算目标 Q 值。算法每隔一定时间用估计网络中的参数来更新目标网络的参数,从而减小当前 Q 值和目标 Q 值之间的相关性。经验池用于存放当前状态、动作、奖励、下一状态信息。智能体训练时,从经验池中随机采样小批量记忆样本进行训练,从而提高学习效率[21]。

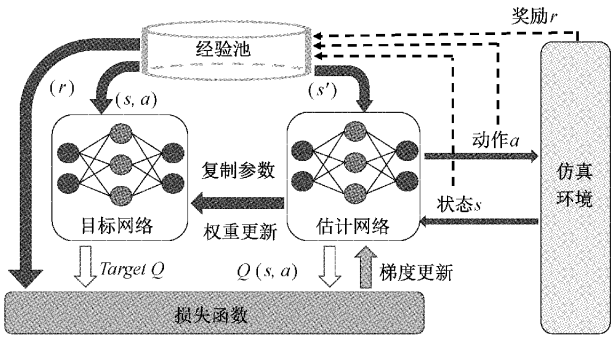

图 3　RGCA-DDQN 算法结构

本文 RGCA-DDQN 算法参数设置如表 2 所示。主要包括经验池大小 *memory_size*,迭代回合数 M,网络参数更新回合数 N,单回合总步长 *steps*,动作探索率 λ,每回合所对应的时间长度 L,状态信息维度大小 *s_dim*,动作集维度大小 *a_dim*。

表 2　算法参数设置

描述	参数值
经验池大小 *memory_size*/个	500
迭代总回合数 M/个	1 000
网络权重更新回合数 N/个	10
单回合总步长 *steps*/步	100
动作探索率 λ	0.5
单回合映射时长 L/s	900
状态维度 *s_dim*	3
动作集维度 *a_dim*	7

RGCA-DDQN 算法步骤如下:

算法 1:RGCA-DDQN 算法

输入:上述表 2 中算法参数。

输出:最优的估计网络参数 w。

1. 随机初始化:估计网络参数 w,目标网络参数 w'。

2. 从 1 到 M(训练总回合)进行迭代。

①初始化最初状态 s;

②估计网络依据式(7)动作策略选择动作 a;

③执行动作 a,由式(1)得到新状态 s',由式(3)得到奖励值 r;

④将[s,a,r,s']四元组存入经验池集合 D 中;

⑤更新状态 $s=s'$;

⑥从经验池集合 D 中随机采样样本,通过目标网络得到目标 Q 值 *TargetQ*;

⑦使用均方差损失函数进行梯度反向传播,更新估计网络参数 w;

⑧若当前迭代回合 m 满足 $m\%N=1$,则使用估计网络权重参数更新目标网络参数,$w'=w$;

⑨若 s 为终止状态,则本轮迭代结束,否则转到步骤②。

3　实验结果与分析

3.1　仿真环境训练

本文仿真实验软件配置为:操作系统为 64 位 Windows 10,编程软件为 Pycharm,深度学习框架为 TensorFlow 1.14,编程语言为 Python3.7。硬件配置为英特尔酷睿 i5-9400F 处理器、16 GB 内存以及英伟达 GTX 1650 4 G 显卡。

强化学习中奖励值是评价智能体控制策略优劣的重要标准。奖励值主要分为 3 个阶段:第 1 阶段,奖励值小于 250 时,奖励值越小越说明智能体不具备有实用经验,在环境中不断试错受到较多惩罚,导致奖励值较低;第 2 阶段,奖励值在 250～750 范围时,说明智能体通过前期试错已具备一定的经验策略,因此获得的奖励值逐渐增加;第 3 阶段,奖励值在 750～1 250 范围,说明智能体已拥有实用的经验,从环境中学习到了较优的控制策略,能在回合中得到

更多的正奖励值，此阶段奖励值大于 900 时表示智能体在回合中已具备有较好的控制能力。

图 4 为传统 DDQN 算法、RG-DDQN 算法和 RGCA-DDQN 算法的奖励值变化曲线，仿真实验结果数据如表 3 所示。从收敛速度上看，传统 DDQN 算法在训练时难以收敛，RG-DDQN 算法和 RGCA-DDQN 算法分别在 325 和 221 回合收敛，RGCA-DDQN 算法收敛速度更快，收敛速度分别提升 67.5%和 77.9%，证明奖励引导机制能够改善奖励稀疏性，提高算法收敛速度。从收敛稳定性上看，传统 DDQN 算法在整个训练过程奖励值波动大，平均奖励值为 413，说明智能体受到较多的惩罚，始终无法学习到较优控制策略；奖励引导 RG-DDQN 算法针对目标进行学习，训练中后期能逐渐收敛，平均奖励值为 741，但奖励值波动依旧明显；而 RGCA-DDQN 算法能够跳出局部最优，所获得奖励值更高，平均奖励值为 929，相较于传统 DDQN 奖励值增加了 516，在整个训练过程的奖励值波动较小，收敛稳定性也更好。

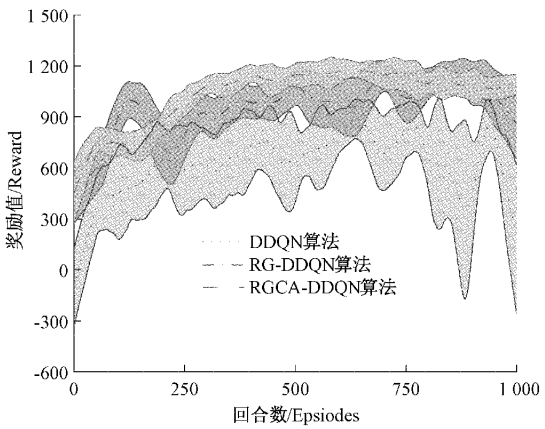

图 4　三种算法奖励曲线

表 3　三种算法仿真训练结果

算法名称	平均奖励值	收敛回合数
DDQN	413	1 000
RG-DDQN	741	325
RGCA-DDQN	**929**	**221**

图 5 所示为 3 种算法的成功率曲线，表示智能体每 100 回合在预设时间步内达到期望温度范围的回合次数。传统 DDQN 算法学习没有导向性，在整个过程中成功率均较小且学习速度慢，训练结束算法成功率仍小于 40%；奖励引导 RG-DDQN 算法成功率随着训练进行逐步上升，但由于智能体还不能平衡经验利用和动作探索之间的关系，导致训练结束后成功率小于 70%；而 RGCA-DDQN 算法在第 300 个回合后成功率快速提升，随着智能体经验的积累，在第 700 个回合成功率已高达 91%，训练结束时成功率达到 95%，相比于传统 DDQN 算法，改进后的 RGCA-DDQN 算法有更高的成功率。

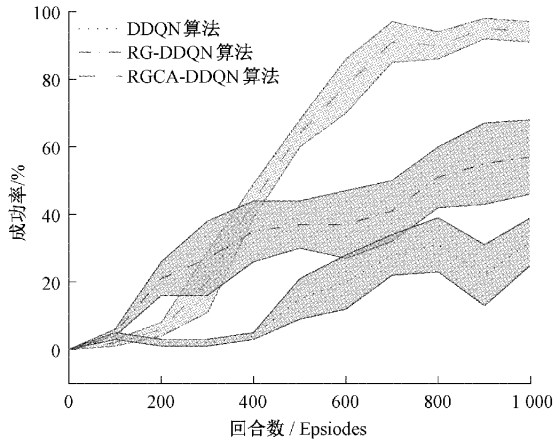

图 5　三种算法的成功率曲线

3.2　仿真环境实验结果

本文以 43 ℃作为期望目标温度进行实验，该温度既是使机体能产生抗炎效应的有效灸温[22]，也是避免皮肤烫伤的安全温度[23]。3 种算法的温度响应曲线和实验数据结果分别如图 6 和表 4 所示，传统 DDQN 算法的温度响应曲线从初始温度到达 42 ℃的爬坡时间最长，爬坡时间为 223 s，最晚达到期望温度区间，其稳态误差为 1.3 ℃，温度波动较大。RG-DDQN 算法能较快到达期望温度区间，爬坡时间为 93 s，但整体平均温度更高，容易超过期望温度上限，且稳态误差为 0.9 ℃。RGCA-DDQN 算法能控制温度较快到达设定区间，同时能更好保持温度平稳，稳态误差为 0.5 ℃，说明 RGCA-DDQN 算法能学习到更好的控制策略，温度控制效果更好。

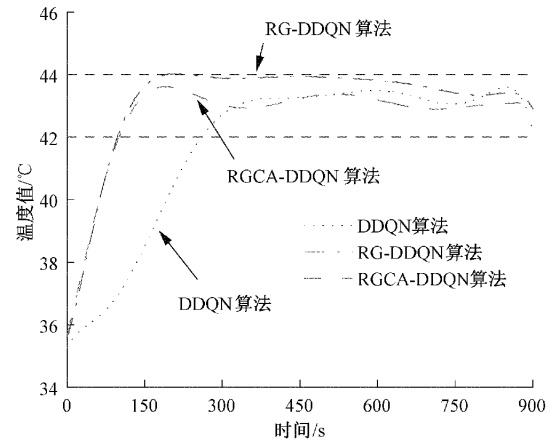

图 6　仿真温度响应曲线图

表 4　三种算法仿真实验结果

算法名称	平均温度/℃	稳态温度/℃	稳态温度误差/±℃	爬坡时间/s
DDQN	41.74	43.21	1.3	223
RG-DDQN	43.13	43.75	0.9	93
RGCA-DDQN	42.64	43.12	0.5	93

3.3　真实环境实验结果

为进一步验证本文算法实际的温度控制性能,将训练后的强化学习模型移植到课题组自研的艾灸机器人硬件平台上进行实验,并将其与传统 PID 算法作为机器人控制器进行对比。艾灸机器人硬件平台[24]如图 7 所示,该硬件平台的机械结构主要包括机械臂本体和工具末端两部分,工具末端是根据艾灸机器人的任务特点而设计,用于携带艾条棒以及部署温度和距离传感器。艾灸机器人采用非接触式红外温度传感器 MLX906140ESF 测量施灸穴位处皮肤温度,通过激光距离传感器 VL53l0X 测量艾条棒到施灸穴位皮肤表面的施灸距离。

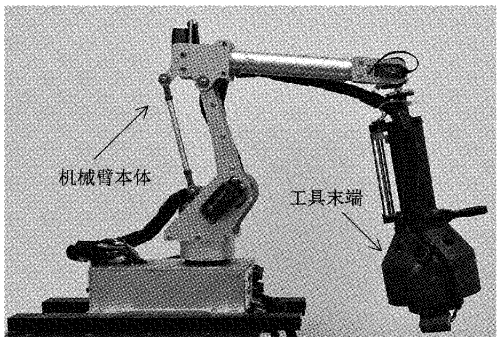

图 7　艾灸机器人硬件平台

两种算法的温度响应曲线实验结果如图 8 所示,两种算法实验的温度数据结果如表 5 所示。首先,传统 PID 算法的稳态温度为 43.26 ℃,RGCA-DDQN 算法的稳态温度为 43.15 ℃,本文算法稳态温度更接近于最佳期望温度 43 ℃;PID 算法稳态温度误差为 0.7 ℃,RGCA-DDQN 算法稳态温度误差为 0.4 ℃,本文算法的稳态温度误差更小;PID 算法的超调量为 0.8 ℃,RGCA-DDQN 算法超调量为 0.2 ℃,本文算法超调量也更小;其次,以整个艾灸过程温度响应曲线稳定性来看,本文 RGCA-DDQN 算法温度波动范围也更小,对被控对象的适应性更好。

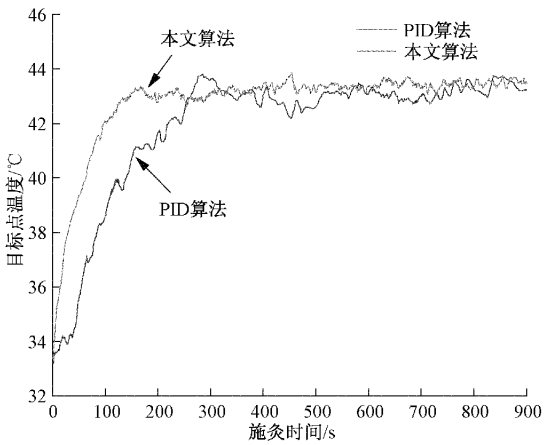

图 8　PID 算法温度响应-距离变化曲线

最后,传统 PID 算法被灸穴位处皮肤温度由初始温度上升到 42 ℃的爬坡时间为 230 s,而本文算法的爬坡时间为 98 s,相比传统 PID 算法爬坡时间缩短了 132 s。本文算法爬坡时间更短,增加了艾灸治疗的有效作用时间,可以使艾灸具有更好的治疗效果。

表 5　真实环境实验结果

算法名称	稳态温度/℃	稳态温度误差/±℃	超调量/℃	爬坡时间/s
PID 算法	43.26	0.7	0.8	230
本文算法	43.15	0.4	0.2	98

以上结果表明,基于 RGCA-DDQN 强化学习算法的艾灸机器人可以将施灸点温度安全快速上升到期望温度区间,并很好保持温度恒定,相比于传统 PID 算法,本文算法可以使艾灸机器人具备更好的温度控制能力。

4　结　　论

针对传统 PID 控制算法参数辨识困难、对不同被控对象缺乏自适应性,以及 DDQN 算法存在奖励稀疏以及容易陷入局部最优等问题,本文提出了一种结合奖励引导机制和余弦退火策略的强化学习控制算法 RGCA-DDQN。与传统 DDQN 算法相比,RGCA-DDQN 算法在收敛速度、收敛稳定性和成功率等方面都有显著提升。与传统 PID 算法相比,本文算法的温度超调量更小,降低了温度超调大可能导致烫伤的风险;并且本文算法稳态温度误差更小,温度响应更平稳,对施灸对象的适应性更好;温度响应的爬坡时间更短,进一步提升了艾灸治疗效果。在今后的研究中,通过艾灸机器人从真实环境采集更多的温度状态数据并保存在经验池中,发挥出强化学习的可持续学习能力,使艾灸机器人具有更好的适应性和控制效果。

参考文献

[1]　张建斌,王玲玲,胡玲,等. 艾灸温通作用的理论探讨[J]. 中国针灸,2011,31(1):51-54.

[2]　程洁,张玉飞,田元生. 针刺联合艾灸涌泉治疗顽固性口腔溃疡案[J]. 中国针灸,2021,41(11):1248.

[3]　郭娅静,孙怡,周竞,等. 一种头部保健艾灸器的设计和应用[J]. 上海针灸杂志,2019,38(7):818.

[4]　赵国友,刘宜成,涂海燕,等. 艾灸机械臂的设计与应用[J]. 针刺研究,2020,45(11):936-940.

[5]　姚勇,毛雷,邢嘉祺,等. 智能艾灸手法模拟系统的设计与实现[J]. 北京生物医学工程,2020,39(6):588-593.

[6]　张艳. 一种智能艾灸仪设计方案[J]. 信息技术与信息化,2020(12):207-208.

[7]　吴金宗,方建啟,邓正明,等. 灸疗研究进展及新型灸疗仪[J]. 中国中医药现代远程教育,2018,16(14):136-138.

[8]　赵鑫,高明,张一鸣,等. 一种基于碳纤维的迷你红外灸疗仪的设计[J]. 中国针灸,2020,40(1):109-111.

[9] 袁浩. 共享艾灸理疗系统的设计与实现[D]. 太原:中北大学, 2019.
[10] 邓丽, 黄炎, 费敏锐, 等. 改进的广义预测控制及其在温度系统中的应用[J]. 仪器仪表学报, 2014,35(5):1057-1064.
[11] 徐新, 邓斌, 王奇, 等. 基于自整定模糊 PID 的艾灸点皮肤温度控制系统[J]. 电子测量技术, 2020,43(22):39-44.
[12] 杨军, 张和生, 潘成. 交通信息采集传感器网络基于强化学习的路由[J]. 电子测量与仪器学报, 2012,26(12):1086-1090.
[13] 卢笑, 曹意宏, 周炫余, 等. 基于深度强化学习的两阶段显著性目标检测[J]. 电子测量与仪器学报, 2021,35(6):34-42.
[14] 王军, 杨云霄, 李莉. 基于改进深度强化学习的移动机器人路径规划[J]. 电子测量技术, 2021,44(22):19-24.
[15] 石欣, 田文彬, 冷正立, 等. 基于 CFD 和 LightGBM 算法的建筑室内温度全局预测模型[J]. 仪器仪表学报, 2021,42(1):237-247.
[16] KARIM F, MAJUMDAR S, DARABI H, et al. LSTM fully convolutional networks for time series classification[J]. IEEE Access, 2018, 6(99):1662-1669.
[17] 许培昌, 李达良, 崔淑丽. 不同施灸距离对人体体表皮肤温度的影响——论施灸的安全距离[J]. 中国针灸, 2012,32(7):611-614.
[18] 路玫, 张丽繁, 袁晔, 等. 隔姜灸、悬灸对不同穴位各时段热感度的对比研究[J]. 中国针灸, 2011,31(3):232-235.
[19] 杨惟轶, 白辰甲, 蔡超, 等. 深度强化学习中稀疏奖励问题研究综述[J]. 计算机科学, 2020,47(3):182-191.
[20] 尹旷, 王红斌, 方健, 等. 基于强化学习的移动机器人路径规划优化[J]. 电子测量技术, 2021,44(10):91-95.
[21] 李珊, 任安虎, 白静静. 基于 DQN 算法的倒计时交叉口信号灯配时研究[J]. 国外电子测量技术, 2021,40(10):91-97.
[22] 周攀, 张建斌, 王玲玲, 等. 不同灸温的艾灸抗炎效应及 TRPV1 作用机制研究[J]. 中国中医基础医学杂志, 2015,21(9):1143-1145.
[23] 高希言, 陈岩, 王鑫, 等. 腹部透灸时温度变化的研究[J]. 中国针灸, 2015,35(1):45-49.
[24] 夏世林, 佃松宜, 张洽芮, 等. 一种适用于多关节艾灸机械臂艾灸器的设计与应用[J]. 中国针灸, 2021,41(2):221-224.

作者简介

张博,硕士研究生,主要研究方向为智能机器人控制。
E-mail: 1432421513@qq.com

涂海燕(通信作者),博士,副教授,硕士生导师,主要研究方向为智能机器人控制、生物医学工程。
E-mail: haiyantu@163.com