文章编号:1671-4598(2018)09-0067-05 DOI:10.16526/j.cnki.11—4762/tp.2018.09.015 中图分类号:TP13 文献标识码:A

改善室内环境舒适度的一种新型控制方法

葛 双,付保川,许馨尹
(苏州科技大学 电子与信息工程学院,江苏 苏州 215009)

摘要:室内环境舒适度直接影响着在室人员的健康和工作效率,需要进行有效的控制以保证其舒适性;论文以某办公室为研究对象,提出一种基于同策蒙特卡罗自适应控制方法对建筑物内空调系统、加湿器、除湿器、照明系统和通风系统等设备优化控制,调节室内温湿度、照度和CO_2浓度提高室内舒适感;建立了有关影响室内环境舒适度的因素的状态变化的数学模型,并利用实验数据验证模型,同时与PID和模糊控制算法进行了仿真对比分析;结果表明:在同时控制影响室内环境的多个因素时,同策蒙特卡罗控制方法能够较好的跟踪实际情况,且该方法稳态误差较小,收敛速度较快。

关键词:室内环境;舒适度;优化控制;仿真分析

A New Control Method to Improve Comfort of Indoor Environment

Ge Shuang, Fu Baochuan, Xu Xinyin
(Suzhou University of Science and Technology, Suzhou 215009, China)

Abstract: Indoor environment comfort directly affects the health and work efficiency of staff in the room. Effective control is needed to ensure its comfort. Taking a certain office as the research object, this paper proposes a Monte Carlo adaptive control method based on the policy to optimize the control of air conditioning system, humidifier, dehumidifier, lighting system and ventilation system in the building and adjust the indoor temperature and humidity, illumination and CO_2 concentration to improve indoor comfort. The mathematic model of the change of the factors affecting the comfort of the indoor environment is set up, and the model is validated by the experimental data. At the same time, the simulation and comparison with the PID and fuzzy control algorithms are carried out. The results show that the On-policy Monte Carlo control method can track the actual situation better while controlling a number of factors that affect the indoor environment at the same time. The steady-state error of this method is small and the convergence speed is fast.

Keywords: indoor environment; comfort; optimization control; simulation analysis

0 引言

室内环境对人体的舒适度有着决定性作用,经济发展和生活水平不断提高使得室内环境问题日益突出[1]。据国内外学者研究发现,若室内环境质量得以改善,其室内工作人员的效率将提高15%~20%[2]。而在室内环境中,室内的热湿环境、光环境和空气品质对人的影响尤为突出。因此通过对室内设备进行调节控制提高室内热湿环境、光环境和空气品质也就意味着提高了在室人员的舒适度。

环境舒适度的控制优化研究,Diouns AI等学者利用一种新的算法Fuzzy-PD用来控制建筑内的有关设备并提高室内舒适度,通过仿真实验表明该算法较传统的模糊算法性更好[3]。FJ Lin等学者结合模糊控制和神经网络,利用该方法对系统输入的参数进行跟踪并对该系统进行控制[4]。段永培等通过改进粒子群算法寻解被控系统的最佳参数,实现动态舒适度的最优控制[5]。刘运城将模糊规则与双线性控制算法相结合对室内温度进行控制,实验结果表明该算法的鲁棒性和稳定性较好[6]。除以上方法外,强化学习和深度学习算法也逐渐用于智能控制领域,Cho S H提出了一种对建筑暖通空调进行控制基于强化学习的控制器的可能性,并对其进行了理论分析[7];Dalamagkids K等学者则通过对强化学习算法改进,提出了一种基于递归最小乘法的强化学习控制器用于建筑设备的控制[8];Bielskis等学者提出一种基于强化学习的室内照明控制器,该算法可自动调节照明系统从而优化室内光环境的同时减少照明能耗[9];Li等学者提出一种能在线学习最优控制策略的多网络Q学习方法用于建筑节能,实验表明该算法的收敛速度比未改进的Q学习算法快[10]。综上可看出,越来越多的学者注重室内环境的质量,运用在建筑设备的控制技术手段也层出不穷。

在研究控制建筑内相关系统时,常见的方法如模糊控制、PID控制[11-12]等,这些传统方法在控制较为复杂的系统或多个被控对象时存在收敛速度慢或者收敛性能较差等缺点。本文提出了一种基于同策蒙特卡罗(On-policy Monte Carlo,OMC)算法的控制器,用于控制建筑内的相关设备,在提供室内人员最基本环境需求的同时提高室内环境舒适度。蒙特卡罗算法是强化学习里的一种算法,通过状态和动作得到奖赏值从而评估策略的好坏[13-14]。

1 基于同策蒙特卡罗算法的优化控制

强化学习就是学习如何将场景映射到动作,以获取最

收稿日期:2018-02-05; **修回日期**:2018-03-12。
基金项目:国家自然科学基金(61672371)。
作者简介:葛 双(1993-),女,江苏南通人,硕士生,主要从事建筑智能化方向的研究。

大数值奖赏信号[15]。强化学习解决问题的过程简单说就是一个智能体（Agent）采取行动（Action）从而改变自身状态（State）获得回报值（Reward）并与环境（Environment）不断的发生交互的一个过程。强化学习包括多种不同的算法，是否需要模型是区别这些算法的一种重要特征，其中同策蒙特卡罗方法是一种不需要模型仅需要经验的算法——从与环境在线或模拟交互中获得状态、动作和回报。

影响室内舒适度的因素主要有室内的热湿环境、光环境和室内空气质量。需要考虑的状态因素有：室内温度，二氧化碳浓度、相对湿度和照度。室内温度需要空调设备对其进行调节控制；二氧化碳浓度通过通风系统进行控制；湿度的改变需通过加湿器和除湿器；调节照度需通过照明设备进行控制。所以动作因素有：空调系统、通风系统、加湿器、除湿器和照明系统这几个设备的运行情况。

改善室内环境舒适度需从室内的热湿环境、光环境和空气质量等因素进行分析。室内的热湿环境中干球温度和相对湿度对人舒适度影响最为突出；室内光环境取决于室内照度情况；而室内空气质量二氧化碳浓度对人舒适度的影响比重最大。对于 Agent，假设外部环境为一个独立的只包含温湿度、照度、二氧化碳浓度这 4 个参数的普通办公室。故涉及到的参数有：室内温度 T（℃），设定范围为 $[T_{\min}, T_{\max}]$，$T_{\min}$是设定的最小温度值，$T_{\max}$是设定的最大温度值；室内相对湿度 h（HR）（相对湿度是用百分比表示的，为了简化参数本文中直接用整数表示）设定范围为 $[h_{\min}, h_{\max}]$，$h_{\min}$是最小值，$h_{\max}$是设定的最大值；室内照度 I（Lx），范围为 $[I_{\min}, I_{\max}]$；室内二氧化碳浓度 ρ_{co_2}(ppm)，设定范围为 $[\rho_{\min}, \rho_{\max}]$。若这些参数的值都超过上述设定的最大值，人将感觉不舒适。为了满足人对环境舒适度的要求，各个参数都需设置一个舒适值并保证该设定值都在给定的范围内。

2 算法框架建模

本文中的被控对象为空调、可调光的照明设备、通风系统、加湿器和除湿器。环境状态的改变需要通过被控对象的状态的改变才能实现。被控对象即被控设备如空调系统、通风系统等根据当前的环境状态对设备动作进行选取从而改变设备状态，如图 1 所示。

$$s_0 \xrightarrow{a_0} s_1 \xrightarrow{a_1} s_2 \xrightarrow{a_2} s_3 \cdots\cdots \xrightarrow{a_{n-1}} s_n$$

图 1 状态更新示意图

某一时刻的状态如室内温湿度、CO_2 浓度和照度等；根据当前时刻的环境状态通过策略选择器和动作选择器产生下一时刻的动作，包括空调系统、通风系统、照明系统、除湿器和加湿器等设备的动作；再通过动作执行器对策略进行评估改进，直至判断是否为最优策略。其基本流程图如图 2 所示。

每个时间步 t，agent 都得到若干环境状态 $s_t \in S$，其中 S 是所有可能状态的集合，在此基础上根据策略（状态到动作的映射，$s \to a$）选择一个动作 $a_t \in A(s_t)$，其中 $A(s_t)$ 是可选动作的集合。一个时间步后，agent 得到一个奖赏值 $r \leftarrow s \times a$，并得到下一个状态 s_{t+1}，根据奖赏值进行策略的评估与改进。

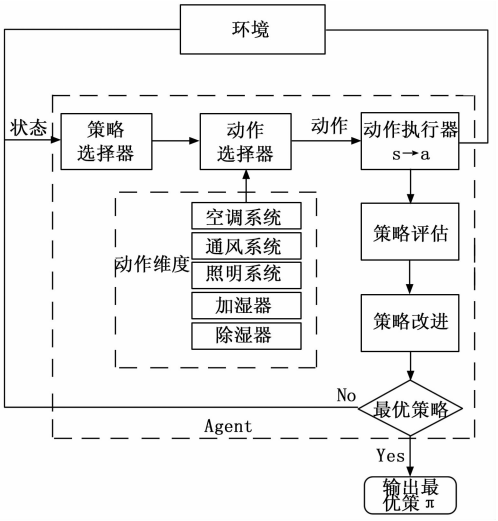

图 2 算法框架流程图

2.1 算法框架设计

算法中的关改变设备状态的动作 a 被建模为一个矩阵。水平维度是五维向量，用来表示各个不同设备的动作。第一维 AC（Air Conditioning）表示空调的动作，可以用 $a_1 = [a_{10}, a_{11}, a_{12}, a_{13}, a_{14}]$ 的向量表示，共有 5 种动作：0 表示关闭，1 表示热风（小风），2 表示冷风（小风），3 表示热风（大风），4 表示冷风（大风）。第二维 VS（Ventilation system）表示通风系统的动作，通风系统的动作向量表示为 $a_2 = [a_{20}, a_{21}, a_{22}]$，共 3 种动作：0 表示关闭，1 表示小档位，2 表示大档位。第三维 H（Humidifier）表示加湿器的动作 $a_3 = [a_{30}, a_{31}, a_{32}]$，共 3 种动作：0 表示关闭，1 表示小档，2 表示大档。第四维 DH（Dehumidifier）表示除湿机的动作：0 表示关闭，1 表示小档，2 表示大挡；除湿器的动作可用向量 $a_4 = [a_{40}, a_{41}, a_{42}]$ 表示。最后一位 L（Light）表示灯的动作，照明设备的动作向量 $a_5 = [a_{50}, a_{51}, a_{52}]$ 表示：0 表示关闭，1 表示提高照度，2 表示降低照度。

OMC 中的环境状态 $s = [T_t, h_t, \rho_t, I_t]$ 这几个参数组成，见式（1）到式（5）。

$$T = \frac{|T_t - T_s|}{T_{\max} - T_s} \tag{1}$$

在式（1）中，T_s 是设置的最舒适温度，$T_{\max}$ 是在范围内的最大值。

$$h = \frac{|h_t - h_s|}{h_{\max} - h_s} \tag{2}$$

h_s 为设室内最合适的相对湿度为，如式（2）所示，分母表示取值范围的最大值 $h_{\max}$ 减去最适湿度值 h_s 的差。

$$I = \frac{|I_t - I_s|}{I_{\max} - I_s} \tag{3}$$

设在这间独立的普通办公室内，照度参考平面及其高度为 0.75 m 水平面，I_s 表示的是设置的室内最佳照度，

I_{max}是设定的最大照度值，照度若超过 I_{max}人眼会感觉不舒适，在式（3）中，分母表示两者之差。

$$\rho_{\omega_2}=\frac{|\rho_t-\rho_s|}{\rho_{max}-\rho_s} \tag{4}$$

在式（4）中，ρ_s 是设定的目标值，是室外 CO_2 浓度可以达到的最低水平；ρ_{max} 是设定的最大值，若超出该值舒适感则会消失。

$$r=-w_1(T)-w_2(h)-w_3(I)-w_4(CO_2) \tag{5}$$

r 值是系统最终的评估标准。在本文中，r 的值被控制在［−1，0］之间，式（5）表示各个参数在不同权重下奖赏值的叠加。在式（1）至（4）中，各个参数的取值偏离设定值越大，r 值就越接近−1（越小），反之越大；所以式（5）中用负号来表示。这里的权重 $w=[0.6,0.1,0.1,0.2]$ 是通过多次实验得到的，这确保了 r 值在［−1，0］之间，并能使系统保持良好的性能。

本算法中状态转移函数如式（6）到（10）。式（6）表示的是温度随时间的变化，但在空调运行时，打开通风系统会影响室内温度，所以在等式中表现通风系统对温度的影响是加入一个弱化参数 0.2。T_c 表示的是温度变化率，它与空调产生的风的强弱有关，见式（7）。式（8）、（9）、（10）分别表示的是湿度、CO_2 浓度和照度的状态转移函数。

$$T(t+1)=T(t)-[(-1)^{AC/2}\times T_c\times(1-0.2\times VS)] \tag{6}$$

$$T_c=\begin{cases}0.001 & weak\\ 0.002 & stong\end{cases} \tag{7}$$

$$h(t+1)=h(t)+0.1\times H-0.1\times DH \tag{8}$$

$$\rho(t+1)=\rho(t)-0.2\times VS \tag{9}$$

$$I(t+1)=I(t)+(-1)^{L\%2}\times 0.1\times L \tag{10}$$

2.2　控制算法

算法流程：

1）初始化 $r=0$，动作 a

2）对于每个情节，初始化状态

$$s_0(T_0,h_0,\rho_0,I_0)$$

3）根据状态转移函数确定下一时刻的状态 s'

4）根据式 1）至 5）更新 r 值

5）对情节中的每个状态 s：

$$a'\leftarrow \mathrm{argmax}_a r(s,a)$$

6）重复每个情节，直至 s 满足终止条件

在强化学习里这个问题没有确定的终止条件，所以为了方便实验，需设置确定的情节数，并置每个情节有 N 个单位时间步数，当 $t+1=N$ 时，结束运行一个情节。

3　仿真结果分析

3.1　仿真步骤

本文中使用了 OMC 算法优化室内环境舒适度，控制室内的相关设备。为了验证该算法的有效性，在 Python2.7 环境中做了仿真实验，具体步骤如下：

步骤 1：建立奖赏函数如公式（1）～（5）、状态转移函数如公式（6）～（10）。

步骤 2：初始化动作值函数 $Q(s_t,a_t)$、学习率 α 和折扣率 γ。其中，s 是状态参数，由室内温度 T_t、室内二氧化碳浓度 ρ_t、室内照度 I_t、室内湿度 H_t 和实时能耗 E_t；a 是动作参数，由空调系统动作、照明系统动作、加湿器和除湿器动作和通风系统动作构成。学习率和折扣率根据经验得到：$\alpha=0.1$，$\gamma=0.9$。

步骤 3：对于每个情节，设置情节的参数包括 $N=4\,000$个单位时间步，令 $t=0$，也就是使各个状态和动作性参数保持初始状态。

步骤 4：在每个情节中每个时间步的运行包括对当前状态 s_t，计算出在该时刻下动作因素 a_t；当采取这个时刻的动作时，根据建立的状态转移函数计算该状态的转移情况，得出下一刻相应的状态 s_{t+1}；然后根据上述建立的奖赏函数公式，计算出在当前状态 s_t 和动作 a_t 下的奖赏值 r_t。

步骤 5：判断终止条件，如下：

对观察所有状态因素下的动作值函数的值判断是否是预设值，若不满足则返回到步骤 3 进行新的情节的运行，若满足则结束循环。

3.2　实验结果分析

本章节主要验证了同策蒙特卡洛控制算法的有效性并将该算法与 PID 控制和模糊控制的收敛性能进行了比较。在本文中，设置了 200 个情节。并将每个情节的步数设置为4 000步。

参考实际情况，各个参数设定的范围为：室内温度 T（℃），设定范围为［0，40］；室内湿度 h（HR），设定范围为［0，100］；室内照度 I（Lx），设定范围为［0，800］；室内二氧化碳浓度 ρ(ppm)，设定范围为［200，1 000］。设定满足室内舒适度时各个参数值为：温度 25 ℃、湿度 50 HR、CO_2 浓度 300 ppm、照度 300 lx。本文做了多组实验，选取其中两组实验说明该算法的收敛性能。实验 a 各个参数设置的初始状态为 $s_a=[35,70,700,100]$，实验 b 的初始状态为 $s_b=[10,20,850,600]$。实验数据如图 3、图 4、图 5、图 6、图 7、图 8 所示。

室内热湿环境是影响室内舒适度的一个重要影响因素，图 3 和 4 分别表示随着步数的增加，在不同控制算法下室内温度和湿度的变化情况。

图 3 是两组温度收敛变化实验图，由图 3（a）可知，OMC 方法在 1 800 步左右收敛到设定的参数值即 25 ℃，并能保持在这个值，具有良好的精度和稳定性；而实验 b 改变了初始状态值，大约在 2 200 步达到收敛预设值，其收敛的效果和实验 1 是一样的。相比较而言，在两组实验中 PID 算法和模糊算法虽然在达到预设值前有更平滑的下降或上升趋势，但是 PID 算法在收敛后的稳定性较差，在设定的温度值上下浮动；而模糊算法的稳定性较好但收敛精度较差，并不能完全收敛到预设值。实验表明，OMC 比 PID 算法和模糊算法具有更好的稳定性和收敛精度。图 4 是两组室内湿度收敛实验图。图 4（a）大约在 1 000 步达到收敛效果，因改变了湿度的初始设定值，图 4（b）大约在 1 600 步收敛到设定的最适湿度值 50 HR；从图 4 中可以看出 PID

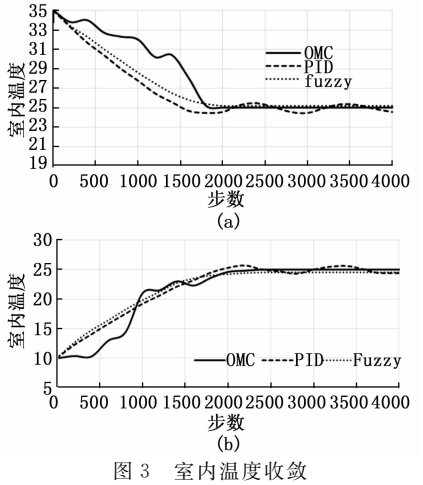

图 3 室内温度收敛

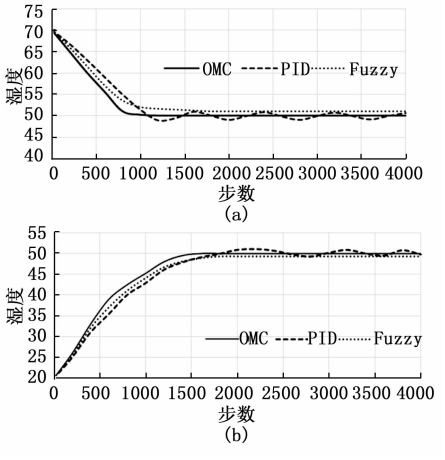

图 4 室内湿度收敛

在实验（a）和实验（b）中大约分别在 1 200 和 1 800 步达到预设的舒适值，但并不能稳定在预设的舒适值；模糊算法分别大约在 1 200 步和 1 600 步开始收敛，但其收敛精度不高。实验结果表明：OMC 方法的性能优于 PID 和模糊控制，能给室内提供良好舒适的热湿环境。

图 5 表示的是随着步数的增加，室内 CO_2 浓度在不同算法控制下的变化情况。通过 CO_2 浓度的不同显示了室内空气质量的品质的高低。

图 5 是 CO_2 浓度的变化图。两个实验的区别在于实验（a）设定的初始状态值不同，实验（a）中设定的初始值比实验（b）的设定的值要低一些。图 5（a）中 OMC 大约在 1 000步收敛，图 5（b）大约在1 400步达到收敛效果。PID 在图 5（a）中大约在1 200步收敛，图 5（b）在1 600步收敛；Fuzzy 在实验（a）中1 000步左右开始收敛，在实验（b）中 1 600步左右收敛。将 OMC 算法与这两种算法相比较，这两种算法的收敛速度和精度都差于 OMC。实验结果表明：OMC 方法在保证拥有良好的室内空气环境同时能在更短的时间内对通风系统进行调节控制，提高室内空气品质。

室内的光环境对在室人员的舒适感也有较大影响，图 6 表示的是随着步数的增加，室内照度在不同控制算法下的变化情况图。

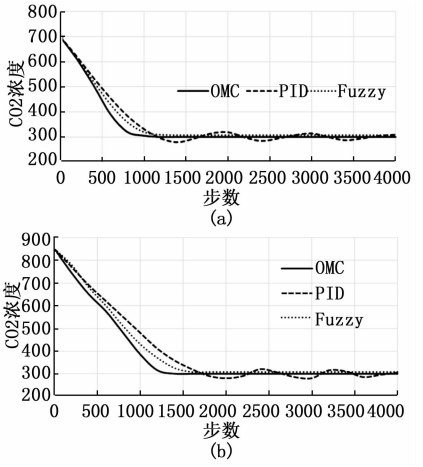

图 5 CO_2 浓度收敛

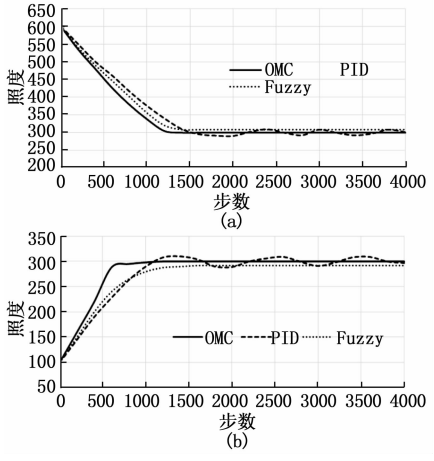

图 6 室内照度收敛

图 6 是控制照度的变化曲线图，由图可知 OMC 方法在第一组实验中的收敛步数大约为 1 000，在第二组实验中的收敛步数大约为 1 400 步。使用 PID 算法，图 6（a）显示在 1 200步左右收敛，图 6（b）在 1 500 步左右收敛并在 300 lx 左右波动；通过 Fuzzy 控制照明系统，如图所示，实验（a）在 1 200 左右开始收敛，实验（b）在 1 600 左右收敛且收敛值与预设值有一定偏差。实验结果表明实验 OMC 算法进行控制，更能提供室内良好的光环境。比较两组实验中各个参数使用 OMC 方法的收敛曲线图，室温的收敛时间最长，其原因可能与通风系统和室内湿度环境有关。更多参数和动作的加入意味着需要更复杂的控制过程和收敛步骤。从上述几组图中，对比 OMC 方法和 PID、Fuzzy 算法，发现 OMC 方法的收敛速度与精度更好。

图 7 是本实验中 200 个情节的奖赏值的收敛变化图。实验（a）在前 50 个情节，回报收敛的波动较大，振动幅度大于2 000，在此期间 agent 处在试错阶段；经过前 60 个情节的学习，回报值渐渐稳定在－7 000 左右。实验（b）是第二组实验过程中的回报收敛图，大约经过 100 个情节的学习，回报值渐渐稳定在－1 300 左右。

图 8 是收敛步数图，表示的是 200 个情节中每个情节的

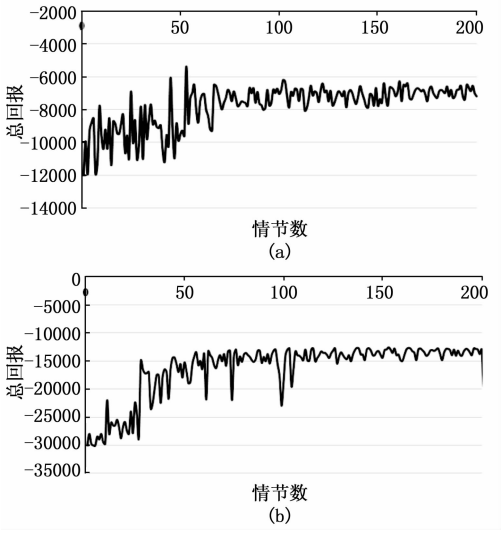

图 7　200 个情节的回报值变化

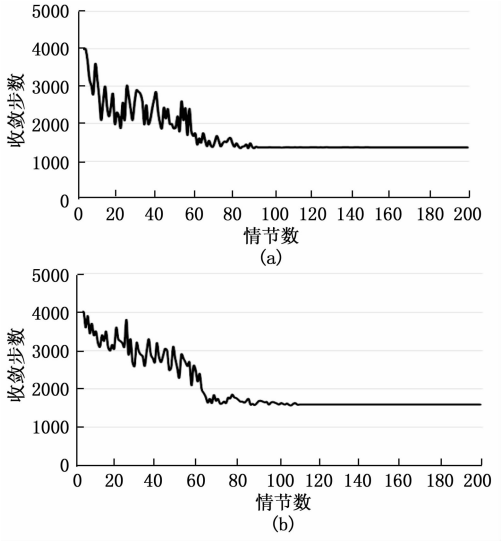

图 8　200 个情节的收敛步数

收敛步数的变化情况。从第一个图中看出，在一开始的几个情节里，收敛步数维持在初设值并没有发生变化，之后情节的收敛步才开始改变。从图中可看出，收敛步数发生较大变化大约在 3～60 个情节之间，说明 OMC 在这个阶段处于学习阶段；大约在 60～90 个情节之后，系统的震荡幅度较小，此时 OMC 处于调整阶段；在 90 个情节之后 OMC 收敛在 1 400 步左右，说明系统找到最优策略。图 8（b）在 70 个情节前震荡幅度较大，在 70～110 个情节期间震荡较小，在 110 个情节后达到收敛，大约在 1 600 步左右。

4　结束语

为了提高人在室内的舒适感，采用基于同策蒙特卡罗算法控制办公室内的空调系统、加湿器、照明系统和通分系统等设备，并对这些设备进行了简单的模型构建。对输入的温湿度、照度和二氧化碳浓度等参数进行智能调整，进而将各个参数值控制在设定的最适值以优化室内舒适度。本文基于构造的模型进行了仿真实验，实验结果表明：（1）该方法在不同参数设置下都能达到良好的收敛性和稳定性，能很好的改善室内环境舒适度；（2）在控制建筑设备等方面，和 PID 算法和模糊控制方法进行比较，发现该算法具有收敛速度较快、鲁棒性好、精度较高等优势。

参考文献：

[1] 杨　娜. 用户偏好的室内环境舒适度智能控制方法仿真研究 [J]. 科学技术与工程，2013，13（25）：7557－7562.

[2] 曹　彬，朱颖心，欧阳沁，等. 公共建筑室内环境质量与人体舒适性的关系研究 [J]. 建筑科学，2010，26（10）：126－130.

[3] Dounis A I，Santamours M J，Lefas C C，et al. Design of a fuzzy-set environment comfort system [J]. Energy and Buildings，1995，22（1）：81－87.

[4] Lin F J，Hwang W J，Wai R J. A supervisory fuzzy neural network control system for tracking periodic inputs [J]. IEEE Transactions on Fuzzy Systems，1999，7（1）：41－52.

[5] 段培永，刘聪聪，段晨旭，等. 基于粒子群优化的室内动态热舒适度控制方法 [J]. 信息与控制，2013，42（1）：100－110.

[6] 刘运城. 智能建筑室内环境恒温优化控制仿真研究 [J]. 计算机仿真，2017，34（1）：318－321.

[7] Cho S H. Application Study of Reinforcement Learning Control for Building HVAC System [J]. International Journal of Air Conditioning and Refrigeration，2006，14（4）：138－146.

[8] Dalamagkids K，Kolokotsa D. Reinforcement learning for Building Environment Control [M]. INTECH Open Access Publisher，2008：283－294.

[9] Bielskis A A，Guseinoviene E，Dzemydiene D，et al. Ambient Lighting Controller Based on Reinforcement Learning Components of Multi－Agents [J]. Elektronika Ir Elektrotechnika，2012，121（5）：79－84.

[10] Li B，Xia L. A multi－grid reinforcement learning method for energy conservation and comfort of HVAC in buildings [A]. IEEE International Conference on Automation Science and Engineering [C]. IEEE，2015：444－449.

[11] Egilegor B，Uribe J P，Arregi G，et al. A Fuzzy control adapted by a neural network to maintain a dwelling within thermal comfort [J]. Proceedings of Building Simulation，1997，97：87－94.

[12] Ulpiani G，Borgognoni M，Romagnoli A，et al. Comparing the performance of on/off，PID and fuzzy controller applied to the heatings system of an energy efficient building [J]. Energy and Buildings，2016，116：1－17.

[13] Zhou Y，Yan Y，Huang X. Error Analysis of Closed Loop Control for Spacecraft Orbital Transfer Using Monte Carlo Method [J]. Advanced Materials Research，2013，765（767）：1998－2003.

[14] Qian X. Control－limit Policy of Condition－based maintenance Optimization for Multi－component System by Means of Monte Carlo Simulation [J]. Iaeng International Journal of Computer Science，2014，41（4）：269－273.

[15] Sutton R S，Barto A G. Reinforcement Learning：An Introduction [M]，Cambridge：MIT Press，1998.