

一种具有发育机制的感知行动认知模型

张晓平¹ 阮晓钢² 王力¹ 李志军¹ 闫佳庆¹ 毕松¹

摘要 针对结构固定认知模型中存在的学习浪费与计算浪费问题,在具有内发动机机制的感知行动认知模型基础上,根据操作条件反射学习特性,借鉴潜在动作原理,建立起一种具有发育机制的感知行动认知模型 D-SSCM (Development-sensorimotor cognitive model),具体为一个 14 元组,包含离散学习时间集、内部可感知离散状态集、可输出动作集、有效输出动作空间集、潜在动作关系集、可输出动作空间探索率集及发育算法等.针对模型发育过程,分别设计了模型结构扩展式发育方法和算法以及缩减式发育方法和算法,定义了模型的发育式学习过程.使用两轮机器人自平衡任务对设计的学习模型进行验证,实验结果表明,发育机制下的感知行动认知模型 D-SSCM 具有更快的学习速度及更稳定的学习效果.

关键词 发育,操作条件反射,潜在动作,感知行动认知,两轮机器人

引用格式 张晓平,阮晓钢,王力,李志军,闫佳庆,毕松.一种具有发育机制的感知行动认知模型.自动化学报,2021,47(2):391-403

DOI 10.16383/j.aas.c180750

A Kind of Sensorimotor Cognitive Model With Developmental Mechanism

ZHANG Xiao-Ping¹ RUAN Xiao-Gang² WANG Li¹ LI Zhi-Jun¹ YAN Jia-Qing¹ BI Song¹

Abstract Aiming at the problems of learning waste and computing waste that exist in the cognitive models with fixed structure, and according to the operate conditioning learning characteristics as well as drawing on the affordance theory, a new kind of sensorimotor cognitive model named D-SSCM with the developmental mechanism is established based on the sensorimotor cognitive model with the mechanism of intrinsic motivation. D-SSCM is a fourteen tuple in specific, including discrete learning time set, internal sensible discrete state set, optional motion set, effective output motion space set, affordance relationship set, optional motion space exploration rate set, developmental algorithm and etc. In view of D-SSCM's developmental learning, extended developmental method and algorithm as well as reduced developmental method and algorithm are designed. Model's developmental learning process is defined. Using two-wheeled robot self-balancing task to test this designed model, results show that D-SSCM is with faster learning speed and more stable learning effect.

Key words Development, operant conditioning, affordance, sensorimotor cognitive, two-wheeled robot

Citation Zhang Xiao-Ping, Ruan Xiao-Gang, Wang Li, Li Zhi-Jun, Yan Jia-Qing, Bi Song. A kind of sensorimotor cognitive model with developmental mechanism. *Acta Automatica Sinica*, 2021, 47(2): 391-403

近年来,人工智能与机器人在各国政府、研究机构及相关企事业单位中的受重视程度持续提升,其中主要研究目标之一是建立类似人或动物可以进行

自主学习的认知型机器人.机器人具有认知能力表现为其能够在与周身环境的交互过程中渐进掌握知识和技能^[1].受人和动物感知运动系统的启发^[2],为机器人感知运动过程设计认知模型成为实现认知机器人的有效方法之一^[3-5].

机器人感知行动认知模型的设计主要包含模型结构及模型算法两方面.模型结构常借鉴神经生物学相关知识^[6-8],而算法方面则以心理学学习机制为指导,常见的有操作条件反射^[9-11]、内发动机^[12]等,其中,内发动机又涉及好奇心^[13-14]、情感^[7,15]等.在操作条件反射学习理论下,Cyr 等基于人工脉冲神经网络 ASNN (Artificial spiking neural networks) 为机器人设计了一种大脑控制器,实现了机器人自主移动^[10];Itoh 等为类人机器人设计了一种行为模型,成功实现了机器人握手行为学习^[11].在内发动机理论下,Ren 等借鉴多巴胺、丘脑、基底神经节-大脑皮层工作机制,针

收稿日期 2018-11-11 录用日期 2019-06-02
Manuscript received November 11, 2018; accepted June 2, 2019
北京市自然科学基金 (4204096, 4184086), 国家自然科学基金 (61903006, 61174116), 北京市教委项目 (KM201610009001), 北京高校高精尖学科建设项目, 北方工业大学青年优秀人才项目, 北方工业大学科研启动基金项目资助
Supported by Beijing Natural Science Foundation (4204096, 4184086), National Natural Science Foundation of China (61903006, 61174116), Beijing Education Commission Project (KM201610009001), Beijing Universities Top Discipline Construction Project, Youth Yuyou Talent Project of North China University of Technology, Research Initial Foundation of North China University of Technology
本文责任编辑 张俊
Recommended by Associate Editor ZHANG Jun
1. 北方工业大学电气与控制工程学院 北京 100144 2. 北京工业大学信息学部 北京 100124
1. School of Electrical and Control Engineering, North China University of Technology, Beijing 100144 2. Faculty of Information Technology, Beijing University of Technology, Beijing 100124

对两轮机器人自平衡学习问题提出一种计算模型 C-DCCM (Curiosity-driven cognitive computing model), 成功实现了机器人的自学习和对环境的自发探索^[6]; Mannella 等为平面机器人提出一种计算模型, 实现了机器人自身运动技能学习^[12]; Baranes 和 Oudeyer 将感知运动理论与内发动机理论结合, 提出一种自适应目标生成-鲁棒智能自适应好奇心算法 SAGG-RIAC (Self-adaptive goal generation-robust intelligent adaptive curiosity), 实现了冗余机器人对自身逆运动学的学习^[13-14]; Castellanos 等则考虑机器人感知行动认知过程中的情感因素, 为机器人提出一种情感评估模型, 实现了机器人的个性化^[15]. 以上工作, 机器人在学习过程中, 其认知模型算法会在一定程度上进行更新, 但模型结构均为固定, 一旦设计完成, 不再发生变化.

2001 年, Weng 在 *Science* 上发文, 首次提出自主心智发育的概念^[16], 后期继续提出一系列发育网络 DN (Development networks) 理论^[17], 奠定了发育机器人研究基础, 其发育网络核心理念就在于学习过程中网络结构可变. Cai 等以学习自动机为数学模型, 结合操作条件反射机制与模糊理论设计的模糊斯金纳操作条件反射自动机 FSOCA (Fuzzy skinner operant conditioning automaton) 就融合了这样的思想, 基于在线聚类算法实现了感知行动映射规则的增加和删除^[9]. 发育理念对于机器人学习非常重要, 以文献 [18] 为例, 其在感知运动系统认知模型中同时结合操作条件反射与内发动机机制, 为两轮机器人设计了一种具有内发动机机制的认知模型, 使得机器人表现出一定的认知能力, 然而, 其模型结构固定, 研究过程中表现出两个问题: 1) 模型需要学习的动作空间需要提前定义, 降低了机器人的智能性, 并且固定的动作学习空间存在大量对无效感知行动映射的探索和学习, 造成学习的浪费, 导致模型学习率低; 2) 固定感知行动映射空间下, 模型需要对当前学习状态下所有的感知行动映射取向性进行更新, 存在计算上的浪费.

受发育理论启发, 本文在文献 [18] 的研究基础上, 借鉴潜在动作理论, 设计了一种新的结构可发育的机器人感知行动认知模型 D-SSCM, 针对模型的发育式学习过程, 分别设计了模型扩展式学习方法和算法以及缩减式学习方法和算法, 同时节省了机器人学习成本和计算成本, 很大程度上提高了机器人的学习速度和学习稳定性. 将文本模型与文献 [18] 在相同实验任务及参数设置下进行对比, 对本文模型的特点及上述优越性进行了说明和验证.

1 潜在动作理论

潜在动作 (Affordance) 理论由美国感知心理

学家 Gibson 于 1977 年提出, 认为婴儿在环境学习过程中, 首先学习的是物体的潜在动作, 如箱子“可堆积”、椅子“可坐”等, 之后才学习物体颜色、大小等属性^[19]. 21 世纪, 潜在动作理论被引入发育机器人研究领域^[20], 对此, 欧盟还成立了专门的研究项目^[21], RSS、ECCV 等会议也举办了相关的 Workshop^[22]. 国内有关潜在动作的相关研究主要来自华南理工大学易长安等的工作^[23-25].

潜在动作理论重点研究机器人与环境之间可能的动作关联, 从而实现机器人完成不同的任务.

1.1 潜在动作定义

潜在动作理论自提出以来, 各学者对其理解不一, Turvey 将潜在动作定义为环境的属性^[26]; 在 Turvey 理论上, Stoffregen 认为潜在动作是存在于动物-环境系统中的某种属性^[27]; Chemero 认为潜在动作是动物属性与环境属性之间的关联^[28]; Steedman 忽略感知作用, 将潜在动作理解为环境与动作的关联^[29]; 2015 年, 易长安等指出, 潜在动作是指机器人结合自身行为能力及感知能力, 判断其在当前环境下可执行的动作^[30]. 机器人在与环境的交互过程中逐渐学习到不同环境中的潜在动作, 从而完成不同的任务, 并在任务学习过程中不断提高自身行为学习能力. 由此可见, 潜在动作是机器人学习到的关于其周身环境的知识, 是机器人获得高级技能的重要基础, 它能够使机器人预测动作结果, 实现高效率的学习^[31].

1.2 潜在动作形式化

2007 年, Sahin 等^[32] 总结潜在动作相关知识, 定义潜在动作是效果和 (实体, 行为) 组之间的关系, 并给出了潜在动作学习的形式化, 具体为一个三元组, 如式 (1) 所示, 为广大学者所采用. 式 (1) 具体表示当智能体对实体 (*entity*) 执行行为 (*behavior*) 后, 产生效果 (*effect*).

$$(effect, (entity, behavior)) \quad (1)$$

2015 年, 易长安等对 Sahin 等的潜在动作学习模型进行扩展, 针对动态环境下的复杂任务, 提出了基于子任务的潜在动作描述方法^[30], 如式 (2) 所示, 其中 *precondition* 表示动作执行前环境需要满足的前置条件, *postcondition* 表示动作执行完成后环境需满足的后置条件. 在该模型之上, 易长安等提出潜在动作预测框架, 集成了分层强化学习、状态抽象机制、任务图和物体属性等, 提高了机器人学习效率.

Affordance :

$$(subtask, action, precondition, postcondition) \quad (2)$$

机器人与环境的每次交互都可以产生一个潜在动作元组, 多次交互可以得到更一般的关联, 从而完成更复杂的任务。

2 具有发育机制的感知行动认知模型

2.1 模型结构

本文在文献 [18] 的基础上, 结合潜在动作理论, 为机器人设计了一种结构可变的具有发育机制的感知行动认知模型 D-SSCM (Development-sensorimotor cognitive model), 其结构如图 1 所示, 包含离散学习时间集 t 、内部可感知离散状态集 S 、可输出动作集 M 、有效输出动作空间集 M_s 、有效感知行动映射取向性集 O_s 、有效感知行动映射学习次数集 N_s 、有效感知行动映射好奇心集 C_s 、状态评价函数 V 、取向函数 V_s 、有效操作函数集 P_s 、有效动作空间取向性学习算法 L_s 、潜在动作关系集 AF 、可输出动作空间探索率集 Exp 以及发育算法 DL 共 14 部分 (下标 s 表示感知 sensory, 后文下标 m 表示运动 motor)。

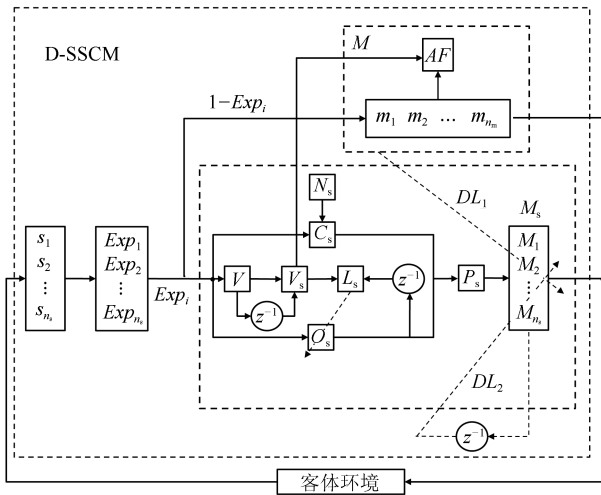


图 1 D-SSCM 结构图

Fig. 1 Structure of D-SSCM

为更清楚地显示 D-SSCM 中各元素的含义, 给出如下定义:

定义 1. 具有发育机制的感知行动认知模型 D-SSCM 是一个 14 元组: $D-SSCM = \langle t, S, M, M_s, O_s, N_s, C_s, V, V_s, P_s, L_s, AF, Exp, DL \rangle$, 其中各元素含义具体如下:

1) $t \in \{0, 1, \dots, n_t\}$: D-SSCM 离散学习时刻集, 其中 $t = 0$ 表示学习初始时刻, n_t 表示最大离散学习时刻数;

2) $S = \{s_i | i = 1, 2, \dots, n_s\}$: D-SSCM 内部可感知离散状态集, 其中 $s_i \in S$ 表示模型第 i 个可感知的内部状态, n_s 为离散状态数;

3) $M = \{m_j | j = 1, 2, \dots, n_m\}$: D-SSCM 可输出动作集, m_j 表示可输出动作集中第 j 个动作, n_m 为动作空间可输出动作数;

4) $M_s = \{M_i | i = 1, 2, \dots, n_s\}$: D-SSCM 有效输出动作空间集, $M_i = \{m_{ik} | k = 1, 2, \dots, n_i\}$ 为状态 s_i 下的有效输出动作空间, $m_{ik} \in M$ 为 D-SSCM 在状态 s_i 下从 M 中学习到的第 k 个有效动作, n_i 为状态 s_i 下学习到的有效动作个数. 状态 s_i 下的有效输出动作指的是该状态下能够使机器人趋向任务目标的动作, M_i 是在机器人对环境的学习过程中不断构建的, 随着 M_i 结构的不断变化, n_i 也随之发生变化, 体现出模型发育的思想. 学习初始时刻, M_i ($i = 1, 2, \dots, n_s$) 均为空, $n_i = 0$ ($i = 1, 2, \dots, n_s$);

不同于文献 [18] 中所设计模型需要学习的动作空间固定, 在 D-SSCM 中各状态 s_i 所对应的动作空间 M_i 并非教师或专家根据经验提前定义的, 而是随着学习过程渐进发育形成的, 该模式下, 有效避免了冗余感知行动映射造成的学习浪费和计算浪费.

5) $O_s = \{O_i | i = 1, 2, \dots, n_s\}$: D-SSCM 有效感知行动映射取向性集, 其中 $O_i = \{o_{ik} | k = 1, 2, \dots, n_i\}$ 为状态 s_i 下的有效感知行动映射取向性集, o_{ik} 为状态 s_i 对其第 k 个有效动作的选择取向性;

6) $N_s = \{N_i | i = 1, 2, \dots, n_s\}$: D-SSCM 有效感知行动映射学习次数集, $N_i = \{n_{ik} | k = 1, 2, \dots, n_i\}$ 为状态 s_i 下模型对其各有效动作的学习次数集, n_{ik} 表示状态 s_i 对动作 m_{ik} 的学习次数, 若 t 时刻, m_{ik} 被选择, 则 $t + 1$ 时刻:

$$n_{ik}(t + 1) = n_{ik}(t) + 1 \quad (3)$$

对于其他所有没有被学习的有效感知行动映射, 对应学习次数保持不变;

7) $C_s = \{C_i | i = 1, 2, \dots, n_s\}$: D-SSCM 有效感知行动映射好奇心集, $C_i = \{c_{ik} | k = 1, 2, \dots, n_i\}$ 为状态 s_i 下模型对其各有效动作的好奇心集, c_{ik} 表示状态 s_i 对动作 m_{ik} 的好奇度, 计算方式同文献 [18], 具体为:

$$c_{ik} = \frac{1}{1 + e^{k_c(n_{ik} - c)}} \quad (4)$$

其中, k_c 和 c 为好奇心参数;

8) V : D-SSCM 状态评价函数, 用来评价模型当前感知状态的理想程度, 机器人越接近学习目标, 模型对应感知状态的状态值越大, 机器人越远离学习目标, 则模型对应感知状态的状态值越小;

9) V_s : D-SSCM 取向函数, 用于决定模型学习

方向, 定义为:

$$V_s(t+1) = V(t+1) - V(t) \quad (5)$$

10) $P_s = \{P_i | i = 1, 2, \dots, n_s\}$: D-SSCM 有效操作函数集, 用于决定模型在有效动作空间内对动作的选择, $P_i = \{p_{ik} | k = 1, 2, \dots, n_i\}$ 为状态 s_i 所对应的有效操作函数集, p_{ik} 为状态 s_i 对动作 m_{ik} 的操作值, 具体为:

$$p_{ik} = \sigma o_{ik} + (1 - \delta)c_{ik} \quad (6)$$

$0 < \sigma < 1$ 为操作函数参数.

11) L_s : D-SSCM 有效动作空间取向性学习算法, 在 D-SSCM 有效动作空间内各动作均能使得 $V_s(t+1) \geq 0$, 因此针对有效动作空间的取向性学习算法 L_s 简单设计为:

$$L_s: \begin{cases} o_{ik}(t+1) = \frac{o_{ik}(t) + o_{ik}(t) \times (1 - e^{-\eta \times V_s(t+1)})}{1 + o_{ik}(t) \times (1 - e^{-\eta \times V_s(t+1)})} \\ o_{ik'}(t+1) = \frac{o_{ik'}(t)}{1 + o_{ik}(t) \times (1 - e^{-\eta \times V_s(t+1)})} \end{cases} \quad (7)$$

其中 o_{ik} ($k \in 1, \dots, n_i$) 对应被选动作 m_{ik} 的取向性值, $o_{ik'}$ 为状态 s_i 下其余动作的取向性值, η 为取向性学习算法参数.

12) $AF = \{AF_{ij} | i = 1, 2, \dots, n_s, j = 1, 2, \dots, n_m\}$: D-SSCM 潜在动作关系集, 受潜在动作理论及其形式化启发, 在此定义 D-SSCM 不同状态与不同动作之间的潜在关系, 具体定义为一个三元组:

$$AF_{ij}: (effect, (s_i, m_j)) \quad (8)$$

AF_{ij} 意义为 D-SSCM 在状态 s_i 下输出 m_j 产生的效果为 $effect$. D-SSCM 在对可输出动作空间进行探索时, 每探索一个新的感知行动映射 (s_i, m_j) , 就会伴随着一个新的潜在动作关系组形成.

针对 D-SSCM 发育式学习过程:

若 $effect = 1$, 表示在感知状态 s_i 下动作 m_j 是可被选择的, 即动作 m_j 是状态 s_i 下的有效动作;

若 $effect = 0$, 表示在状态 s_i 下, 动作 m_j 是不可取的, 会使得学习偏离目标;

若 $effect$ 值为空, 即 $effect = \emptyset$, 则表示对应的感知行动映射还没有被学习.

依据操作条件反射机制及定义的取向函数, $effect$ 值计算如下:

$$effect = \begin{cases} 1, & V_s \geq 0 \\ 0, & V_s < 0 \end{cases} \quad (9)$$

13) $Exp = \{Exp_i | i = 1, 2, \dots, n_s\}$: D-SSCM 可输出动作空间探索率集, Exp_i 表示模型在状态 s_i 下对可输出动作空间 M 的探索率, 可通过式 (10)

进行计算:

$$Exp_i = \frac{Num_i\{effect \neq \emptyset\}}{n_m} \quad (10)$$

其中, $Num_i\{effect \neq \emptyset\} = n_m - Num_i\{effect = \emptyset\}$ 表示状态 s_i 下可输出动作空间中潜在动作关系元组 $effect \neq \emptyset$ 的个数, 亦即状态 s_i 对可输出动作空间已经进行探索的感知行动映射数.

D-SSCM 中, 在任意状态 s_i 下, 模型都可以选择对该状态下已发育形成的有效输出动作空间集 M_i 进行学习, 或对该状态下可输出动作空间 M 的剩余空间进行探索. 在此规定, D-SSCM 学习过程中, 在状态 s_i 下模型总是以概率 $1 - Exp_i$ 对 M 剩余空间进行探索, 以概率 Exp_i 对其有效输出动作空间 M_i 进行学习. 特别地:

a) 在学习初始时刻 $t = 0$ 时, 模型没有任何环境知识, 任意状态 s_i 下, 其潜在动作关系元组 AF_{ij} ($j = 1, 2, \dots, n_m$) 中 $effect$ 值均为 \emptyset , M_i 也为 \emptyset , 此时 D-SSCM 以

$$1 - Exp_i = 1 - \frac{Num_i\{effect \neq \emptyset\}}{n_m} = 1 - \frac{0}{n_m} = 1 \quad (11)$$

的概率从可输出动作空间集 M 中探索动作, 以完成对 M_i 的构建.

b) 当学习进行到某个时刻, 若状态 s_i 已经完成了对可输出动作空间 M 的全部探索, 则 D-SSCM 在该状态下继续对 M 进行探索的概率为:

$$1 - Exp_i = 1 - \frac{Num_i\{effect \neq \emptyset\}}{n_m} = 1 - \frac{n_m}{n_m} = 0 \quad (12)$$

表明此时模型不再会对 M 进行探索.

14) DL : D-SSCM 发育算法, 包含 DL_1 和 DL_2 两部分, 其中 DL_1 为 D-SSCM 探索可输出动作空间 M 后需要对有效输出动作空间集 M_s 进行扩展的发育算法, DL_2 是 D-SSCM 对有效输出动作空间集 M_s 不断学习后需要对其进行缩减的发育算法.

2.2 发育算法设计及分析

2.2.1 扩展发育算法 DL_1

在状态 s_i 下, 若 D-SSCM 以概率 $1 - Exp_i$ 对可输出动作集 M 剩余动作空间进行了探索, 可能出现两种情况:

1) t 时刻在当前状态下探索某一动作后, $t + 1$ 时刻模型受到负强化 $V_s(t+1) < 0$, 则对应感知行

动映射潜在动作关系元组 $effect = 0$, 表明该动作在当前状态下是不应该被选择的, 针对该情况, $t + 1$ 时刻不需要对已建立的有效感知行动映射动作集进行更新;

2) 若 t 时刻模型在当前状态下探索某一动作后, $t + 1$ 时刻获得正强化 $V_s(t + 1) \geq 0$, 则对应感知行动映射潜在动作关系元组 $effect = 1$, 证明该动作是当前状态下的一个有效动作, $t + 1$ 时刻需要对 M_s 进行扩展发育。

以状态 s_i 为例, D-SSCM 在情况 2) 下 M_s 的结构扩展发育过程如图 2 所示。

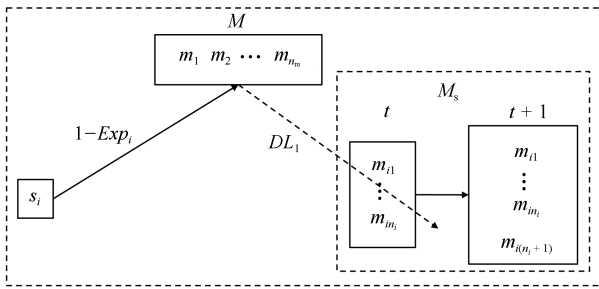


图 2 D-SSCM 扩展发育原理图

Fig. 2 D-SSCM extended development diagram

更具体的, 假设 t 时刻, 状态 s_i 所对应有效输出动作空间集 M_i 中的有效动作个数为 n_i , 模型以 $1 - Exp_i$ 探索了可输出动作集 M 剩余空间中某一动作, 设为 m_j , $t + 1$ 时刻获得正强化, 因此需要对 M_i 进行扩展发育, 具体如下:

步骤 1. 动作扩展: 将动作 m_j 扩展为动作集 M_i 第 $n_i + 1$ 个有效动作:

$$m_j \rightarrow m_{i(n_i+1)} \quad (13)$$

步骤 2. 取向性更新: 按照发育算法 DL_1 对扩展后的动作集 M_i 取向性集 O_i 进行更新, 其中 DL_1 算法学习过程具体如下:

步骤 2.1. 首先针对新增加动作 $m_{i(n_i+1)}$, 定义其在状态 s_i 下的取向性:

$$o_{i(n_i+1)}(t + 1) = \frac{1}{n_i + 1} \quad (14)$$

步骤 2.2. M_i 中原有动作取向性 o_{ik} ($k = 1, \dots, n_i$) 更新如下:

$$o_{ik}(t + 1) = \frac{n_i}{n_i + 1} o_{ik}(t) \quad (15)$$

步骤 3. 好奇心激活: 伴随着状态 s_i 下新动作 $m_{i(n_i+1)}$ 的增加, 除取向性按式 (14) 和式 (15) 进行更新外, 状态 s_i 对 $m_{i(n_i+1)}$ 的好奇心也随之被激活, 用于计算动作好奇度的感知行动映射学习次数

$n_{i(n_i+1)} = 1$, 并在以后的学习中不断更新. 可以看出, 任何新发育的动作, 其所对应的好奇心值均较大, 结合对新增动作定义的初始取向性值, 能够保证模型对该有效动作的充分学习。

步骤 4. 结构发育:

$$n_i \leftarrow n_i + 1 \quad (16)$$

DL_1 算法分析: D-SSCM 在结构不需要发育时, 其取向性学习算法如 L_s 所示, 算法有效性在文献 [18] 中给出了相关证明, 在此不再赘述. $t + 1$ 时刻, 若模型结构需要发育, 则发育后模型中动作的取向性应满足 L_s 的学习条件, 即: 1) $0 \leq o_{ik}(t + 1) \leq 1$ ($k = 1, 2, \dots, (n_i + 1)$), 2) $\sum_{k=1}^{n_i+1} o_{ik}(t + 1) = 1$.

针对条件 1): $t + 1$ 时刻, 对于新扩展的动作 $m_{i(n_i+1)}$ 的取向性有:

$$0 \leq o_{i(n_i+1)}(t + 1) = \frac{1}{n_i + 1} \leq 1 \quad (17)$$

满足条件, 对于 M_i 中原有动作的取向性有:

$$o_{ik}(t + 1) = \frac{n_i}{n_i + 1} o_{ik}(t), k = 1, \dots, n_i \quad (18)$$

在 $0 \leq o_{ik}(t) \leq 1$ 的情况下, 因为 $0 \leq \frac{n_i}{n_i+1} \leq 1$, 可知 $0 \leq o_{ik}(t + 1) \leq 1$ 成立, 因此 DL_1 发育算法满足条件 1)。

针对条件 2), $t + 1$ 时刻:

$$\begin{aligned} \sum_{k=1}^{n_i+1} o_{ik}(t + 1) &= \sum_{k=1}^{n_i} o_{ik}(t + 1) + o_{i(n_i+1)}(t + 1) = \\ &= \sum_{k=1}^{n_i} \frac{n_i}{n_i + 1} o_{ik}(t) + \frac{1}{n_i + 1} = \\ &= \frac{n_i}{n_i + 1} \sum_{k=1}^{n_i} o_{ik}(t) + \frac{1}{n_i + 1} = 1 \end{aligned} \quad (19)$$

可见 DL_1 发育算法满足条件 2)。

从以上扩展发育算法 DL_1 可以看出, 在潜在动作关系元组引入的条件下, 相比较于文献 [18] 认知模型, D-SSCM 对各动作的初步探索可一次完成, 通过计算 $effect$ 的值仅对各状态下的有效动作进行发育用于后期的学习, 大大降低了对感知行动映射的学习成本以及取向性更新时的计算成本, 同时模型对整个可输出动作集的探索依概率进行, 从另一方面缩减了计算成本。

2.2.2 缩减发育算法 DL_2

D-SSCM 认知模型在状态 s_i 下对其有效感知行动映射动作集 M_i 进行学习, 遵循内动机下的主动学习机制. M_i 中, 对所有的动作都有潜在动作

关系元组:

$$(1, (s_i, m_{ik})) \quad (20)$$

成立,但在操作条件反射机制下,模型总是趋向于选择获得更大正强化值的动作,随着学习的进行,某些有效动作的取向性会随着学习的进行不断降低,当其取向值下降到一定程度以下时,D-SSCM在好奇心作用下,依据内发动机机制继续对其进行学习会同时造成学习和计算的浪费,此时需要对模型结构进行缩减发育。

D-SSCM在对其有效输出动作空间集 M_s 的学习过程中,如果 t 时刻,当前状态 s_i 下某动作 m_{ik} 所对应的取向性 $o_{ik}(t)$ 满足:

$$o_{ik}(t) \leq \frac{1}{10 \times n_i} \quad (21)$$

则认为动作 m_{ik} 不属于当前状态下的可选潜在最优动作,需要将其从 M_i 中剪除.不同于结构扩展发育时直接将有效动作 m_j 作为 M_i 第 $n_i + 1$ 个有效动作,D-SSCM的结构缩减发育经历一个比较复杂的过程,具体如图3所示。

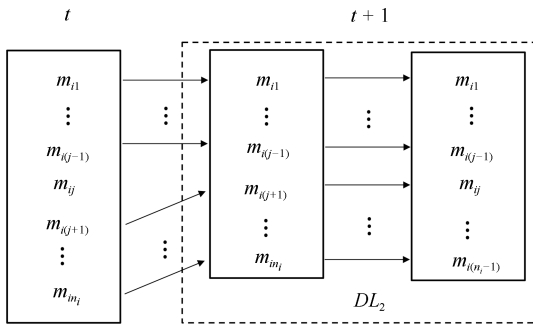


图3 D-SSCM 缩减发育原理图

Fig.3 D-SSCM reduced development diagram

以 t 时刻,状态 s_i 下的有效输出动作集 M_i 中第 j ($j \in 1, \dots, n_i$) 个动作 m_{ij} 需要被剪除为例,D-SSCM的缩减发育原理具体如下:

步骤 1. 动作剪除: 首先将 m_{ij} 从 M_i 中剪除。

步骤 2. 取向性更新: M_i 中动作 m_{ij} 被剪除后,对其剩余动作取向性按发育算法 DL_2 进行更新,具体为:

$$o_{ik}(t+1) = \frac{o_{ik}(t)}{1 - o_{ij}(t)} \quad (22)$$

步骤 3. 结构发育: 对剪除动作 m_{ij} 后的 M_i 结构进行更新。

步骤 3.1. 对于 $k < j$ 的动作各自排序不变,对于 $k > j$ 的动作,依次向前移动:

$$m_{ik} \rightarrow m_{i(k-1)} \quad (23)$$

直至

$$m_{in_i} \rightarrow m_{i(n_i-1)} \quad (24)$$

步骤 3.2. 将 M_i 空间由 n_i 维降低为 $n_i - 1$ 维:

$$n_i \leftarrow n_i - 1 \quad (25)$$

结构发育过程中,各动作相关性质如取向性、好奇心等随动作更新。

DL_2 算法分析:与 DL_1 算法相同, DL_2 算法同样需要保证 L_s 的学习条件: 1) $0 \leq o_{ik}(t+1) \leq 1$ ($k = 1, 2, \dots, (n_i - 1)$), 2) $\sum_{k=1}^{n_i-1} o_{ik}(t+1) = 1$ 成立.在结构缩减发育下,可简单描述为: $t+1$ 时刻上述步骤 2 中 1) $0 \leq o_{ik}(t+1) \leq 1$, ($k = 1, \dots, n_i$), 且 $k \neq j$, 2) $\sum_{k=1}^{j-1} o_{ik}(t+1) + \sum_{k=j+1}^{n_i} o_{ik}(t+1) = 1$.

针对条件 1),由于对任意动作 m_{ik} ,其取向性满足 $o_{ik}(t) \leq 1 - o_{ij}(t)$,因此有:

$$0 \leq o_{ik}(t+1) = \frac{o_{ik}(t)}{1 - o_{ij}(t)} \leq 1 \quad (26)$$

成立,满足条件。

针对条件 2), $t+1$ 时刻:

$$\begin{aligned} & \sum_{k=1}^{j-1} o_{ik}(t+1) + \sum_{k=j+1}^{n_i} o_{ik}(t+1) = \\ & \sum_{k=1}^{j-1} \frac{o_{ik}(t)}{1 - o_{ij}(t)} + \sum_{k=j+1}^{n_i} \frac{o_{ik}(t)}{1 - o_{ij}(t)} = \\ & \frac{\sum_{k=1}^{j-1} o_{ik}(t) + \sum_{k=j+1}^{n_i} o_{ik}(t)}{1 - o_{ij}(t)} = \\ & \frac{1 - o_{ij}(t)}{1 - o_{ij}(t)} = 1 \end{aligned} \quad (27)$$

条件 2) 成立。

学习中的某一时刻,若有效感知行动映射取向性集 O_i 中同时存在多个需要被剪除的动作,则按顺序依次重复上述过程。

2.3 学习过程

具有发育机制的感知行动认知模型 D-SSCM 学习步骤可总结如下:

步骤 1. 初始化. 学习初始时刻 $t = 0$ 时,对 D-SSCM 中各元素进行初始化,具体包括:定义模型内部可感知离散状态集 S 及可输出动作集 M ,定义模型状态评价函数 V ,设置学习相关参数,设置学习终止条件。

不同于文献 [18] 认知模型, D-SSCM 中,其要学习的感知行动映射结构是可发育的,有效感知行动映射相关元素是随着学习不断变化的,无需提前

定义, 以节省计算空间, 具体包括有效感知行动映射取向性集 O_s 、有效感知行动映射学习次数集 N_s 、有效感知行动映射好奇心集 C_s 、有效操作函数集 P_s 。

步骤 2. 状态感知. 观察学习 t 时刻 D-SSCM 模型状态 $s_i(t)$, 并计算当前状态取向值 $V(t)$ 。

步骤 3. 计算当前状态下的探索率 Exp_i . 依概率 $1 - Exp_i$ 对 M 剩余动作空间进行探索, 依概率 Exp_i 对 M_i 有效动作空间进行学习. 初始时刻, D-SSCM 没有任何先验知识, M_i 为 \emptyset , 模型以概率 1 探索 M , 即模型需要从 M 中获取知识, 同时, 在此规定, 在学习任意时刻, 若 M_i 为 \emptyset , 则模型以概率 1 探索 M 剩余动作空间。

所谓状态 s_i 下 M 的剩余动作空间, 即状态 s_i 在可输出动作集 M 中所有潜在动作关系元组 $(effect, (s_i, m_j))$ 中 $effect$ 值为 \emptyset 的动作, $effect$ 值为 \emptyset 表示感知行动映射 (s_i, m_j) 没有被探索。

t 时刻, 若 D-SSCM 依概率 $1 - Exp_i$ 对剩余动作空间进行了探索, 则执行步骤 4.1.1 ~ 步骤 4.1.5:

步骤 4.1.1. 选择动作并输出. D-SSCM 在当前状态的 M 剩余动作空间中随机选择某个动作并输出;

步骤 4.1.2. 状态发生转移. t 时刻, 模型在当前状态 $s_i(t)$ 下从 M 剩余动作空间中随机选择了某个动作, 假设为 m_j 作用于客体环境, 状态发生转移, 观测 $t + 1$ 时刻模型的新状态, 并计算其状态值 $V(t + 1)$;

步骤 4.1.3. 计算取向函数值 $V_s(t + 1)$. 根据 t 时刻及 $t + 1$ 时刻观测到的状态值计算模型的取向函数值 $V_s(t + 1)$, 此处用于计算模型潜在动作关系 $effect$ 值;

步骤 4.1.4. 更新模型潜在动作关系集. 根据计算获得的取向函数 $V_s(t + 1)$ 值更新模型 M 空间潜在动作关系元组 $(effect, (s_i, m_j))$, 若 $V_s(t + 1) \geq 0$, 则 $effect = 1$, 若 $V_s(t + 1) < 0$, 则 $effect = 0$;

步骤 4.1.5. 判断是否扩展发育. 根据新生成的潜在动作关系元组的 $effect$ 值判断是否需要状态 s_i 下的有效动作空间 M_i 进行扩展发育, 若 $effect = 0$, 表明所探索的动作在当前状态下是无效的, $t + 1$ 时刻无需对 M_i 进行扩展发育, D-SSCM 有效感知行动映射结构不变; 若 $effect = 1$, 则表明 t 时刻探索到当前状态下一个有效动作, $t + 1$ 时刻需要将探索到的动作扩展到 M_i 中, 具体按第 2.2.1 节中扩展发育算法对模型结构及相关属性进行更新。

t 时刻, 若 D-SSCM 依概率 Exp_i 对当前状态 s_i 下有效动作空间 M_i 进行学习, 则执行步骤 4.2.1 ~ 步骤 4.2.7, D-SSCM 对 M_i 中各动作的学习依内发动机机制进行^[18], 具体为:

步骤 4.2.1. 计算当前状态下的好奇心集 $C_i(t)$ 。

在感知状态 $s_i(t)$ 下, D-SSCM 有效输出动作空间 M_i 中某一动作, 设为 m_{ik} ($k \in 1, 2, \dots, n_i$) 随机引起了模型对其进行学习的好奇心, 其好奇心被激发, 计算该好奇心值 $c_{ik}(t)$, 对于没有引起模型好奇心的其余动作, $c_{ik'}(t) = 0$;

步骤 4.2.2. 计算操作函数集 $P_i(t)$. 结合模型当前状态下的有效感知行动映射取向性集 $O_i(t)$ 及好奇心集 $C_i(t)$, 计算当前状态下的有效操作函数集 $P_i(t)$;

步骤 4.2.3. 选择动作并输出. 依据内发动机机制, 选择 M_i 中操作函数值最大的动作, 设为 m_{ib} ($b \in 1, 2, \dots, n_i$) 作用于环境中;

步骤 4.2.4. 状态发生转移. 感知模型 $t + 1$ 时刻新状态, 计算其状态值 $V(t + 1)$;

步骤 4.2.5. 计算取向函数值 $V_s(t + 1)$. 计算 $t + 1$ 时刻模型取向函数值 $V_s(t + 1)$, 此处用于决定模型已有结构下取向性学习方向;

步骤 4.2.6. 更新有效感知行动取向性映射集. 根据式 (7) 对有效感知行动取向性映射集 O_i 进行更新;

步骤 4.2.7. 判断是否缩减发育. 根据更新后的取向性映射集 O_i 判断是否需要 M_i 进行缩减发育, 当 O_i 中存在需要被剪除的动作时, 依据第 2.2.2 节缩减发育算法对模型结构及相关属性进行更新。

步骤 5. 判断学习结束条件. 根据设定的学习终止条件判断学习是否结束, 若满足条件, 则结束, 否则返回步骤 2。

D-SSCM 的学习过程可用流程图 4 更清楚直观地描述。

3 两轮机器人自平衡实验

3.1 D-SSCM 模型设置

针对两轮机器人自平衡任务, 首先需要对 D-SSCM 模型进行设置。

1) 定义模型可感知离散状态. 本文中, 机器人的自平衡主要指其在俯仰方向上的平衡. 通过倾角传感器和陀螺仪, 机器人可获知自身的倾角角度 (用 φ 表示) 以及倾角角速度 (用 $\dot{\varphi}$ 表示). 以平衡位置 (竖直向上) 为中心, 采用对称法, 可在俯仰方向上将机器人身姿角度 φ 和身姿角速度 $\dot{\varphi}$ 各划分为 12 个离散状态, 如表 1 所示, 因此 D-SSCM 共有 $n_s = 12 \times 12 = 144$ 个可感知状态。

2) 两轮机器人通过控制轮子的运动实现平衡, 因此模型可输出动作设定为轮子的转矩, 可输出动作集设计为 $M = \{-10, -5, -2, -1, -0.1, 0, 0.1, 1, 2, 5, 10\}$ (N·m), 机器人共有 $n_m = 11$ 个可输出动作;

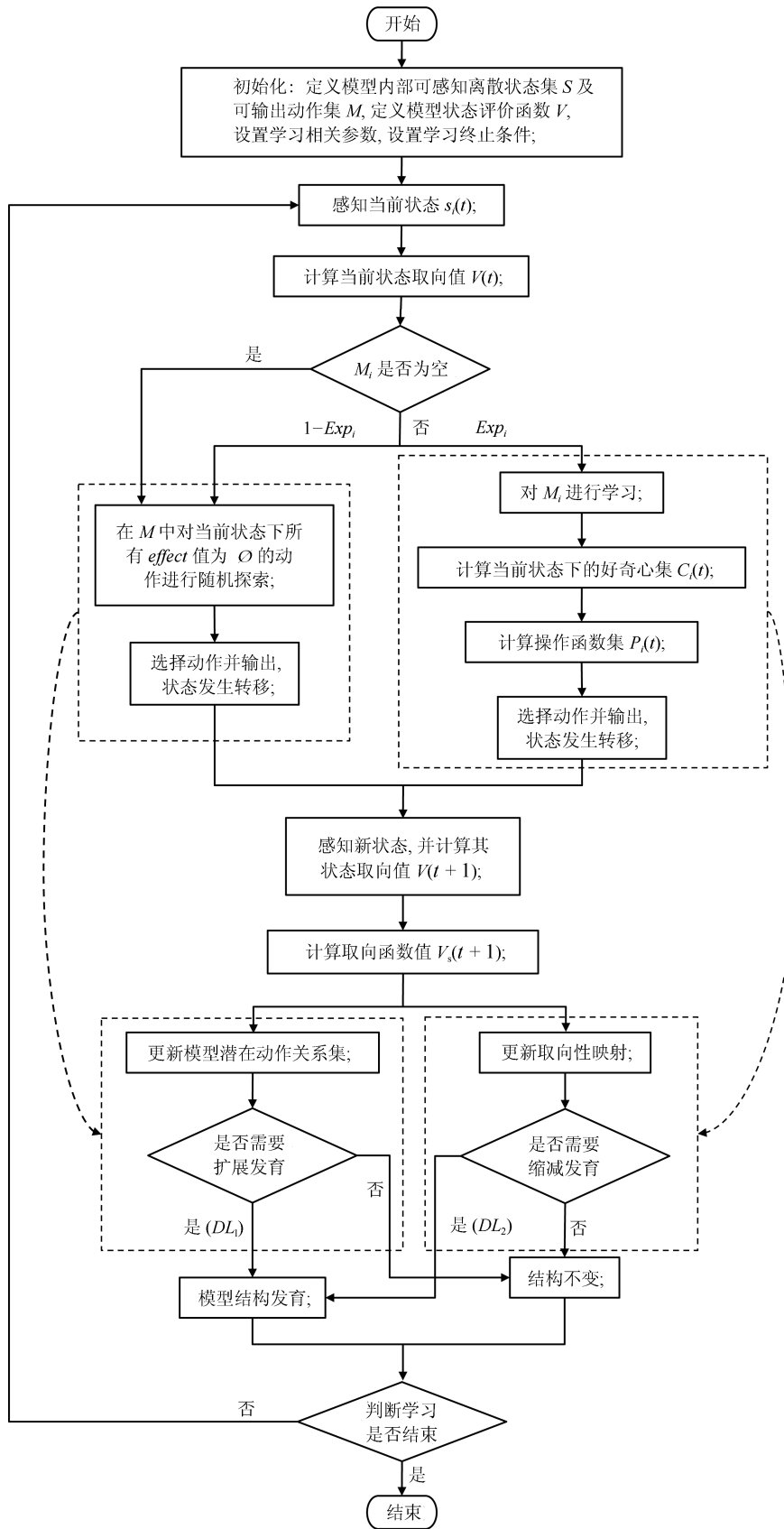


图 4 D-SSCM 学习流程图
Fig. 4 Learning flowchart of D-SSCM

表 1 D-SSCM 状态划分
Table 1 D-SSCM state division

$\varphi (^{\circ})$	$\dot{\varphi} (^{\circ}/s)$
$(-\infty, -17.5)$	$(-\infty, -100)$
$[-17.5, -12.5)$	$[-100, -50)$
$[-12.5, -7.5)$	$[-50, -20)$
$[-7.5, -2.5)$	$[-20, -5)$
$[-2.5, -0.5)$	$[-5, -2)$
$[-0.5, 0)$	$[-2, 0)$
$[0, 0.5)$	$[0, 2)$
$[0.5, 2.5)$	$[2, 5)$
$[2.5, 7.5)$	$[5, 20)$
$[7.5, 12.5)$	$[20, 50)$
$[12.5, 17.5)$	$[50, 100)$
$[17.5, +\infty)$	$[100, +\infty)$

3) 针对两轮机器人自平衡任务, 根据经验, 模型状态评价函数设计为:

$$v_i(t) = -225\varphi^2(t) - 81\varphi(t)\dot{\varphi}(t) - 9\dot{\varphi}^2(t) \quad (28)$$

4) 其他各相关参数具体设定为 $k_c = 0.05$, $c = 1$, $\delta = 0.7$, $\eta = 0.1$.

3.2 实验结果与分析

为表明 D-SSCM 学习的优越性, 将其与文献 [18] 认知模型在如上相同设置下进行对比实验验证. 方便起见, 文献 [18] 具有内发动机机制的感知行动认知模型简称为 IM-SSCM (Intrinsic motivation-sensorimotor cognitive model)

1) 基本学习过程: 令机器人由初始倾斜角度 -10° 开始学习, 采样时间为 $0.01s$, 学习过程中, 如果机器人身姿角度 $|\varphi(t)| > 15^{\circ}$, 则认为机器人发生倾倒, 将其拉回初始状态继续学习. 如图 5~图 7 所示分别为两轮机器人在 50000 步学习过程中其身姿倾斜角度、角速度以及轮子输出转矩的变化曲线. 可以明显看出: 1) D-SSCM 具有更快的学习速度: 从图 5 及图 6 机器人身姿角度和角速度曲线可以看出, 在 IM-SSCM 指导下, 机器人大约经过 150s 可进入平衡位置, 而在结构可发育的 D-SSCM 指导下, 机器人大约经过 40s 即可进入平衡位置, 这是因为在发育机制下, 机器人只对有效的感知行动映射进行学习, 大大节省了探索成本. 2) D-SSCM 具有更稳定的学习效果: 从图 5 及图 6 中可以看出, IM-SSCM 学习过程阶段性比较明显, 主要表现为学习初期 ($0s \sim 150s$) 以较大的好奇心对不同的感知行动映射进行探索, 学习中期 ($150s \sim 450s$), 好奇心得到一定下降, 模型以较小的好奇心对感知行动映射继续进行探索, 因此在学习中期依旧可能存在一

些小的波动, 相比之下, D-SSCM 对感知行动映射的学习仅在其有效输出动作空间中进行, 其中所有的动作都使得机器人趋向目标, 学习一旦完成, 机器人不会发生晃动, 该结论在图 7 机器人轮子转矩输出中体现更加明显, 从图 7 中可以看出, IM-SSCM 认知模型在 $150s \sim 450s$ 之间尽管对 ± 10 、 ± 5 的选择有所减少, 但是在好奇心作用下依旧可能会对不良感知行动映射进行尝试, 而在 D-SSCM 发育认知模型下, 机器人通过一次探索获知动作结果后, 后期就不再会选择各状态下的不良动作, 不至学习偏离目标.

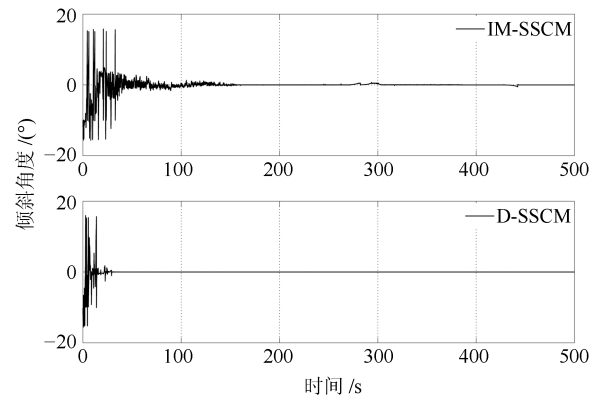


图 5 两轮机器人倾斜角度

Fig. 5 Angle of two-wheeled robot

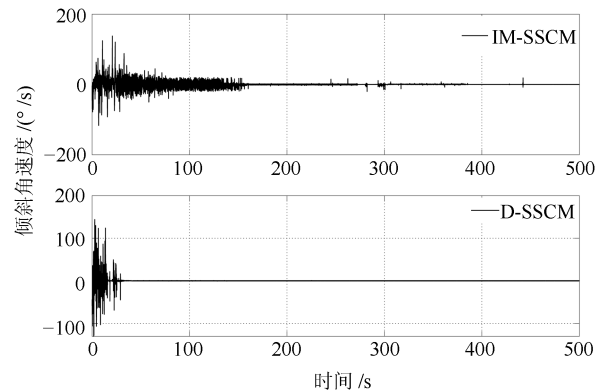


图 6 两轮机器人倾斜角速度

Fig. 6 Angular velocity of two-wheeled robot

从以上结果可以看出, D-SSCM 指导下机器人学习速度非常快, 大约 40s 即可完成, 为说明 D-SSCM 发育机制下的学习特点, 对机器人前 100s 学习过程中对 M 空间下感知行动映射探索次数及 M_s 空间有效感知行动映射构建数进行了记录, 结果如图 8 所示. 根据实验数据结果, D-SSCM 在初始 100s 学习中共探索 M 空间 660 次, 最终在 M_s 中形成有效感知行动映射数 179 条. 在 IM-SSCM 认知模型中, 机器人在整个探索阶段中需要探索的感知行动映射数始终为 $n_s \times n_m = 144 \times 11 = 1584$,

相比较之下, D-SSCM 在发育机制下, 对 M 空间下各感知行动映射的探索只需要一次, 通过计算所得的潜在动作关系元组 *effect* 值决定是否对当前感知行动映射进行发育, 其需要学习的感知行动映射数仅为 M_s 空间下的有效感知行动映射, 学习空间大大缩小.

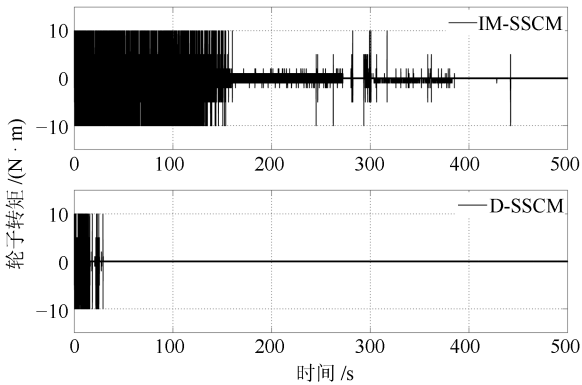


图 7 两轮机器人轮子转矩

Fig. 7 Wheel's torque of two-wheeled robot

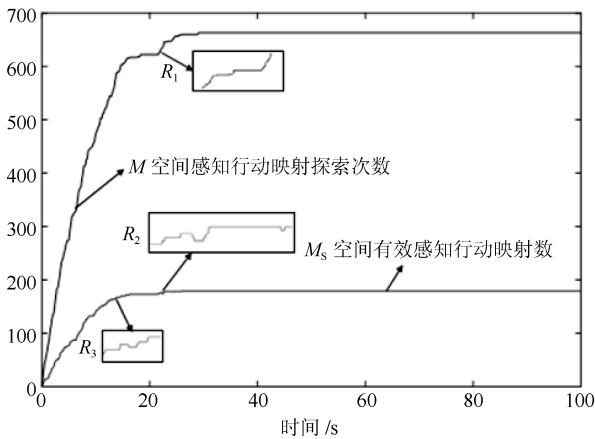


图 8 D-SSCM 发育过程实验结果图

Fig. 8 Experiment results figure of D-SSCM's development process

对 M 空间感知行动映射探索次数曲线进行分析, 首先可以看到在前 40s 的学习过程中, 曲线整体呈上升趋势, 表明机器人在不断探索 M 空间下的感知行动映射以获取新知识, 进而完成对 M_s 的构建. 此外, 曲线存在如图中 R_1 区域所示水平直线部分, M 空间感知行动映射探索次数曲线中水平直线部分的出现意味着该阶段 D-SSCM 正在对其所构建的有效输出动作空间集 M_s 进行学习, 表明模型对 M 空间的探索是依概率的, 而非遍历的. 智能系统学习的目标在于寻找某种可以达到目标的策略, 该策略不一定是最优的, 该目标下, 遍历式的搜索策略是不必要的, 反而会降低学习效率, D-SSCM 对 M 空间的依概率探索则很好地避免了以上问题.

对 M_s 空间有效感知行动映射数曲线进行分析. 不同于 M 空间感知行动映射探索次数表现为不减曲线, M_s 空间有效感知行动映射数在缩减式发育下可能会出现下降. 与 R_1 区域所处学习阶段对应, M_s 空间有效感知行动映射数变化过程如图中 R_2 区域所示, 需要指出的是, 该阶段中, R_2 区域出现了有效感知行动映射数减小的情况, 从实验角度体现了模型的缩减式发育过程, 同样的过程也发生在模型学习前期阶段, 具体如图 8 中 R_3 区域所示.

图 8 清楚的显示了 D-SSCM 的发育过程, 同时包含扩展式发育及缩减式发育.

2) 轮次学习: 两轮机器人的自平衡过程关键在于其由初始状态运动到平衡位置期间. 为更清楚地显示 D-SSCM 的学习能力, 令其不断在前一次学习的基础上从初始状态开始运动, 观察其学习结果, 并与 IM-SSCM 进行对比. 在此设定轮次学习步数为 3000 步, 结果如图 9~11 所示, 分别为两种模型下机器人第 1 轮、第 2 轮及第 3 轮的学习结果, 从图中可以看出, 相比较于 IM-SSCM, D-SSCM 具有更快的学习速度. 实验结果显示在具有发育机制的感知行动认知模型 D-SSCM 指导下, 机器人

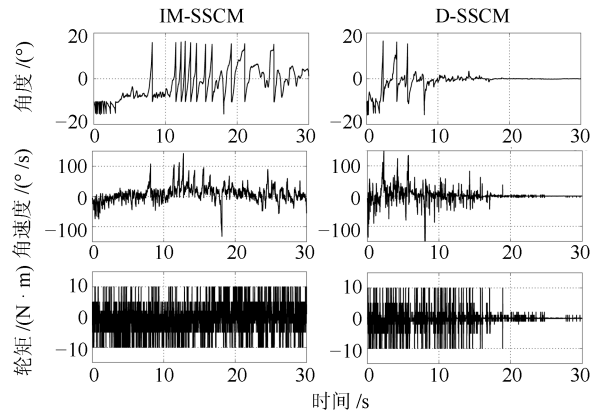


图 9 第 1 轮学习结果

Fig. 9 Learning results of the 1st round

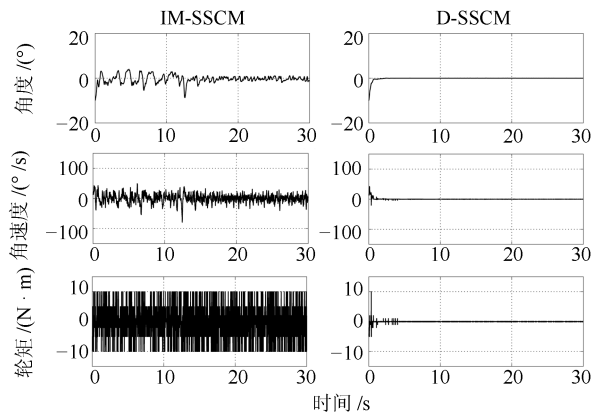


图 10 第 2 轮学习结果

Fig. 10 Learning results of the 2nd round

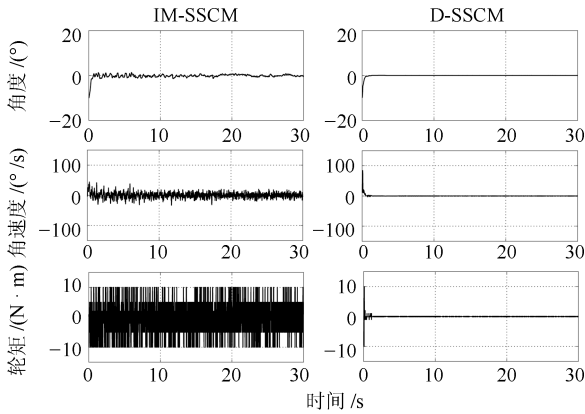


图 11 第 3 轮学习结果

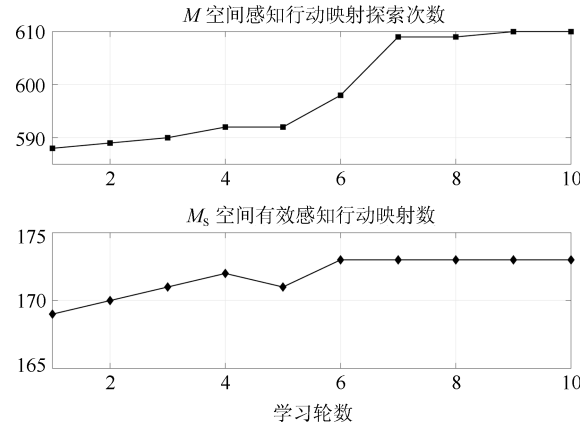
Fig. 11 Learning results of the 3rd round

一般经过一轮对知识的探索, 在第二轮就能够快速从初始倾斜角度运动到平衡状态, 该学习速度较其他已知操作条件反射相关认知模型有显著优势.

令机器人连续学习 10 轮, 对各轮学习过后 D-SSCM 模型对 M 空间感知行动映射的累计探索次数 (用 n_M 表示) 及其自身 M_s 空间有效感知行动映射数 (用 n_{M_s} 表示) 进行记录, 结果如图 12 所示, 可以看出, 不同轮次下, 机器人在不同程度上完成了对 M 空间的探索和对 M_s 空间的构建.

更具体地, 表 2 中数据与图 12 相对应, 对其进行分析, 首先, 经过第 1 轮的学习, 机器人探索了 M 空间下的 588 条感知行动映射, 并经过对其潜在动作关系进行分析, 在 M_s 空间下构建了 169 条有效感知行动映射; 进入第 2 轮及第 3 轮, n_M 数与 n_{M_s} 数均增加 1, 表明在这两轮学习过程中, 机器人都探索了一次 M 空间, 同时探索的感知行动映射有效, 对 M_s 进行了扩展发育; 进入第 4 轮, n_M 数增加 2, 而 n_{M_s} 仅增加了 1, 说明 D-SSCM 探索的两条感知行动映射中一条有效, 一条无效; 从第 4 轮到第 5 轮, n_M 数不变, n_{M_s} 减 1, 说明该轮次中, 模型主要在学习 M_s 有效感知行动映射空间, 并且在学习期间, 存在感知行动映射取向性小于一定值的情况, 因

此对模型进行了缩减式发育; 从第 6 轮到第 10 轮, n_M 值都不同程度的增加, n_{M_s} 没有再发生变化, 说明模型结构没有再发生变化. 表 2 中 D-SSCM 从第 1 轮到第 10 轮中 n_M 和 n_{M_s} 的变化情况, 更好、更清楚地说明了 D-SSCM 的发育过程.

图 12 10 轮学习中的 n_M 及 n_{M_s} 数Fig. 12 n_M and n_{M_s} in 10 learning rounds

4 结论

机器人感知行动认知模型结构固定情况下, 多存在学习浪费及计算浪费的问题, 对此, 本文在文献 [18] 的基础上, 结合潜在动作理论, 为机器人行为学习过程提出了一种结构可发育的感知行动认知模型 D-SSCM, 能够在探索可输出动作空间过程中自建需要学习的有效感知行动映射, 同时节省了学习成本和计算成本, 在实现机器人自主学习的同时, 提高了机器人学习速度和稳定性. D-SSCM 中, 需要学习的有效感知行动映射通过自建形成, 一定程度上提升了系统的智能性, 但是其中状态评价函数依旧是教师依据经验设计, 此外, D-SSCM 中, 机器人可输出动作为离散, 在一定程度上都限制了机器人的智能性和学习效果, 这都将成为本文下一步工作的重点.

表 2 10 轮学习中的 n_M 及 n_{M_s} 数Table 2 n_M and n_{M_s} in 10 learning rounds

学习轮数	1	2	3	4	5	6	7	8	9	10
M 空间感知行动映射探索次数	588	589	590	592	592	598	609	609	610	610
M_s 空间有效感知行动映射数	169	170	171	172	171	173	173	173	173	173

References

- 1 Tang H J, Yan R, Tan K C. Cognitive navigation by neuro-inspired localization, mapping and episodic memory. *IEEE Transactions on Cognitive and Developmental Systems*, 2018, 10(3): 751–761
- 2 Piaget J. *The Origins of Intelligence in Children*. New York: International Universities Press, 1952
- 3 Alexandrov A V, Lippi V, Mergner T, Frolov A A, Hettich G, Husek D. Human-inspired eigenmovement concept provides coupling-free sensorimotor control in humanoid robot. *Frontiers in Neurobotics*, 2017, 11: 22

- 4 Mirus F, Axenie C, Stewart T C, Conradt J. Neuromorphic sensorimotor adaptation for robotic mobile manipulation: From sensing to behaviour. *Cognitive Systems Research*, 2018, **50**: 52–66
- 5 Martius G, Fiedler K, Herrmann J M. Structure from behavior in autonomous agents. In: Proceedings of the 2008 IEEE/RSJ International Conference on Intelligent Robots and Systems. Nice, France: IEEE, 2008. 858–862
- 6 Ren H G, Liu C, Shi T. A computational model of cognitive development for the motor skill learning from curiosity. *Biologically Inspired Cognitive Architectures*, 2018, **25**: 101–106
- 7 Vallverdú J, Talanov M, Distefano S, Mazzara M, Tchitchigin A, Nurgaliev I. A cognitive architecture for the implementation of emotions in computing systems. *Biologically Inspired Cognitive Architectures*, 2016, **15**: 34–40
- 8 Bing Z S, Meschede C, Röhrbein F, Huang K, Knoll A C. A survey of robotics control based on learning-inspired spiking neural networks. *Frontiers in Neurobotics*, 2018, **12**: 35
- 9 Cai J X, Hong L, Cheng L N, Yu R H. Skinner operant conditioning model and robot bionic self-learning control. *Technical Gazette*, 2016, **23**(1): 65–75
- 10 Cyr A, Boukadoum M, Thériault F. Operant conditioning: a minimal components requirement in artificial spiking neurons designed for bio-inspired robot's controller. *Frontiers in neurobotics*, 2014, **8**: 21
- 11 Itoh K, Miwa H, Matsumoto M, et al. Behavior model of humanoid robots based on operant conditioning. In: Proceedings of the 5th IEEE-RAS International Conference on Humanoid Robots. Tsukuba, Japan: IEEE, 2005. 220–225
- 12 Mannella F, Santucci V G, Somogyi E, Jacquey L, O'Regan K J, Baldassarre G. Know your body through intrinsic goals. *Frontiers in Neurobotics*, 2018, **12**: 30
- 13 Baranes A, Oudeyer P Y. Intrinsically motivated goal exploration for active motor Learning in robots: a case study. In: Proceedings of IEEE/RSJ 2010 International Conference on Intelligent Robots and Systems. Taipei, China: IEEE, 2010. 1766–1773
- 14 Baranes A, Oudeyer P Y. Active learning of inverse models with intrinsically motivated goal exploration in robots. *Robotics and Autonomous Systems*, 2013, **61**(1): 49–73
- 15 Castellanos S, Rodríguez L F, Gastro L A, Gutierrez-Garcia J O. A computational model of emotion assessment influenced by cognition in autonomous agents. *Biologically Inspired Cognitive Architectures*, 2018, **25**: 26–36
- 16 Weng J. Artificial intelligence: autonomous mental development by robots and animals. *Science*, 2001, **291**(5504): 599–600
- 17 Weng J. A model for auto-programming for general purposes. arXiv preprint arXiv: 1810.05764, 2018.
- 18 Zhang Xiao-Ping, Ruan Xiao-Gang, Xiao Yao, Zhu Xiao-Qing. Establishment of a two-wheeled robot's sensorimotor system with mechanism of intrinsic motivation. *Acta Automatica Sinica*, 2016, **42**(8): 1175–1184
(张晓平, 阮晓钢, 肖尧, 朱晓庆. 两轮机器人具有内发动机机制的感知运动系统的建立. 自动化学报, 2016, **42**(8): 1175–1184)
- 19 Gibson J J. *The Ecological Approach to Visual Perception*. Houghton Mifflin, 1979
- 20 Hsiao S W, Hsu C F, Lee Y T. An online affordance evaluation model for product design. *Design Studies*, 2012, **33**(2): 126–159
- 21 EU Project MACS [Online], available: <http://www.macs-eu.org/>, July 1, 2018
- 22 Af Net 2.0: The Affordance Network [Online], available: <http://affordances.info/workshops>, September 3, 2018
- 23 Yi C, Min H, Luo R, et al. A novel formalization for robot cognition based on affordance model. In: Proceedings of the 2000 IEEE International Conference on Robotics and Biomimetics. Guangzhou, China: IEEE, 2000. 145–146
- 24 Min H, Yi C, Luo R, et al. Affordance learning based on subtask's optimal strategy. *International Journal of Advanced Robotic Systems*, 2015, **12**: 8
- 25 Min H, Yi C, Luo R, et al. Goal-directed affordance prediction at the subtask level. *Industrial Robot*, 2016, **43**(1): 48–57
- 26 Turvey M T. Affordances and prospective control: An outline of the ontology. *Ecological Psychology*, 1992, **4**(3): 173–187
- 27 Stoffregen T A. Affordances as properties of the animal-environment system. *Ecological Psychology*, 2003, **15**(2): 115–134
- 28 Chemero A. An outline of a theory of affordances. *Ecological Psychology*, 2003, **15**(2): 181–195
- 29 Steedman M. Plans, affordances, and combinatory grammar. *Linguistics and Philosophy*, 2002, **25**(5/6): 723–753
- 30 Yi Chang-An, Min Hua-Qing, Luo Rong-Hua. Robot's affordance prediction based on the subtask. *Journal of Huazhong University of Science and Technology (Natural Science Edition)*, 2015, **s1**: 412–415
(易长安, 闵华清, 罗荣华. 基于子任务的机器人潜在动作预测. 华中科技大学学报(自然科学版), 2015, **s1**: 412–415)
- 31 Shen Xiao-Wen. The Research and Application of Hierarchical Reinforcement Learning and Affordance Model [Ph. D. dissertation], South China University of Technology, China, 2014
(沈孝文. 分层强化学习与潜在动作模型的研究与应用 [博士学位论文], 华南理工大学, 中国, 2014)
- 32 Sahin E, Cakmak M, Doğar M R, Uğur E, Ücoluk G. To afford or not to afford: A new formalization of affordances toward affordance-based robot control. *Adaptive Behavior*, 2007, **15**(4): 447–472



张晓平 北方工业大学电气与控制工程学院讲师. 主要研究方向为认知机器人. 本文通信作者.

E-mail: zhangxiaoping369@163.com

(**ZHANG Xiao-Ping** Lecturer at the School of Electrical and Control Engineering, North China University of Technology. Her research interest covers cognitive robotics. Corresponding author of this paper.)



李志军 北方工业大学电气与控制工程学院副教授. 主要研究方向为智能机器人与智能学习系统.

E-mail: lzj78@ncut.edu.cn

(**LI Zhi-Jun** Associate professor at the School of Electrical and Control Engineering, North China University of Technology. His research interest covers intelligent robot and intelligent learning system.)



阮晓钢 北京工业大学信息学部教授. 主要研究方向为人工智能与机器人.

E-mail: adrxg@bjut.edu.cn

(**RUAN Xiao-Gang** Professor at the Faculty of Information Technology, Beijing University of Technology. His research interest covers artificial intelligence and robotics.)



闫佳庆 北方工业大学电气与控制工程学院副教授. 主要研究方向为神经信息学. E-mail: yjq@ncut.edu.cn

(**YAN Jia-Qing** Associate professor at the School of Electrical and Control Engineering, North China University of Technology. His main research interest is neuroinformatics.)



王力 北方工业大学电气与控制工程学院教授. 主要研究方向为智能交通.

E-mail: Li.wang@ncut.edu.cn

(**WANG Li** Professor at the School of Electrical and Control Engineering, North China University of Technology. His main research interest is intelligent transportation.)



毕松 北方工业大学电气与控制工程学院副教授. 主要研究方向为智能机器人. E-mail: bisongo@163.com

(**BI Song** Associate professor at the School of Electrical and Control Engineering, North China University of Technology. His main research interest is intelligent robot.)