文章编号: 1001-0920 (2016) 02-0303-07 DOI: 10.13195/j.kzyjc.2014.1835

基于学习自动机的具有内发动机的感知运动系统的建立

阮晓钢[1], 张晓平[1], 武 璇[1], 庞 涛[1,2]

(1. 北京工业大学 电子信息与控制工程学院，北京 100124；2. 沈阳航空航天大学 电信学院，沈阳 110136)

摘 要: 以学习自动机为数学模型, 结合斯金纳操作条件反射, 建立一种人工感知运动系统, 称为感知运动自动机(SMA). 该系统包括感知状态集合、动作集合、感知运动取向性映射集合等9部分. 系统引入好奇心和取向性概念, 设计具有主动学习环境的内发动机机制, 定义并分析了取向性学习过程, 证明了系统熵的收敛性. 通过模拟斯金纳鸽子实验表明了系统的可行性和有效性, 仿真结果表明系统具有较好的自学习和自组织特性, 同时稳定性较高.

关键词: 感知运动系统；学习自动机；操作条件反射；内发动机

中图分类号: TP242 **文献标志码:** A

Establishment of a sensorimotor system with mechanism of intrinsic motivation based on the learning automaton

RUAN Xiao-gang[1], *ZHANG Xiao-ping*[1], *WU Xuan*[1], *PANG Tao*[1,2]

(1. College of Electronic Information and Control Engineering, Beijing University of Technology, Beijing 100124, China; 2. College of Electronic Information Engineering, Shenyang Aerospace University, Shenyang 110136, China. Correspondent: ZHANG Xiao-ping, E-mail: zhangxiaoping369@163.com)

Abstract: By using the learning automaton as a mathematical model, and combing the Skinner's operant conditioning theory, a sensorimotor system is established. The system consists of nine parts: the sensory state set, the motion set, the sensorimotor orientation mapping set and etc. The concepts of curiosity and orientation are introduced to the system, and the mechanism of intrinsic motivation is designed which can learn the environment actively. The orientation learning process is defined and analyzed, and the convergence of the system's entropy is proved. Based on the famous Skinner's pigeon experiment, the system's feasibility and effectiveness are proved. The experiment results show that the system has good properties of self-learning and self-organizing as well as high stability.

Keywords: sensorimotor system; learning automaton; operant conditioning; intrinsic motivation

0 引 言

机器人是20世纪人类最伟大的发明之一, 自20世纪60年代初问世至今已取得了很大进步. 目前, 机器人学研究的两大热点是: 1) 对机器人肢体化能力的研究[1-2]; 2) 对机器人智能问题的探索. 对于智能机器人的研究已由早期的示教再现型机器人及具有简单感知能力的机器人发展到今天的认知发育型机器人[3].

1952年, 日内瓦大学心理学教授Piaget[4]指出, 人类认知发育的第1阶段主要通过其感知运动技能获得. 而感知运动技能的习得需要感知器官和运动器官协调完成, 这就涉及到感知运动系统. 早在20世纪80年代初, 美国哈佛大学的Houk等[5]针对感知运动神经信号在小脑皮质中的处理过程进行了研究. 1999年, Lee等[6]采用嵌入式系统所设计的人工感知运动系统, 能够使移动机器人通过来自视觉和听觉的信息成功跟踪特定目标. 2005年, Natale等[7]为人形机器人设计的感知运动系统, 能通过认知, 由简单的初始化形态形成感知运动神经协调机制. 2009年, Hülse等[8]从生物学角度出发, 设计了一种自主机器人手眼协调感知运动神经系统及其认知算法, 通过发育成功实现手眼协调机制. 2011年, Mathews[9]等构建了一个名为PASAR的人工感知运动系统, 可以模拟预测、期望、感觉、注意、响应等认知机制. 以上工作

收稿日期: 2014-12-02；**修回日期:** 2015-03-20.

基金项目: 国家自然科学基金项目(61375086)；国家973计划项目(2012CB720000)；北京市自然科学基金项目/北京市教育委员会科技计划重点项目(KZ201210005001).

作者简介: 阮晓钢(1958−), 男, 教授, 博士生导师, 从事机器人、自动控制与人工智能等研究；张晓平(1991−), 女, 博士生, 从事人工智能与机器人、智能控制的研究.

都对感知运动系统进行了相关的理论研究和建模, 但都是针对具体任务或目标的, 并未给出统一的、形式化的数学模型, 不具备泛化能力.

1938 年, 美国哈佛大学心理学教授 Skinner[10] 首次提出了“操作条件反射”的概念, 此后其操作条件反射理论被广泛应用于认知机器人学习过程中. 1988 年, Rosen 等[11]利用斯金纳操作条件反射原理, 成功实现了倒立摆在一定距离内的平衡控制. 2011 年, 蔡建羡等[12]以概率自动机为平台, 基于操作条件反射原理, 设计了一种自主学习系统, 用于两轮机器人自主学习过程, 并成功实现了两轮机器人的自平衡控制, 同年, 其团队又结合遗传算法, 构建了基于遗传算法的 Skinner 操作条件反射学习模型[13], 同样有效实现了两轮机器人的自平衡控制. 以上相关工作对斯金纳操作条件反射理论进行了研究, 但均未涉及感知运动系统, 同时设计的系统或模型在达到一定的稳定状态后无法杜绝小概率事件的发生. 2014 年, Shi 等[14]基于斯金纳操作条件反射理论建立了自平衡机器人感知运动系统的认知模型, 使机器人能够在移动过程中通过自学习获得运动平衡技能, 所采用的动作选择机制同样会引起小概率事件发生.

感知运动系统和操作条件反射理论对于人或动物运动技能的习得都有着指导性的意义, 将这种感知运动能力复制到机器人上, 使机器人能够主动探索外部世界, 学习世界知识, 对认知发育机器人的研究有着重要意义. 本文借鉴斯金纳操作条件反射理论, 以学习自动机为数学模型, 建立一种新的具有内发动机机制的人工感知运动系统. 通过著名的斯金纳鸽子实验对系统的自学习和自组织能力进行了验证, 实验结果表明了系统的可行性和有效性. 系统引入了好奇心和取向性概念, 设计了能够主动学习环境的内发动机机制, 可以提高系统的学习效率, 成功避免小概率事件发生的情况, 提高了系统的稳定性.

1 感知运动自动机 (SMA)

1.1 Skinner 操作条件反射

条件反射分为经典条件反射和操作条件反射两种重要理论. 经典条件反射由俄国生理学家巴甫洛夫于 1927 年提出, 适用于应答性行为, 是一种“刺激-反应”的过程. 在此基础上, 斯金纳根据自己创制的斯金纳箱对白鼠和鸽子进行实验, 提出了操作条件反射理论. 该理论适用于操作性行为, 由经典条件反射的“刺激-反应”模式深入为“反应-刺激-反应”模式. 斯金纳认为, 人或动物执行某一行为后, 如果行为的后果能促使这一行为的发生, 则称之为正强化, 该行为之后发生的概率将有所增加; 如果行为的后果不容易使这一行为发生, 则称为负强化, 该行为之后发生的概率将减小. 换言之, 正强化促进某一行为的发生, 负强化避免某一行为的发生.

Skinner 操作条件反射理论的核心是: 动物在与环境的接触过程中, 不断学习, “感知-运动-感知”不断循环, 最终形成感知序列与动作序列之间的映射关系.

1.2 SMA 模型定义

从神经生理学角度看, 感知运动系统主要包括: 感觉神经系统、运动神经系统、小脑皮质中的感觉运动认知学习单元 3 部分. 基于神经生理学和神经心理学知识, 所构建的感知运动系统定义为一个九元组模型 SMA $= \langle S, M, O, C, V, V_s, P, F, E\rangle$, 各元素含义具体如下.

1) S: SMA 内部离散感知状态集合, $S = \{s_i | i = 1, 2, \cdots, n_s\}$, $s_i \in S$ 为第 i 个感知状态, n_s 为可感知到的离散状态的个数.

2) M: SMA 动作集合, $M = \{M_i | i = 1, 2, \cdots, n_s\}$, $M_i = \{m_{ij} | j = 1, 2, \cdots, n_i\}$, m_{ij} 表示 SMA 在第 i 个感知状态下第 j 个可选动作, n_i 为第 i 个状态下可选动作的个数.

3) O: SMA 的“感知-运动”取向性映射集合, $O = \{O_i | i = 1, 2, \cdots, n_s\}$, $O_i = \mathrm{diag}([o_{i1}, \cdots, o_{ij}, \cdots, o_{in_i}])_{n_i \times n_i}$ 为第 i 个状态对该状态下可选动作的取向性映射矩阵, diag 表示括号里的元素以对角阵的方式储存, $o_{ij}\,(i \in (1, 2, \cdots, n_s), j \in (1, 2, \cdots, n_i))$ 表示一条“感知-运动”映射, 表征的是 SMA 在感知状态 $s_i \in S$ 下对动作 m_j 的取向性, 或称感知状态 s_i 与动作 m_j 的感知运动取向性为 o_{ij}, 满足 $0 \leqslant o_{ij} \leqslant 1$ 且 $\sum_{j=1}^{n_i} o_{ij} = 1$.

4) C: 好奇心, $C = \{c_i | i = 1, 2, \cdots, n_s\}$, c_i 为系统第 i 个状态下的好奇心, 从生物学角度出发, 动物在某一状态下的好奇心随探索该状态次数的增加而下降, 基于此, 好奇心设计如下:

$$c_i = \frac{1}{1 + \mathrm{e}^{k(N_i - c)}}. \tag{1}$$

其中: N_i 为至 t 时刻系统探索状态 s_i 的次数; k、c 为好奇心参数.

取向性和好奇心是影响生物选择动作的两个内在因素, 基于此, 本系统的内发动机机制设计为选择所处状态下取向性和好奇心之和最大的动作.

5) V: 系统状态取向值, $V = \{V_i | i = 1, 2, \cdots, n_s\}$, 用来决定取向函数的值, 与系统感知状态一一对应. 其中 $V_i \in [-1, 1]$, -1 为最差状态的状态取向值, 1 为最理想状态的状态取向值.

6) V_s: 取向函数, $V_s = aV_n + b(V_n - V_o)$, V_o 和 V_n 分别表示执行某一动作之前和之后的状态, 其中

$a \geqslant 0, b \geqslant 0$ 为取向函数参数, 其取值应保证取向函数的正负号不改变 $(V_n - V_o)$ 的正负号, 且满足 $a+b=1$, 一般可以通过学习得到.

7) P: 取向性学习矩阵, $P = \{P_i | i = 1, 2, \cdots, n_s\}$, 作用是依据取向函数所提供的信息, 对取向性映射进行更新调整, 其中 $P_i = \mathrm{diag}([p_{i1}, \cdots, p_{ij}, \cdots, p_{in_i}])_{n_i \times n_i}$, 各参数意义与 3) 中相同, 不再赘述. 设 t 时刻系统在感知状态 s_i 下的取向性映射为 $O_i(t)$, 执行动作 m_j 后, 在该感知状态下的取向性映射变为 $O_i(t+1)$, 则取向性映射更新方法如下:

$$\begin{cases} p_{ij}(t) = 1 + \mathrm{Sign}(V_s(t))(1 - \mathrm{e}^{-\eta |V_s(t)|}); \\ p_{ik}(t) = 1,\ k \in (1, 2, \cdots, n_i),\ k \neq j; \end{cases} \tag{2}$$

$$\mathrm{Sign}(x) = \begin{cases} 1,\ x > 0; \\ 0,\ x = 0; \\ -1,\ x < 0; \end{cases} \tag{3}$$

$$O_i(t+1) = \frac{1}{\sum\limits_{j=1}^{n_i} o_{ij}(t) p_{ij}(t)} O_i(t) P_i(t). \tag{4}$$

其中 $\eta > 0$ 为取向性学习参数.

8) F: SMA 内部状态转移函数, $F(s(t), m(t)) = s(t+1)$, 表示 t 时刻在感知状态为 $s(t)$ 下执行动作 $m(t)$ 后状态转移为 $s(t+1)$.

9) E: 感知运动系统的知识熵, $E = \{E_i | i = 1, 2, \cdots, n_s\}$, 用来描述系统对知识的学习程度, 表征系统的自学习和自组织特性. 系统 t 时刻的知识熵定义为

$$E(t) = \sum_{i=1}^{n_s} E_i(t), \tag{5}$$

$$\begin{aligned} & E_i(t) = E_i(m_j(t)|s_i) = \\ & -\sum_{j=1}^{n_i} o_{ij}(t) \log_2 o_{ij}(t) = \\ & -\sum_{j=1}^{n_i} o_{ij}(m_j(t)|s_i) \log_2 o_{ij}(m_j(t)|s_i). \end{aligned} \tag{6}$$

1.3　算法流程

SMA 的基本工作原理可以简述如下: t 时刻, 系统感知到其内部状态 $s_i \in S$, 计算该时刻当前状态对各动作的取向性映射矩阵 $O_i(t)$ 及该状态下的好奇心 $c_i(t)$; 将好奇心随机加在任意动作之上, 依照系统内发动机机制选取取向性和好奇心之和最大的动作; 执行该动作, 状态发生转移; 计算新状态的状态取向值及取向函数的值; 根据取向函数提供的信息按式 (2) 计算取向性学习矩阵, 按式 (4) 更新取向性映射矩阵, 获得新的 "感知-运动" 映射; 如此循环, 直至取向性不再发生变化或学习时间 t 大于终止时间 T_0 时, 学习结束. 算法流程如图 1 所示.

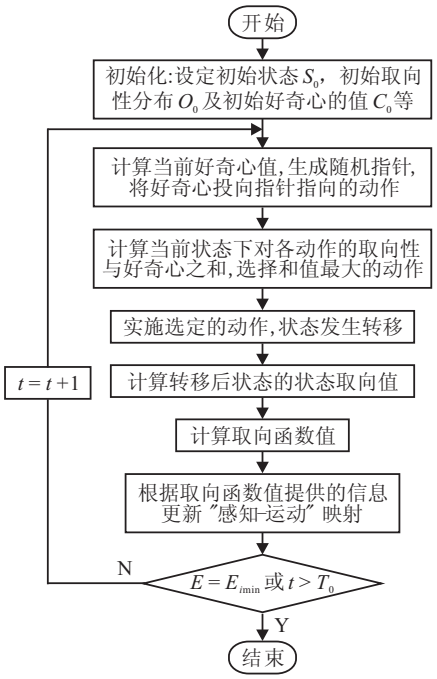


图 1　SMA 算法流程

2　学习算法分析证明

定理 1　设 SMA 是一个人工感知运动系统, 则下式成立:

$$\lim_{N_i \to \infty} o_{ik}(m_{ik}(N_i) | s_i(N_i)) \to 1, \tag{7}$$

$$\lim_{N_i \to \infty} o_{ik'}(m_{ik'}(N_i) | s_i(N_i)) \to 0. \tag{8}$$

其中: m_{ik} 为感知状态 s_i 下使取向函数 V_s 值为正的某个动作; $m_{ik'}$ 为其余动作, 包括两部分: 一部分为使取向函数值为负的所有动作, 一部分为除动作 m_{ik} 后剩余的使取向函数 V_s 值为正的动作.

定理 1 表示, 当状态 s_i 被学习的次数趋于无穷时, SMA 随学习进程, 依取向性 1 选择某个取向性好的动作 m_{ik}, 以取向性 0 选择其余动作.

证明　SMA 系统依据设计的内发动机机制选择动作, 每个时刻选择取向性与好奇心之和最大的动作, 其中好奇心是一个随 N_i 下降的函数, 当 $N_i > N_T$ 后, 好奇心不再影响动作选择, 动作的选择将主要依赖于系统的取向性. 设 m_{ib} 为状态 s_i 下所有使取向函数 V_s 值为正的动作, m_{iw} 为状态 s_i 下所有使取向函数 V_s 值为负的动作.

1) 当 $0 \leqslant N_i \leqslant N_T$ 时, SMA 系统对各动作的 "好奇心" 较大, 因此, 对动作的选择具有随机性:

若 $m_{ij}(N_i) \in m_{ib}(N_i)$, 则根据取向性映射更新机制, 有

$$V_s(N_i) > 0, \tag{9}$$

$$\begin{cases} p_{ij}(N_i) = 1 + (1 - \mathrm{e}^{-\eta |V_s(N_i)|}) = 1 + \Delta > 1, \\ p_{ij'}(N_i) = 1, \end{cases} \tag{10}$$

其中 $0<\Delta=1-\mathrm{e}^{-\eta|V_s(N_i)|}<1$. 因此, 有

$$\begin{cases} o_{ij}(N_i+1)=\dfrac{o_{ij}(N_i)p_{ij}(N_i)}{\sum\limits_{l=1}^{n_i}o_{il}(N_i)p_{il}(N_i)}=\\ \qquad\qquad \dfrac{o_{ij}(N_i)+\Delta\times o_{ij}(N_i)}{1+\Delta\times o_{ij}(N_i)},\\ o_{ij}(N_i)<o_{ij}(N_i+1)<1; \end{cases} \tag{11}$$

$$\begin{cases} o_{ij'}(N_i+1)=\dfrac{o_{ij'}(N_i)p_{ij'}(N_i)}{\sum\limits_{l=1}^{n_i}o_{il}(N_i)p_{il}(N_i)}=\\ \qquad\qquad \dfrac{o_{ij'}(N_i)}{1+\Delta\times o_{ij}(N_i)},\\ 0<o_{ij'}(N_i+1)<o_{ij'}(N_i). \end{cases} \tag{12}$$

若 $m_{ij}(N_i)\in m_{iw}(N_i)$, 则根据取向性映射更新机制, 有

$$V_s(N_i)<0, \tag{13}$$

$$\begin{cases} 0<p_{ij}(N_i)=1-(1-\mathrm{e}^{-\eta|V_s(N_i)|})=1-\Delta'<1,\\ p_{ij'}(N_i)=1, \end{cases} \tag{14}$$

其中 $0<\Delta'=1-\mathrm{e}^{-\eta|V_s(N_i)|}<1$. 因此, 有

$$\begin{cases} o_{ij}(N_i+1)=\dfrac{o_{ij}(N_i)p_{ij}(N_i)}{\sum\limits_{l=1}^{n_i}o_{il}(N_i)p_{il}(N_i)}=\\ \qquad\qquad \dfrac{o_{ij}(N_i)-\Delta'\times o_{ij}(N_i)}{1-\Delta'\times o_{ij}(N_i)},\\ 0<o_{ij}(N_i+1)<o_{ij}(N_i); \end{cases} \tag{15}$$

$$\begin{cases} o_{ij'}(N_i+1)=\dfrac{o_{ij'}(N_i)p_{ij'}(N_i)}{\sum\limits_{l=1}^{n_i}o_{il}(N_i)p_{il}(N_i)}=\\ \qquad\qquad \dfrac{o_{ij'}(N_i)}{1-\Delta'\times o_{ij}(N_i)},\\ o_{ij'}(N_i)<o_{ij'}(N_i+1)<1. \end{cases} \tag{16}$$

该阶段学习的结果使得系统对好的行为的取向性高于初始取向性, 差的行为的取向性低于初始取向性.

2) 当 $N_i>N_T$ 时, 好奇心将不再影响动作的选择, 系统此后总以最大的取向性选择 m_{ib} 中的某一个动作 m_{ik}. 即当 $N_i>N_T$ 时, 系统对动作的选择具有了确定性, 小概率事件不再发生, 该动作满足式 (9)～(11), 其取向性不断增大. 从数学角度分析, $o_{ik}(N_i)$ 为一个单调递增函数, 且有上界 1, 可知 $o_{ik}(N_i)$ 将增长直至 $o_{ik}(N_i)=1$ 为止, 从而得证

$$\lim_{N_i\to\infty}o_{ik}(m_{ik}(N_i)|s_i(N_i))\to 1.$$

而对于动作 $m_{ik'}(N_i)$, 满足式 (12), $m_{ik'}(N_i)$ 为单调递减函数且有下界 0, 可知 $m_{ik'}(N_i)$ 将减小直至 $m_{ik'}(N_i)=0$ 为止, 从而得证

$$\lim_{N_i\to\infty}o_{ik'}(m_{ik'}(N_i)|s_i(N_i))\to 0. \qquad \square$$

定理 2 SMA 是一个自学习自组织的系统, 是消除不确定性的过程, 系统的知识熵将随学习进程趋于极小, 具体阐述如下: 设 $\mathrm{SMA}=\langle S,M,O,C,V,V_s,P,F,E\rangle$ 是一个感知运动自动机, 其知识熵 $E_i(m_j(N_i)|s_i(N_i))$ 随学习进程收敛至极小, 即 $\lim\limits_{N_i\to\infty}E_i(N_i)\to E_{i\min}$.

证明 初始时刻, 系统没有任何先验知识, 因此, 在状态 s_i 下对所有动作取向性相同, 知识熵最大. 随着学习的进行, 取向性矩阵被更新, 由定理 1 可知当系统状态被充分学习即 $N_i\to\infty$ 时, 每个状态 s_i 下, 系统均以取向性 1 选择某个动作 m_{ik}, 以取向性 0 选择剩下的动作 $m_{ik'}$, 因此, 有

$$\begin{aligned} &\lim_{N_i\to\infty}E_i(N_i)=\\ &\lim_{N_i\to\infty}E_i(m_j(N_i)|s_i(N_i))=\\ &\lim_{N_i\to\infty}\Big(-\sum_{j=1}^{n_i}o_{ij}(N_i)\log_2 o_{ij}(N_i)\Big)=\\ &-\lim_{N_i\to\infty}\Big(o_{ik}(N_i)\log_2 o_{ik}(N_i)+\\ &\sum_{j=1,j\neq k}^{n_i}o_{ij}(N_i)\log_2 o_{ij}(N_i)\Big)\to\\ &E_{i\min}=0. \end{aligned} \tag{17}$$

由此定理 2 得证. $\square$

3 仿真实验

3.1 斯金纳鸽子实验简介

斯金纳鸽子实验是关于操作条件反射理论的著名实验, 很多学者用其来证明所设计学习系统的自学习和自组织特性[15].

斯金纳鸽子实验是在斯金纳设计的一种动物实验仪器即著名的斯金纳箱中进行的, 箱内放进一只鸽子, 并设定 3 个按键: 红、黄、蓝, 箱子的构造尽可能排除一切外部刺激, 鸽子在箱内可以自由活动, 其原理如图 2 所示[15]. 当鸽子啄红色按键时可以获得食物, 啄黄色按键时无任何结果, 啄蓝色按键时给予电击.

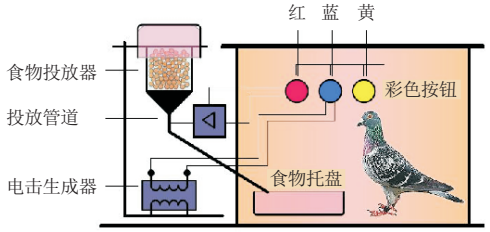


图 2 斯金纳箱

使用该系统来验证所设计的感知运动自动机的自学习和自组织特性, 认为鸽子在任意时刻能够感

受到的是其自身内部状态, 如饱、半饱、饥饿. 开始时, 鸽子没有任何知识, 其按键具有随机性, 人们希望鸽子在与环境的接触过程中能够习得环境知识.

3.2 实验设置

针对本实验, 感知运动自动机各元素设置如下.

1) SMA 内部离散感知状态: s_1 饥饿, s_2 半饱, s_3 饱, $n_s = 3$.

2) SMA 的动作集合. 在本实验中, 鸽子在任何感知状态下都只有 3 个可选动作, 因此, 不再区分不同状态下的动作集, 统一表示为 m_1: 按红色按键, m_2: 按黄色按键, m_3: 按蓝色按键, $n_1 = n_2 = n_3 = 3$.

3) SMA初始"感知-运动"取向性为 $o_{ij}(0) = 1/3$. 其中: $i = 1, 2, 3$; $j = 1, 2, 3$.

4) 好奇心. 所有状态下使用统一的好奇心函数. 通过实验发现: 好奇心参数 k 值越大, 系统学习越快, 但对状态的学习越不完全; 值越小, 对状态学习越完全, 但学习速度下降. 针对本实验, 好奇心参数最终设置为 $k = 0.01$, $c = 1$.

5) SMA 状态取向值, 本实验中状态取向值设置为: $V_1 = -1$, $V_2 = 0$, $V_3 = 1$.

6) 取向函数参数取 $a = 0.1$, $b = 0.9$.

7) 初始取向性学习矩阵元素值均为 1, 即 $p_{ij} = 1$. 其中: $i = 1, 2, 3$; $j = 1, 2, 3$; 取向性学习参数 $\eta = 0.1$.

8) SMA 的内部状态转移函数, 认为鸽子如果在 t 时刻没有获得食物, 则 $t+1$ 时刻就会变到下一个差的状态; 如果在 t 时刻获得食物, 则 $t+1$ 时刻就变到下一个好的状态. 具体表示为

$$F(s_1, m_1) = s_2,\ F(s_2, m_1) = s_3,\ F(s_3, m_1) = s_3,$$
$$F(s_1, m_2) = s_1,\ F(s_2, m_2) = s_1,\ F(s_3, m_2) = s_2,$$
$$F(s_1, m_3) = s_1,\ F(s_2, m_3) = s_1,\ F(s_3, m_3) = s_1.$$

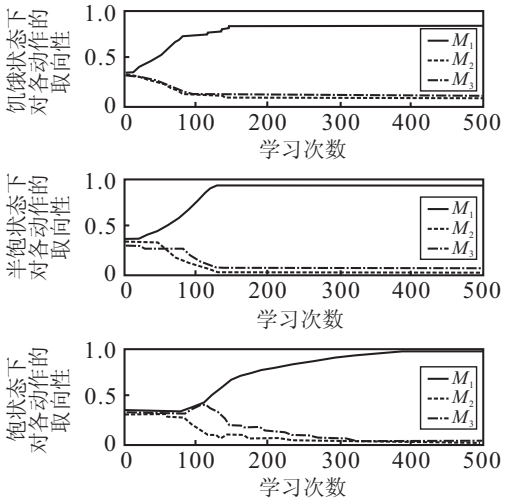
9) 初始时刻熵值最大, 为

$$E_i(0) = E_i(m_j(0)|s_i) = -\sum_{j=1}^{n_i} o_{ij}(0)\log_2 o_{ij}(0) = -\sum_{j=1}^{3} \frac{1}{3}\log_2\frac{1}{3} \approx 1.585, \tag{18}$$

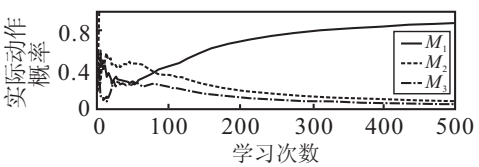
其中 $i = 1, 2, 3$.

3.3 实验结果与分析

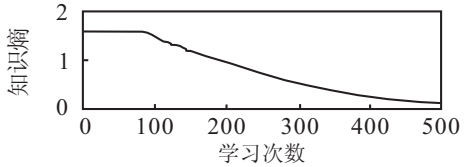
图 3 给出了 SMA 的基本学习过程. 图 3(a) 为各状态下的取向性变化曲线, 可以看出, 刚开始, 鸽子对外部世界没有任何认识, 在每个状态下, 对 3 个动作的取向性相等, 均为 1/3. 随着学习的进行, 鸽子对不同动作的取向性发生了变化, 在所有状态下, 鸽子对于按红色按键的取向性不断增大, 对蓝色按键和黄色按键的取向性不同程度地减小, 表明鸽子在与环境的接触过程中学会了一定的知识, 明白了按红色按键能够得到食物. 实验数据表明, 在学习 146 步和 147 步后, 鸽子在状态 s_1 和 s_2 下对按红色按键的取向性分别保持 0.824 9 和 0.924 0 不再发生变化. 这是因为从 148 步开始, 鸽子在状态 s_3 下对红色按键、黄色按键和蓝色按键的取向性分别为 0.678 0, 0.102 6, 0.219 4, 此时该状态下的好奇心值下降到 0.186 9, 不再影响动作选择. 此后, 鸽子均以最大取向性选择红色按键, 从而停留在理想状态 s_3, 不再返回状态 s_1 和 s_2. 但在状态 s_1 和 s_2 下, 鸽子同样学会了以较大的取向性选择红色按键这一知识, 此时鸽子已经学会通过按红色按键获得食物使自己保持在饱的状态. 图 3(b) 为学习过程中的实际动作概率曲线, 可见, 在学习初始阶段, 鸽子选择按键具有随机性, 随着学习的进行, 选择红色按键的次数明显高于其他两个按键. 图 3(c) 以状态 s_3 为例, 给出了该状态下的知识熵曲线, 其值由开始的最大值 1.585 0 下降到学习后的 0.121 4, 同样体现了系统的自学习和自组织特性.



(a) 取向性变化曲线



(b) 实际动作概率曲线

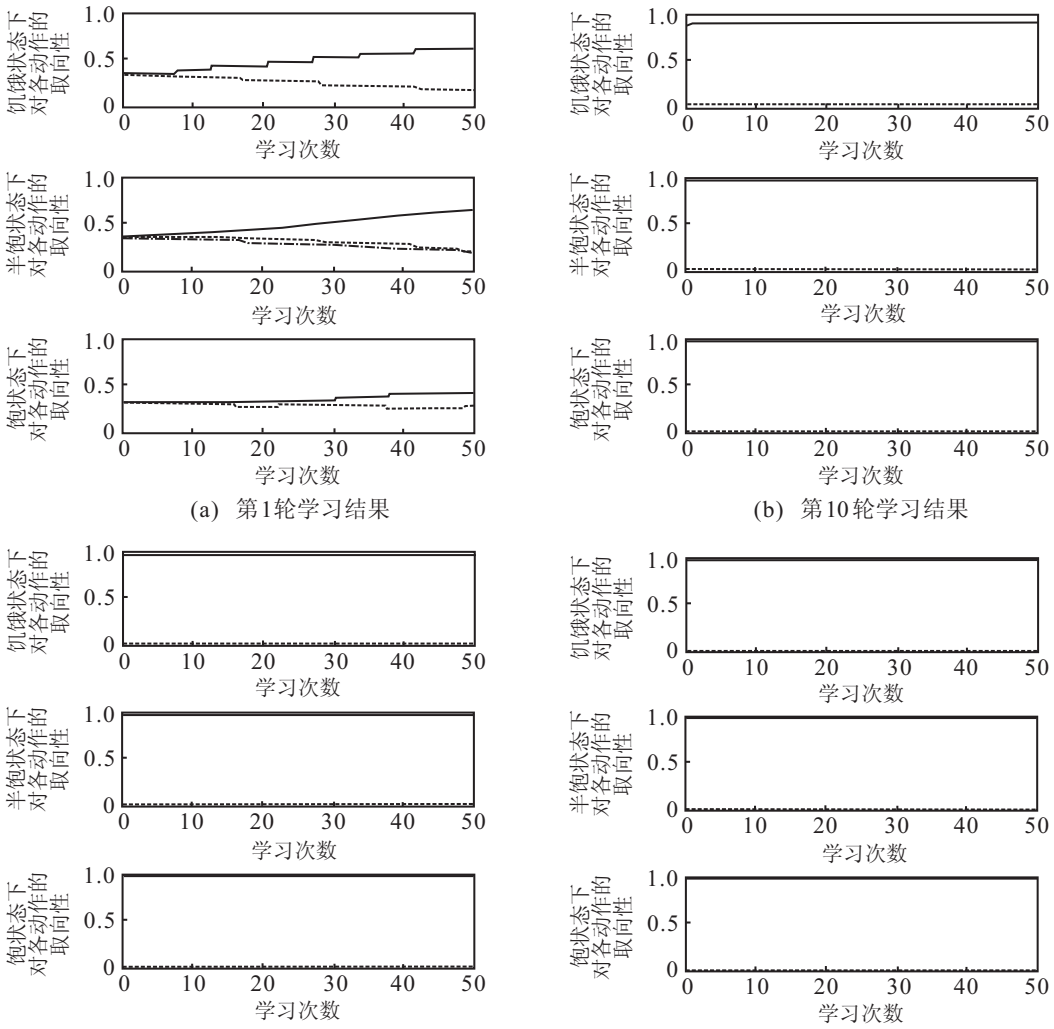


(c) 状态s_3下的知识熵曲线

图 3　SMA 基本学习过程

如前所述, 鸽子在理想状态 s_3 下学会不断按红色按键以保持饱的状态, 不再对状态 s_1 和 s_2 进行学习, 尽管在状态 s_1 和 s_2 下, 鸽子已经学会以较大的取

向性选择红色按键. 为了说明系统的充分收敛性, 在此设定轮次学习, 每轮学习 50 次, 之后在前一轮的学习基础上重新学习, 结果如图 4 所示, 分别为第 1 轮、第 10 轮、第 20 轮和第 30 轮的学习结果.



(a) 第1轮学习结果　(b) 第10轮学习结果

(c) 第20轮学习结果　(d) 第30轮学习结果

图 4　不同轮次下的学习结果

以状态 s_1 为例, 系统在第 1 轮、第 10 轮、第 20 轮和第 30 轮学习后的取向性映射如表 1 所示. 由表 1 可以看出, 经过不同轮次的学习, 系统的取向性在不断向好的方向发展, 说明只要让系统充分地学习, 系统一定能够学习到最理想的取向性映射.

表 1　不同轮次学习后状态 s_1 下的取向性映射

学习轮次	o_{11}	o_{12}	o_{13}
1	0.607 0	0.191 6	0.201 4
10	0.877 1	0.058 4	0.064 5
20	0.942 3	0.027 5	0.030 2
30	0.973 8	0.012 4	0.013 8

将本文主动学习环境的内发动机机制 (以下简称主动学习机制) 与依概率学习机制 (以文献 [12-13] 中的 OCPA 为例) 进行比较, 以状态 s_2 下的学习结果为例, 结果如图 5 所示. 图 5(a) 中粗线为主动学习机制下鸽子对各动作的取向性曲线, 细线为依概率学习机制下选择各动作的概率曲线, 实线、虚线和点划线分别对应选择红色按键、黄色按键和蓝色按键. 图 5(b) 中实线为主动学习机制下的知识熵曲线, 虚线为依概率学习下的操作行为熵曲线. 从图 5(a) 和图 5(b) 均可以看出, 与依概率学习相比, 主动学习机制学习速度更快, 具有更高的学习效率. 图 5(c) 和图 5(d) 对两种学习机制达到稳定状态所需学习步数和达到稳定状态后的小概率事件进行了统计, 实线对应主动学习机制, 虚线对应依概率学习机制. 在此认为若鸽子能连续 5 步选择同一动作, 则系统进入了稳定状态. 结果表明: 在主动学习机制下, 所设计的感知运动系统平均经过 21.55 步达到稳定状态, 最高步数为 25 步, 最低步数为 19 步, 学习过程比较稳定, 达到稳定状态后不再发生小概率事件; 而依概率学习机制设计的仿生自主系统平均经过 35.65 步进入稳定状态, 最低步数为 22 步, 最高步数为 54 步, 学习过程不稳定, 且后期发生小概率事件平均 2.25 次, 最高时发生 9 次. 从工程角度出发, 小概率事件往往具有破坏性的

结果, 系统进入理想状态后, 应尽量避免小概率事件的发生, 相比之下, 主动学习机制的设计更稳定, 学习结果更可靠.

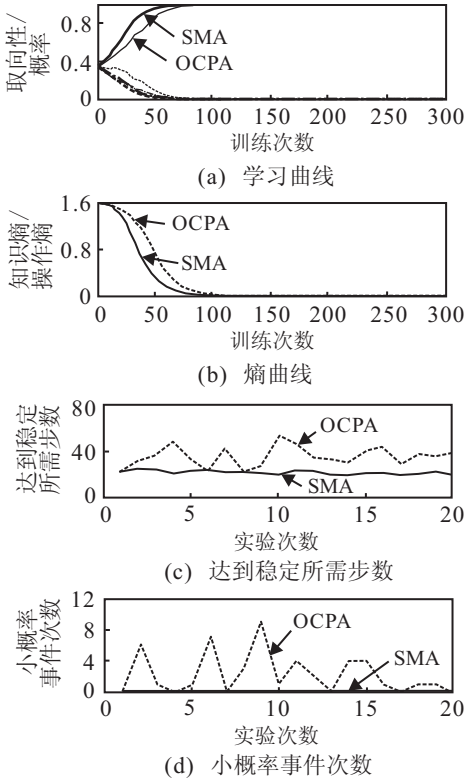


图 5　主动学习与依概率学习对比实验

4　结　论

感知运动技能的获得是智能体认知的基础, 为机器人建立一种人工感知运动系统, 对于建立具有认知发育能力的机器人具有重要意义. 本文从神经系统科学角度出发, 以学习自动机为数学模型, 借鉴斯金纳操作条件反射理论, 建立了一种具有内发动机机制的感知运动系统, 通过著名的斯金纳鸽实验表明了所设计系统具有较强的学习能力和较好的稳定性.

该模型是对利用学习自动机建立人工感知运动系统的初步探索, 取向性学习矩阵的设置为后期继续设计具有预测和记忆能力的感知运动系统提供了空间. 下一步的工作是将其运用到实体机器人上, 进一步验证其主动学习和自组织等特性, 同时基于该系统设计具有认知能力的智能机器人.

参考文献(References)

[1] Kuniyoshi Y, Sangawa S. Early motor development from partially ordered neural-body dynamics: Experiments with a cortico-spinal-musculo-skeletal model[J]. Biological Cybernetics, 2006, 95(6): 589-605.

[2] Fukuoka Y, Akama J. Dynamic bipedal walking of a dinosaur-like robot with an extant vertebrate's nervous system[J]. Robotica, 2014, 32(6): 851-865.

[3] Weng J, McClelland J, Pentland A, et al. Autonomous mental development by robots and animals[J]. Science, 2001, 291(5504): 599-600.

[4] Piaget J. The origins of intelligence in children[M]. New York: International Universities Press, 1952: 1-20.

[5] Houk J C, Gibson A R. Sensorimotor processing through the cerebellum[J]. New Concepts in Cerebellar Neurobiology, 1987: 387-416.

[6] Lee D D, Seung H S. Learning in intelligent embedded systems[C]. Proc of USENIX Workshop on Embedded Systems. Cambridge, 1999: 133-139.

[7] Natale L, Orabona F, Berton F, et al. From sensorimotor development to object perception[C]. The 5th IEEE-RAS Int Conf on Humanoid Robots. New York: IEEE, 2005: 226-231.

[8] Hülse M, McBride S, Lee M. Robotic hand-eye coordination without global reference: A biologically inspired learning scheme[C]. The 2009 IEEE Int Conf on Development and Learning. Shanghai: IEEE, 2009: 1-6.

[9] Mathews Z, i Badia S B, Verschure P F M J. PASAR: An integrated model of prediction, anticipation, sensation, attention and response for artificial sensorimotor systems[J]. Information Science, 2011, 186(1): 1-19.

[10] Skinner B F. The behavior of organisms: An experimental analysis[M]. New York: Appleton-Century Company, 1938: 61-74.

[11] Rosen B E, Goodwin J M, Vidal J J. Machine operant conditioning[C]. Annual Int Conf of the IEEE Engineering in Medicine and Biology Society. Piscataway: IEEE, 1998: 1500-1501.

[12] 蔡建羡, 阮晓钢. OCPA 仿生自主学习系统及在机器人姿态平衡控制中的应用[J]. 模式识别与人工智能, 2011, 24(1): 138-146.
(Cai J X, Ruan X G. OCPA bionic autonomous learning system and its application on robot poster balance control[J]. Pattern Recognition and Artificial Intelligence, 2011, 24(1): 138-146.)

[13] 蔡建羡, 阮晓钢. 基于遗传算法的 Skinner 操作条件反射学习模型[J]. 系统工程与电子技术, 2011, 33(6): 1370-1376.
(Cai J X, Ruan X G. Skinner operant conditioning learning model based on genetic algorithm[J]. Systems Engineering and Electronics, 2011, 33(6): 1370-1376.)

[14] Shi Tao, Yang Weidong, Ren Hongge. A study on the cognitive model of robot sensorimotor system[J]. J of Intelligence & Fuzzy Systems, 2015, 28(5): 1955-1968.

[15] Ruan X G, Wu X. The skinner automaton: A psychological model formalizing the theory of operant conditioning[J]. Science China Technological Sciences, 2013, 56(11): 2745-2761.

(责任编辑：李君玲)