文章编号: 1001-0920 (2013) 06-0930-05

操作条件反射学习自动机及其在机器人平衡控制中的应用

郜园园, 阮晓钢, 宋洪军

(北京工业大学 电子信息与控制工程学院，北京 100124)

摘 要: 针对两轮机器人的平衡控制问题, 在学习自动机理论的框架中, 提出一种基于操作条件反射学习自动机的仿生学习模型. 该模型引入认知学习单元和取向单元, 分别用来实现操作行为学习和指导系统进化的方向. 模拟两轮自平衡机器人的平衡控制仿真实验表明, 该学习模型具有可行性和有效性, 能使机器人自主学会平衡控制技能, 并使其具有高度的自适应能力.

关键词: 操作条件反射；学习自动机；仿生；机器人平衡控制

中图分类号: TP273 **文献标志码:** A

Operant conditioning learning automatic and its application on robot balance control

GAO Yuan-yuan, *RUAN Xiao-gang*, *SONG Hong-jun*

(College of Electronic Information and Control Engineering，Beijing University of Technology，Beijing 100124, China．Correspondent：GAO Yuan-yuan，E-mail：yuangao84@163.com)

Abstract: A biomimetic learning model is proposed based on the operant conditioning learning automatic(OCLA) within the structure of learning automatic theory for the balance control of a two-wheeled robot. A cognitive learning unit and a tropism unit are introduced in this model, and they are used to evaluate the operant behavior learning and to direct the evolution of the system respectively. Finally, the simulation experiments on the two-wheeled robot show the feasibility and effectiveness of the proposed algorithm, and the robot learns the ability of balance control and has high self-adaptive ability.

Key words: operant conditioning；learning automatic；biomimetic；robot balance control

0 引 言

两轮自平衡机器人是典型的欠驱动系统, 近年来, 针对两轮机器人的运动平衡控制问题已提出了较多控制方法, 如PID控制、LQR控制、模糊自适应控制、鲁棒控制和强化学习等. 但是, 这些方法均以传统的控制理论和控制技术为主, 导致动态性能和静态性能都较差.

操作条件反射(OC)[1]是一种重要的条件反射理论, 视为生物系统最基本的学习形式, 这是因为人或动物的平衡控制技能在较大程度上是基于这种学习机制自组织地渐近形成、发展和完善的. 国内外许多学者对于OC的仿生学习模型进行了相关研究, 期望这种模型能够复制动物学习操作或控制的实验. 也有学者以概率自动机为平台模拟操作条件反射机制, 设计了相应的仿生系统, 并成功实现了倒立摆和机器人的平衡控制[2-3]. 但是, 这些计算理论和计算模型没有给出具体的数学计算模型, 不具备泛化能力, 应用受到限制[4-7], 所以研究具有更加普适性的操作条件反射仿生学习模型具有重要的研究价值.

学习自动机(LA)可以作为机器人特别是认知发育机器人的数学抽象和形式化的工具, 近年来在机器人学中的应用研究逐渐增多. LA可以描述机器人及其环境[8], 用于机器人环境勘测、绘制环境地图[9], 帮助机器人进行行为选择[10]和最优控制[11]等. 针对两轮机器人的平衡控制问题, 以学习自动机为框架建立操作条件反射机制的数学模型是一个重要的研究思路, 也是本文的主要研究内容.

基于上述现状, 本文将学习自动机与操作条件反射的思想相结合, 设计一种仿生的学习模型, 并将其应用于两轮机器人的平衡控制问题. 认知学习单元主

收稿日期: 2012-02-01；**修回日期:** 2012-05-02.

基金项目: 国家自然科学基金项目(61075110)；国家863计划项目(2007AA04Z226)；北京市自然科学基金项目(4102011)；北京市教委重点项目(KZ201210005001).

作者简介: 郜园园(1984−), 女, 博士生, 从事智能控制、机器学习的研究；阮晓钢(1958−), 男, 教授, 博士生导师, 从事机器人、神经网络等研究.

要执行操作行为实现功能, 以逼近动力学系统的非线性部分; 取向单元执行操作行为的评价功能, 利用评价机制产生的取向性信息对操作行为产生网络进行调整, 使机器人具有仿生的特性. 通过实现机器人的平衡任务仿真实验, 表明了该算法的可行性和有效性.

1　OCLA 算法的设计

1.1　OCLA 的框架结构

OCLA 是一个离散时间系统, 它是一个有限状态自动机, 如图 1 所示, 由输入符号集合 A、内部状态集合 S、内部操作集合 O、输出符号集合 Z、随机"条件-操作"规则集合 R、状态空间单元 f、取向单元 ψ 和认知学习单元 δ 组成. 系统开始随机执行一个操作行为动作, 由当前时刻输入符号集合、内部状态集合和内部操作集合共同决定系统下一时刻的状态, 不断修改或补充系统内部状态集合 S. 通过与环境交互, 利用评价单元反馈该行为动作执行的好坏, 并将该信息传递给认知学习单元, 通过概率选择机制进行系统的学习与训练, 不断修改或补充规则集合 R 来满足外界未知的环境信息, 实现状态到行为的最佳映射, 从而使系统得到进化或发育.

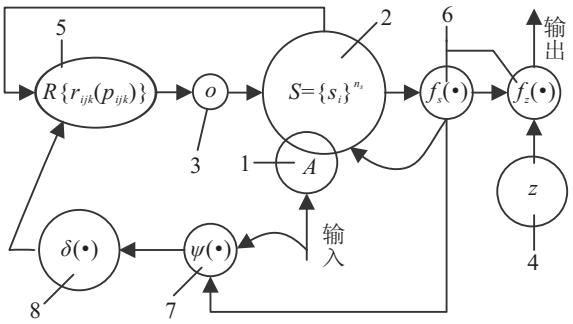


图 1　操作条件反射学习自动机结构

定义 OCLA 为八元组模型 $\mathrm{OCLA}=\langle A,S,O,Z,R,f,\psi,\delta\rangle$. 其中:

1) $A=\{a_j|j=1,2,\cdots,n_A\}$, a_j 为 OCLA 第 j 个输入符号, n_A 为输入集合的维数.

2) $S=\{s_i|i=1,2,\cdots,n_S\}$, s_i 为 OCLA 第 i 个状态符号, n_S 为内部状态集合的维数.

3) $O=\{o_k|k=1,2,\cdots,n_O\}$, o_k 为 OCLA 第 k 个操作符号, n_O 为内部操作集合的维数.

4) $Z=\{z_m|m=1,2,\cdots,n_Z\}$, z_m 为 OCLA 第 m 个输出符号, n_Z 为输出集合的维数.

5) R=$\{r_{ijk}\}$, R 的每一个元素 r_{ijk}∈R 代表一条随机"条件-操作"规则, r_{ijk} : $s_i\times a_j\to o_k$ (p_{ijk}), 即 OCLA 在状态为 s_i、输入为 a_j 的条件下依概率 p_{ijk} 实施操作 o_k, $p_{ijk}=p(o_k|\,s_i,a_j)$ 为规则 r_{ijk} 的激发概率.

6) OCLA 的状态空间方程为

$$f:\begin{cases} f_S:\ S(t)\times A(t)\times O(t)\to S(t+1),\\ f_Z:\ S(t)\times A(t)\times O(t)\to Z(t+1),\end{cases}$$

f_S 为状态转移方程, f_Z 为输出方程.

7) ψ : $S\times A\to[h,q]$, h 为取向性最差的取向函数值, q 为取向性最好的取向函数值(在生物学意义上, 生物的取向性是指由环境决定生物进化的方向). p 和 q 可根据所处理的具体对象进行取值. $\psi_{ij}=\psi(s_i,a_j)$ 表示处于状态 s_i 和输入 a_j 的取向值, 若 $\psi_{ij}<0$, 则 s_i 是 OCLA 在输入为 a_j 时的负取向状态; 若 $\psi_{ij}=0$, 则为零取向状态; 若 $\psi_{ij}>0$, 则为正取向状态.

8) δ : $R(t)\mapsto R(t+1)$, 设 OCLA t 时刻的状态为 $s(t)=s_a\in S$, 输入 $a(t)=a_b\in A$, 依集合 R 中随机"条件-操作"规则选中的操作为 $o(t)=o_c\in O$, 实施操作后观测到 $t+1$ 时刻的状态 $s(t+1)=s_d\in S$, 则基于操作条件反射原理, 操作集合 R 中随机"条件-操作"规则 $p_{abk}(k=1,2,\cdots,n_O)$ 的激发概率依下式进行调节:

$$\delta:\begin{cases} p_{abk}(t+1)=p_{abk}(t)-\xi(\boldsymbol{\psi}_{abc})p_{abk}(t),\ k\neq c;\\ p_{abk}(t+1)=\max\min(p_{abk}(t+1),0,1),\\ p_{abc}(t+1)=1-\displaystyle\sum_{k\neq c}p_{abk}(t+1),\ k=c.\end{cases}$$

$\boldsymbol{\psi}_{abc}$ 为取向值增量函数, $\boldsymbol{\psi}_{abc}=\psi(s_d,a_b)-\psi(s_a,a_b)$ 为 OCLA 在状态 s_a 和输入 a_b 的条件下, 实施操作 o_c 后状态转移为 s_d 时取向函数值的变化量, 由该值可以判断该操作的好坏程度. $p_{abc}(t)$ 为 t 时刻处于状态 s_a 和输入 a_b 时实施操作 o_c 的概率值 $p(o_c|s_a,a_b)$, 有

$$\xi(\boldsymbol{\psi}_{abc})=\xi(\psi(s_d,a_b)-\psi(s_a,a_b))=\lambda\boldsymbol{\psi}_{abc}/r,\quad(1)$$

$\xi(\cdot)$ 为单调增函数, $\xi(\boldsymbol{\psi}_{abc})=0$ 当且仅当 $\boldsymbol{\psi}_{abc}=0$, r 为操作行为的个数, λ 为学习率. 若 $\boldsymbol{\psi}_{abc}<0$, 则表明实施操作 o_c 且转移状态为 s_d 后取向函数值变小, 取向性变差, 相应地 $p_{abc}(t+1)<p_{abc}(t)$, 即下一时刻选择操作 o_c 的概率减小; 若 $\boldsymbol{\psi}_{abc}=0$, 则表明取向性不变, 下一时刻选择操作 o_c 的概率也不变; 若 $\boldsymbol{\psi}_{abc}>0$, 则表明取向性较好, 相应地, 选择操作 o_c 的概率也增大. $\max\min(p_{abk}(t+1),0,1)$ 可以保证 $p_{abk}(t+1)\in[0,1]$, 且 $\displaystyle\sum_{k=1}^{n_o}p_{abk}(t)=1$.

1.2　OCLA 的递归学习算法流程

OCLA 算法的具体流程如下.

Step 1: 初始化. 给定 OCLA 的初始状态 $s(0)$ 和输入 $a(0)$, 选取初始激发概率 $p_{ijk}(0)=1/r$, 由实验要求和环境确定学习率 λ、总迭代步数 T_f 和最优行为选择概率最大阈值 p_ε.

Step 2: 随机选择操作行为. 依据 OCLA t 时刻的状态 $s(t)$ 和输入 $a(t)$, 按 t 时刻状态下各操作概率值

$p_{ijk}(t)$ 的分布, 随机选择 t 时刻的操作 $o_k(t)$.

Step 3: 实施操作行为. 实施选取的操作行为 $o_k(t)$, 并依据 $f_S: S(t)\times A(t)\times O(t)\to S(t+1)$ 状态转移方程观测 $t+1$ 时刻 OCLA 的状态.

Step 4: 取向值增量计算. 依据取向单元中取向函数 ψ, 分别计算状态 $s(t)$ 和 $s(t+1)$ 的取向值, 得到取向值增量函数 $\boldsymbol{\psi}(\cdot)$.

Step 5: 操作条件反射. 由认知学习单元 δ 按照式 (1) 对随机"条件-操作"规则 r_{ijk} 激发概率 $p_{ijk}(t)$ 进行调节.

Step 6: 对外输出. 由系统的输出方程 $f_Z: S(t)\times A(t)\times O(t)\to Z(t+1)$ 对外输出 $Z(t+1)$.

Step 7: 条件停止. 重复 Step 2 ~ Step 6, 直至达到迭代学习次数 T_f 或 $p_{abc}(t+1)>p_\varepsilon$.

2 算法收敛性分析

引理 1 设 OCLA $=\langle A,S,O,Z,R,f,\psi,\delta\rangle$ 是一个操作条件反射学习自动机, 其状态转移过程 f: $S\times A\times O\to S$, $S\times A\times O\to Z$ 是确定的, 设 $p\neq 0$ 和 O_i^+ 是系统处于状态 s_i 下表现为正取向性 ($\boldsymbol{\psi}_{ijk}>0$) 的操作行为集合, 且 $O_i^+(t=0)\neq\varnothing$, $O_i^+\subset O$, 则有

$$p_{ijk}=p(O_i^+|s_i,a_j)=1,\ t\to\infty. \tag{2}$$

即当 $t\to\infty$ 时, OCLA 依概率 1 选取正取向状态转移操作.

证明 令 l 为 s_i 出现次数的序号, $p_{ijk}(l)$ 为状态 s_i 第 l 次出现时集合 O_i^+ 中随机操作被选中的概率, 初始概率为 $p_{ijk}(0)$.

因为 $O_i^+(t=0)\neq\varnothing$, 所以 $p_{ijk}(0)\neq 0$, 又有 $\Delta p_{ijk}(l)\geqslant 0$, 故 $\forall l=0,1,\cdots$, $p_{ijk}(l)\neq 0$. 因为 $p_{ijk}(l)\neq 0$, 当 $t\to\infty$ 时, OCLA 状态 s_i 出现的频次趋于无穷, 同时 $p_{ijk}(l)$ 使得 O_i^+ 中操作行为被选中的频次也趋于无穷.

假定 OCLA 在第 l 次处于状态 s_i 和输入 a_j 的条件下选择操作 $o_k\in O$, 若 $o_k\in O_i^+$, 则 $\boldsymbol{\psi}_{ijk}>0$, $\xi(\boldsymbol{\psi}_{ijk})>0$, 所以有

$$\Delta p_{ijk}(l)=p_{ijk}(l+1)-p_{ijk}(l)=\xi(\boldsymbol{\psi}_{ijk})\sum_{m\neq k}p_{ijm}(l)\geqslant 0. \tag{3}$$

$\Delta p_{ijk}(l)\geqslant 0$ 表示当 $t\to\infty$ 时 O_i^+ 被选中的频次趋于无穷. 当 $t\to\infty$ 时, $\Delta p_{ijk}(l)>0$ 的情形可发生任意多次, 因为 $p_{ijk}(l)$ 有上界且为 1, $p_{ijk}(l)$ 增加到 1 为止. □

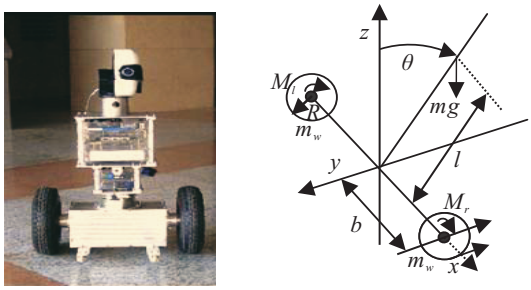
定义 1 设 OCLA $=\langle A,S,O,Z,R,f,\psi,\delta\rangle$ 是一个操作条件反射自动机, 其操作熵 H 可定义为

$$H=H(O|S,A)=\sum_{i=0}^{N_S}p_iH_i=\sum_{i=0}^{N_S}\sum_{j=0}^{N_A}p(s_i,a_j)H_i(O|s_i,a_j)=-\sum_{i=0}^{N_S}\sum_{j=0}^{N_A}p(s_i,a_j)\sum_{k=1}^{N_O}p(o_k|s_i,a_j)\log_2 p(o_k|s_i,a_j), \tag{4}$$

其中 H_i 为 OCLA 处于状态 s_i 条件下的操作熵.

3 基于 OCLA 的两轮机器人平衡控制

两轮自平衡机器人是本质上多变量、强耦合和非线性的复杂动态系统, 其核心问题是运动平衡控制. 本文以北工大人工智能与机器人研究所研制的两轮机器人为研究对象, 机器人系统见图 2, 主要物理参数见表 1.



(a) 实物图 (b) 受力分析

图 2 两轮机器人系统

表 1 两轮机器人的主要物理参数

物理量	符号	数值
机身的质量	m	10 kg
机身高度	H	0.65 m
质心距	l	0.35 m
车体绕竖直轴的转动惯量	I_m	0.589 3 kg·m^2
左右轮子的质量	m_w	1 kg
轮子半径	R	0.15 m
轮轴长度	$2b$	0.44 m
左右轮的转动惯量	I_w	0.011 3kg·m^2
重力加速度	g	9.8 m/s^2
直流电机电枢电阻	R_a	0.317 18 Ω
直流电机反电动势系数	K_e	0.030 6 V·s/rad
直流电机电磁转矩系	K_m	0.030 2 N·m/A
减速比	N	28

分别对机器人车轮、本体进行受力分析, 以牛顿经典力学为基础建立两轮机器人的数学模型. 通过对电机建模, 得到左右轮电机输出转矩 M_l 和 M_r 与电机控制电压 u_l 和 u_r 之间的关系, 分别为

$$M_l=-\frac{0.105K_m^2N^2}{R_aR}\dot{x}_+\frac{NK_m}{R_a}u_l, \tag{5}$$

$$M_r=-\frac{0.105K_m^2N^2}{R_aR}\dot{x}_+\frac{NK_m}{R_a}u_r. \tag{6}$$

通过式 (6) 和 (7) 将左右轮转矩控制转化为电压控制. 设 $X=[x\ \dot{x}\ \theta\ \dot{\theta}]^{\mathrm{T}}$, 线性化后经过整理得到系统的状态空间表达式

$$\dot{X}=\begin{bmatrix}\dot{x}\\ \ddot{x}\\ \dot{\theta}\\ \ddot{\theta}\end{bmatrix}=\begin{bmatrix}0 & 1 & 0 & 0\\ 0 & -b_1b_3/a_2 & a_1/a_2 & 0\\ 0 & 0 & 0 & 1\\ 0 & b_2b_3/a_2 & a_3/a_2 & 0\end{bmatrix}\begin{bmatrix}x\\ \dot{x}\\ \theta\\ \dot{\theta}\end{bmatrix}+$$
$$\begin{bmatrix}0 & 0\\ b_1b_4/a_2 & b_1b_4/a_2\\ 0 & 0\\ -b_2b_4/a_2 & -b_2b_4/a_2\end{bmatrix}\begin{bmatrix}u_l\\ u_r\end{bmatrix}.$$

其中

$$a_1=-2m^2l^2R^2g,$$
$$a_2=(2ml^{^2}+I_m)(mR^2+2m_wR^2+2I_w)-2m^2l^2R^2,$$
$$a_3=mgl(mR^2+2m_wR^2+2I_w),$$
$$b_1=2ml^2R+I_mR,\ b_2=mlR,$$
$$b_3=\frac{0.21K_m^2N^2}{R_aR},\ b_4=\frac{NK_m}{R_a}.$$

因为所设计的学习系统无需外部教师信号, 是一个高度的自治系统, 所以 OCLA 中的输入符号集合 A 为空集. OCLA 的八元组实际意义见表 2. OCLA 自动机的学习目标是确定一个操作行为序列, 以实现两轮机器人的姿态平衡控制.

表 2　机器人平衡控制问题中 OCLA 八元组实际意义

符号	实际意义
A	外部教师信号
S	机器人倾角和倾角速度, $n_S=3$
O	电机电压值, $n_O=3$
Z	机器人控制效果, $n_Z=2$
R	机器人在状态 s_i 和输入 a_j 条件下依概率 p_{ijk} 实施操作 $o_k(o_k\in O)$
f	机器人状态与控制效果转移方程
ψ	$\psi: S\times A\to(-1,0,1)$
δ	具体描述见式 (1)

4　仿真结果与分析

4.1　仿真实验参数设置

OCLA 自动机的仿真参数设置如下: 采样时间为 $t_s=0.01$ s, 实际状态 θ 的上限值为 $\theta_{\max}=\pi/2$, 下限值为 $\theta_{\min}=-\pi/2$, 输入状态 θ 离散化个数为 $n=5$ (θ 在 $[-\pi/2\ \ -\pi/3)$ 或 $[\pi/3\ \ \pi/2]$ 时为 s_1, θ 在 $[-\pi/3\ -\pi/6)$ 或 $[\pi/6\ \pi/3)$ 时为 s_2, θ 在 $[-\pi/6\ \pi/6)$ 时为 s_3). 实际状态 $\dot{\theta}$ 的上限为 $\dot{\theta}_{\max}=\pi$, 下限为 $\theta_{\min}=-\pi$. 将输入状态 $\dot{\theta}$ 离散化的数目等间隔分为 3 段, 初始状态 $\theta=0.1$ rad, 其他为 0. 用电压表示操作行为集合 $O=\{-5,0,5\}$, 每个行为的初始概率为 $p_{ik}(0)=1/3,\ i=1,2,3,\ k=1,2,3$. 对应初始操作行为熵为

$$H=-\sum_{k=1}^{3}p_{ik}\times\log_2 p_{ik}\big|_{p_{ik}=\frac{1}{3}}\approx 1.585.$$

此时熵值最大.

当某一行为选择概率值接近于 1 时易发生小概率事件, 从而导致系统的不稳定性增加, 故本文在设计中增加了最优行为选择概率阈值 p_ε. 该值大小接近于 1, 当机器人的行为选择概率学习到该值时, 认为已习得某一行为, 从而停止随机选择过程, 以避免小概率事件的发生. 学习率 λ 的选择与学习时间有关, 取值较小时学习时间较长. 当 t 时刻满足 $|\theta|\leqslant 0.2°$, $|\dot{\theta}|\leqslant 2°/\mathrm{s}$ 且 $p_{ik}(t)>p_\varepsilon$ 时, 机器人能够通过学习实现其自主平衡控制, 之后机器人在此状态下持续选择操作 o_k, 直到达到迭代次数 T_f. 仿真中取 $\lambda=0.05$, $p_\varepsilon=0.94$, $T_f=1\,000$.

4.2　结果与分析

经过 20 轮学习后, 系统最终趋于离散状态 s_3, 图 3 为该状态下映射的3个行为选择概率曲线. 初始时刻行为选择概率相同, 机器人状态出现较大振荡. 经过 400 次学习后, 操作行为 o_2 为 0 的概率值逐渐增加, 其他两个行为的概率值相应减小, 机器人此时倾向于选择好的行为 o_2, 从而使其能够保持在平衡位置. 小概率事件的存在 (在某一时刻机器人选择了小概率的行为) 导致系统状态出现不稳定, 如图 3 中机器人在 420 次训练时选择了小概率行为 o_3, 从而导致图 4 系统出现暂时的动荡, 但随着学习的进行又趋向于稳定状态. 当检测到机器人 t 时刻满足系统状态 $|\theta|\leqslant 0.2°$, $|\dot{\theta}|\leqslant 2°/\mathrm{s}$ 和 $p_{32}(t)>p_\varepsilon$ 时, 机器人已能通过学习实现自主平衡, 然后按平衡状态下的确定性选择操作 o_2. 该约束条件的加入避免了小概率事件易导致的破坏性实验.

为了验证本文方法的有效性, 将本文方法与传统的 LQR 方法进行对比仿真研究. 由图 4 可见, 在学习的初始阶段, OCLA 没有任何经验, 振荡较大, 学习速度也较慢, 因此在经过约 430 次学习训练后, OCLA 机器人逐渐学会了平衡. 当机器人达到平衡后, 在第 600 次训练时施加了一个正脉冲干扰信号. 由仿真结果可以看出, OCLA 学习系统能使机器人在约 70 次学习训练后达到平衡, 而 LQR 控制则需要大约 150 次训练次数才能恢复平衡状态. 这说明, 当外界环境发生变化时, OCLA 能使机器人快速地适应环境并作出响应, 具有较好的鲁棒性.

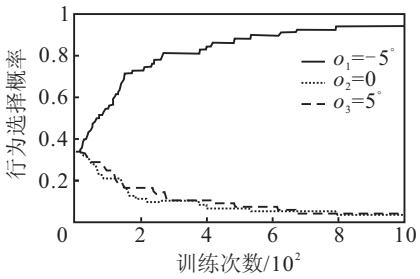
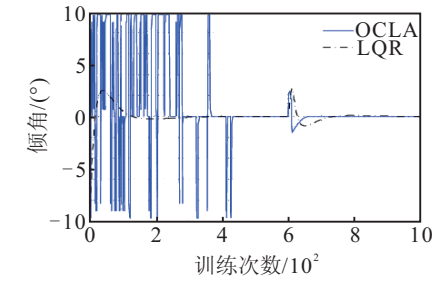
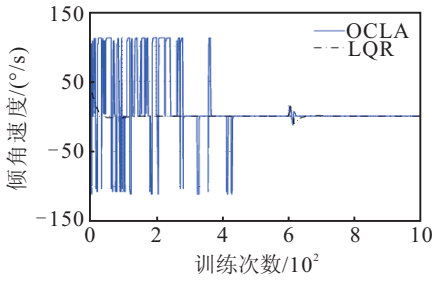


图 3　行为选择概率曲线



(a) 倾角



(b) 倾角速度

图 4 机器人倾角和倾角速度变化曲线

学习结束后各操作行为的概率值为 $p(o_1|s_3) = 0.001$, $p(o_2|s_3) = 0.98$, $p(o_3|s_3) = 0.019$. 代入操作信息熵得到

$$H = -\sum_{k=1}^{3} p_{ik} \times \log_2 p_{ik} \approx 0.069,\ i = 3.$$

图 5 为机器人操作熵的变化曲线, 用熵来反应机器人系统的状态, 开始时因其状态出现的概率相等使得此时熵值最大. 机器人通过取向信息不断调整行为选择的概率, 使机器人慢慢学习不断调整, 操作熵的值也逐渐减小并趋于 0, 最终系统趋于一种稳态, 达到控制平衡的目的.

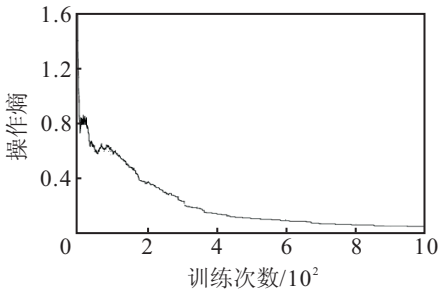


图 5 操作熵变化曲线

由仿真结果可以看出, 虽然本文所设计的操作条件反射学习自动机模型在初始阶段的控制效果稍差于经典的控制方法, 但能有效表达出生物渐近的学习特性, 当环境发生变化时具有快速响应性, 表现出较好的控制性能.

5 结 论

本文在学习自动机理论的框架中基于操作条件反射理论构造了一种仿生学习模型 OCLA, 并将其应用于两轮机器人的平衡控制. 仿真结果表明, OCLA 能使机器人自组织地渐近地学习平衡的技能, 具有仿生的特性. 该方法由于采取状态离散和行为离散的方式, 简单易行, 但泛化能力较弱, 且划分时需要更多的先验知识, 对于更为复杂的系统不易得到合理的划分结果, 下一步将重点研究如何提高其泛化能力.

参考文献(References)

[1] Skinner B F. The behavior of organisms[M]. New York: Appleton Century Crofts, 1938: 18-32.

[2] 任红格, 阮晓钢. 基于 Skinner 操作条件反射的两轮机器人自平衡控制[J]. 控制理论与应用, 2010, 27(10): 1423-1428.
(Ren H G, Ruan X G. Self-balance control of two-wheeled robot based on Skinner's operant conditioned reflex[J]. Control Theory & Applications, 2010, 27(10): 1423-1428.)

[3] 阮晓钢, 陈静. 基于滑模思想和 Elman 网络的操作条件反射学习控制方法[J]. 控制与决策, 2011, 26(9): 1398-1401.
(Ruan X G, Chen J. Operant conditioning reflex learning control scheme based on SMC and Elman network[J]. Control and Decision, 2011, 26(9): 1398-1401.)

[4] Touretzky D S, Saksida L M. Operant conditioning in skinnerbots[J]. Adaptive Behavior, 1997, 5(3/4): 219-247.

[5] Saksida L M, Raymond S M, Touretzky D S. Shaping robot behavior using principles from instrumental conditioning[J]. Robotics and Autonomous Systems, 1998, 22(3/4): 231-249.

[6] Gaudiano P, Chang C. Adaptive obstacle avoidance with a neural network for operant conditioning: Experiments with real robots[C]. IEEE Int Symposium on Computational Intelligence in Robotics and Automation. New York: IEEE Press, 1997: 13-18.

[7] Itoh K, Miwa H, Matsumoto M, et al. Behavior model of humanoid robots based on operant conditioning[C]. The 5th IEEE-RAS Int Conf on Humanoid Robots. Tsukuba: Institute of Electrical and Electronic Engineers Computer, 2005: 220-225.

[8] Pierce D, Kuipers B. Learning to explore and build maps[C]. Proc of the National Conf on Artificial Intelligence. Seattle: AAAI, 1994: 1264-1271.

(下转第939页)