

# 控制与决策

Control and Decision

## 一种结合内在动机理论的移动机器人环境认知模型

阮晓钢, 张家辉, 黄静, 柴洁, 武悦

引用本文:

阮晓钢, 张家辉, 黄静, 等. 一种结合内在动机理论的移动机器人环境认知模型[J]. *控制与决策*, 2021, 36(9): 2211–2217.

在线阅读 View online: <https://doi.org/10.13195/j.kzyjc.2019.1744>

---

## 您可能感兴趣的其他文章

Articles you may be interested in

### 基于改进RRT\*FN算法的机器人路径规划

Robot path planning based on improved RRT\*FN algorithm

控制与决策. 2021, 36(8): 1834–1840 <https://doi.org/10.13195/j.kzyjc.2019.1713>

### 移动机器人运动规划中的深度强化学习方法

Deep reinforcement learning for motion planning of mobile robots

控制与决策. 2021, 36(6): 1281–1292 <https://doi.org/10.13195/j.kzyjc.2020.0470>

### 基于16方向24邻域改进蚁群算法的移动机器人路径规划

Mobile robots path planning based on 16-directions 24-neighborhoods improved ant colony algorithm

控制与决策. 2021, 36(5): 1137–1146 <https://doi.org/10.13195/j.kzyjc.2019.0600>

### 基于虚拟结构法的多移动机器人分布式预测控制

Distributed predictive control of multiple mobile robots based on virtual structure method

控制与决策. 2021, 36(5): 1273–1280 <https://doi.org/10.13195/j.kzyjc.2019.1136>

### 凸优化与A\*算法结合的路径避障算法

Convex optimization and A-star algorithm combined path planning and obstacle avoidance algorithm

控制与决策. 2020, 35(12): 2907–2914 <https://doi.org/10.13195/j.kzyjc.2019.0351>

# 一种结合内在动机理论的移动机器人环境认知模型

阮晓钢, 张家辉, 黄静<sup>†</sup>, 柴洁, 武悦

(1. 北京工业大学 信息学部, 北京 100124; 2. 北京工业大学  
计算智能与智能系统北京市重点实验室, 北京 100124)

**摘要:** 为了探讨动机在环境认知中所起作用, 模仿生物环境认知行为, 以学习自动机为框架, 提出一种结合内在动机理论的移动机器人环境认知模型 (CM-LA)。该模型以访问次数及访问时间间隔共同定义内在动机, 并以操作条件反射原理描述外在动机的影响。同时, 所设计的动作选择机制可以使机器人根据已有的环境认知对新增障碍物进行有效避让。实验结果表明, 所提出模型能够在两种动机的驱使下对环境进行有效的自主认知, 并提高机器人在未知环境中的探索效率, 搜索到的路径长度较短, 且具有一定的环境自适应能力。

**关键词:** 内在动机; 操作条件反射; 环境认知; 认知计算; 移动机器人

中图分类号: TP273

文献标志码: A

DOI: 10.13195/j.kzyjc.2019.1744

开放科学(资源服务)标识码(OSID):



**引用格式:** 阮晓钢, 张家辉, 黄静, 等. 一种结合内在动机理论的移动机器人环境认知模型 [J]. 控制与决策, 2021, 36(9): 2211-2217.

## An environment cognition model combined with intrinsic motivation for mobile robots

RUAN Xiao-gang, ZHANG Jia-hui, HUANG Jing<sup>†</sup>, CHAI Jie, WU Yue

(1. Faculty of Information Technology, Beijing University of Technology, Beijing 100124, China; 2. Beijing Key Laboratory of Computational Intelligence and Intelligent System, Beijing University of Technology, Beijing 100124, China)

**Abstract:** In order to explore the role of motivation in environmental cognition and imitate the environment cognitive behavior of creatures, this paper proposes a mobile robot environment cognitive model (CM-LA) combining the theory of intrinsic motivation with the framework of learning automata. The number of visits and the time interval between visits define the intrinsic motivation, and describe the impact of extrinsic motivation based on the principle of operant conditioning. At the same time, the designed action selection mechanism enables the robot to effectively avoid new obstacles based on existing environmental awareness. The experimental results show that the model can effectively and autonomously recognize the environment driven by two motivations, and improve the exploration efficiency of the robot in an unknown environment. The path length searched using this method is shorter with the ability of self-adaptation to the environment.

**Keywords:** intrinsic motivation; operant conditioning; environment cognition; cognitive computing; mobile robot

## 0 引言

智能移动机器人是一种能够对周围环境进行感知, 并完成自主导航任务的智能体。其中, 对未知环境的认知正是机器人进行环境建模、定位、规划等任务的基本前提<sup>[1]</sup>。如何采用仿生的方式, 赋予机器人类似人和动物的环境认知能力是当前研究的重点。

动机是驱使人或动物进行环境探索的根本动力, 主要可以分为外在动机和内在动机<sup>[2]</sup>。关于外在动机

的研究成果已经较为广泛<sup>[3-4]</sup>。其中, Skinner<sup>[5]</sup>提出的操作条件反射理论体现了生物在与环境交互过程中通过“反应-刺激-反应”模式逐渐产生对环境的自适应性, 是一种依靠外在动机不断强化而成的自主学习方法。阮晓钢等<sup>[6]</sup>提出一种基于操作条件反射的学习模型 OCLM, 在仿真环境中使机器人以“试错式”的方式在环境中学习并成功避障到达目标, 再现了人或动物的自主学习行为; 黄静等<sup>[7]</sup>借鉴该理论, 提出

收稿日期: 2019-12-13; 修回日期: 2020-06-05.

基金项目: 国家自然科学基金项目(61773027); 工信部 2018 年工业互联网创新发展工程项目(Z135060009002).

责任编辑: 高会军.

<sup>†</sup>通讯作者. E-mail: huangjing@bjut.edu.cn.

了一种具有操作条件反射机能的人工感觉系统,并成功模拟了斯金纳鸽子实验及维纳机器蠕虫实验,使机器人具有生物的自主认知能力。

内在动机是指生物为了获得自身的内在满意,而非为了某种可分离的外在奖励所做出的行为<sup>[8-10]</sup>。Oudeyer等<sup>[11]</sup>提出智能自适应好奇心(intelligent adaptive curiosity)机制,使机器人在没有先验知识的情况下不断探索新颖场景,最大化自身学习进程;Baranes等<sup>[12]</sup>对IAC机制进行改进,提出一种鲁棒的智能自适应好奇心(R-IAC),并在复杂感知运动空间验证了该算法能够有效降低平均预测误差;Hester等<sup>[13]</sup>提出一种具有内在动机的环境搜索算法,利用模型预测的不确定性 & 环境新奇度作为探索动机,使机器人在没有外部奖励的情况下准确预测环境信息;张晓平等<sup>[14]</sup>将内在动机引入操作条件反射理论,并应用在机器人非特定环境中的路径规划问题,然而其所引入的好奇心函数仅由动作访问次数决定,不能充分反应好奇心在机器人环境认知中的作用。

为了进一步验证动机在移动机器人环境认知方面的可靠性,本文以学习自动机为数学模型,将外在动机中的取向性与内在动机中的好奇心相结合作为机器人的探索动机。其中,好奇心函数依文献[15]所提出的好奇心唤起理论设计与动作访问次数及访问时间间隔相关的函数,提高机器人的环境探索效率。同时,依据对环境已有的认知,重新设计路径规划时的动作选择机制,以避让原有路径上的新增障碍物。实验结果表明,本文方法较已有的自主学习模型在未知环境探索时具有更强的主动性,且搜索到的平均路径长度较短,算法稳定程度较高,具有一定环境自适应能力。

## 1 模型构建

本文提出一种结合内在动机理论的学习自动机CM-LA,该模型可用8元组进行描述,CM-LA =  $\langle S, A, O, M, N, C, \varepsilon, P \rangle$ ,具体内容如下。

1)  $S$ : 状态空间集合。  $S = \{s_i | i = 1, 2, \dots, n_s\}$ 。由于实际状态空间为连续环境,需要将其进行离散化,  $s_i \in S$  为环境中第  $i$  个状态空间,  $n_s$  表示环境的状态数目。

2)  $A$ : 动作集合。假定机器人在环境中的线速度保持不变,则在任意的状态空间中机器人只需要旋转一定的角度  $\Delta\theta$ 。动作集合表示为  $A = \{a_j | j = 1, 2, \dots, n_a\}$ 。其中:  $a_j$  为智能体的第  $j$  个离散动作,  $n_a$  为智能体总的动作数。

3)  $O$ : 取向性集合。取向性是由环境决定的生物

进化方向,即生物取向性<sup>[16]</sup>,是机器人环境认知中的外在动机。本文将此概念具体解释为:在移动机器人导航问题中,机器人在某一状态空间  $s_i$  时对其可以执行行为  $a_j$  的偏好程度。当机器人对一种动作的取向性增加时,相应对其他动作的取向性会降低,反之取向性则会增大。在自主学习过程中,不断变化的取向性集合即为机器人通过探索逐渐形成的认知,是对环境信息的表征。取向性集合可以表示为  $O = \{o_{ij} | i = 1, 2, \dots, n_s, j = 1, 2, \dots, n_a\}$ 。为了方便进行数学表示,假定机器人在某一状态下对所有动作的取向性之和为1,即满足  $\sum_{j=1}^{n_a} o_{ij} = 1$ , 且  $0 < o_{ij} < 1$ 。

4)  $M$ : 取向函数。取向函数用以引导移动机器人在环境中的学习方向,且机器人依据前后位置的取向函数变化大小对离散空间中的取向性进行更新。通常情况下,当机器人靠近目标点时,取向函数的值较高,而当机器人靠近障碍物时,取向性函数的值较低。鉴于此,本文对取向函数进行如下建模:

$$M(t) = m_1 d_g(t) - m_2 e^{m_3 d_0(t)}. \quad (1)$$

其中:  $d_g(t)$  为机器人在  $t$  时刻离目标点的距离,  $d_0(t)$  为  $t$  时刻机器人距离最近障碍物的距离,  $m_1$ 、 $m_2$ 、 $m_3$  分别为终点及障碍物的修饰权值。由此可以看出,取向函数的设计并不依赖于具体的环境信息,而是通过机器人与环境的不断交互而产生的,是具有普遍意义而非面向特定任务的。

5)  $N$ : 动作访问次数集合。  $N = \{n_{ij} | i = 1, 2, \dots, n_s, j = 1, 2, \dots, n_a\}$  表示智能体在状态  $s_i$  下对动作  $a_j$  的访问次数。当智能体在状态  $s_i$  下执行动作  $a_j$  后,则认为其完成了一次状态动作访问,同时  $n_{ij} = n_{ij} + 1$ 。

6)  $C$ : 好奇心函数。依据文献[15]提出的好奇心唤起理论设计好奇心函数。该理论认为:生物对某种动作的好奇心大小与相同刺激产生的频率及它们之间的时间间隔成反比。据此,本文将好奇心定义为与机器人在任意状态空间中的访问次数及在状态空间下执行某种动作的时间间隔相关的函数。对好奇心函数的建模表示如下:

$$c_{ij} = \frac{1}{2}(e^{-c_1 n_{ij}} + (1 - e^{-c_2 \Delta t})). \quad (2)$$

其中:  $n_{ij}$  为当前状态  $s_i$  下动作  $a_j$  的访问次数;  $\Delta t$  为上一次出现该刺激的时间间隔;  $c_1$  为访问次数的衰减率,满足  $0 < c_1 < 1$ ;  $c_2$  为时间间隔参数,满足  $0 < c_2 < 1$ 。

7)  $\varepsilon$ : 取向性学习方法。通常情况下,智能体会增加使自己处于较优状态的行为取向,而减小使自己处

于较差状态的行为取向,符合操作条件反射原理,因此可以用该理论对取向性学习方法进行设计.具体来说,智能体对自身状态的判断是由相邻两次取向函数值的变化 $\Delta M = M(t+1) - M(t)$ 决定的.当 $M(t+1) > M(t)$ 时,说明机器人在 $t$ 时刻的动作 $a_t$ 使得自己处于更优位置,则其取向性会相应增加;当 $M(t+1) < M(t)$ 时,说明机器人在 $t$ 时刻的动作 $a_t$ 使得自己处于更差位置,则其取向性会相应减少;当 $M(t+1) = M(t)$ 时,说明机器人在 $t$ 时刻的动作 $a_t$ 没有改变自己的状态,则取向性保持不变.设 $\eta$ 为取向性学习比例系数,且 $\eta > 0$ ,用来决定相邻两个位置之间取向函数值的变化量对智能体取向性变化的影响.

当 $M(t+1) > M(t)$ 时,有

$$\begin{cases} o_{ij}(t+1) = \frac{o_{ij}(t) \left( \frac{2}{1 + e^{-\eta\Delta M}} \right)}{1 + o_{ij}(t) \left( \frac{2}{1 + e^{-\eta\Delta M}} - 1 \right)}, \\ o_{ij'}(t+1) = \frac{o_{ij'}(t)}{1 + o_{ij}(t) \left( \frac{2}{1 + e^{-\eta\Delta M}} - 1 \right)}; \end{cases} \quad (3)$$

当 $M(t+1) < M(t)$ 时,有

$$\begin{cases} o_{ij}(t+1) = \frac{2o_{ij}(t) \left( 1 - \left( \frac{1}{1 + e^{\eta\Delta M}} \right) \right)}{1 - o_{ij}(t) \left( \frac{2}{1 + e^{\eta\Delta M}} - 1 \right)}, \\ o_{ij'}(t+1) = \frac{o_{ij'}(t)}{1 - o_{ij}(t) \left( \frac{2}{1 + e^{\eta\Delta M}} - 1 \right)}; \end{cases} \quad (4)$$

当 $M(t+1) = M(t)$ 时,有

$$\begin{cases} o_{ij}(t+1) = o_{ij}(t), \\ o_{ij'}(t+1) = o_{ij'}(t). \end{cases} \quad (5)$$

其中: $o_{ij}$ 为状态 $s_i$ 时使自身取向函数值增大的某个动作 $a_j$ 的取向值, $o_{ij'}$ 为状态 $s_i$ 时动作集合 $A$ 中除动作 $a_j$ 外其他动作的取向值.

8)  $P$ :动作概率集合.动作概率集合决定了机器人在环境中不同位置时最终选择的行为.在自主学习过程中,取向性及好奇心均会对其行为产生影响.而路径规划任务是自主学习结果的体现,可以依据机器人对环境的取向性及传感器信息完成.具体来说:当机器人移动至状态空间 $s_i$ 时,选取满足 $d_o(t) > d_{sh}$ 下取向性最大的动作.因此,动作概率可以写为

$$p_{ij} = \begin{cases} \frac{o_{ij} + c_{ij}}{\sum_{i=1}^{n_s} \sum_{j=1}^{n_a} (o_{ij} + c_{ij})}, & 1 \leq t \leq T_{sh}; \\ \max o_{ij} \text{ when } d_o(t) > d_{sh}, & \text{otherwise.} \end{cases} \quad (6)$$

其中: $T_{sh}$ 为学习轮数阈值, $d_{sh}$ 为机器人与障碍物间

的安全距离阈值.

为了进一步说明CM-LA的自主学习过程,下面给出算法流程,如图1所示.

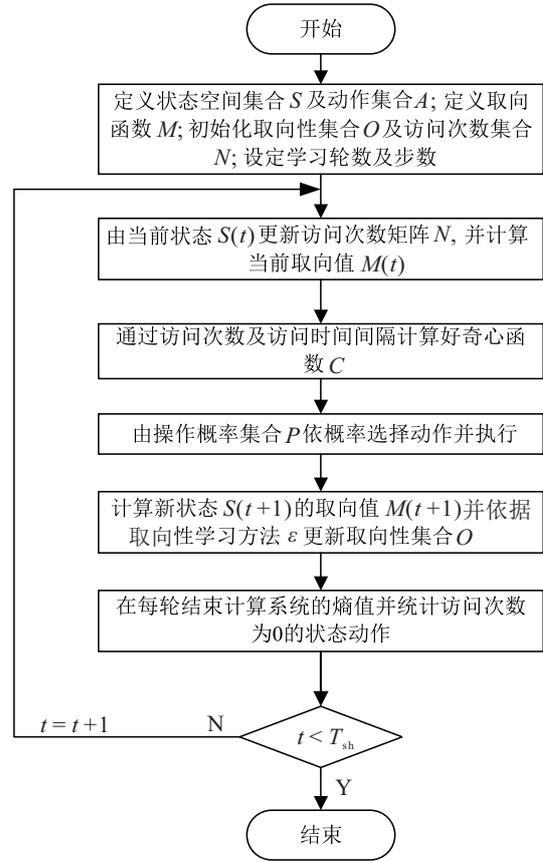


图1 CM-LA自主学习流程

## 2 仿真实验

### 2.1 实验对象及实验环境

实验对象为圆形移动机器人,其运动学模型为

$$\begin{cases} \theta_{t+1} = \theta_t + \Delta\theta, \\ x_{t+1} = x_t + v\Delta t \cos \theta_t, \\ y_{t+1} = y_t + v\Delta t \sin \theta_t. \end{cases} \quad (7)$$

其中: $\Delta\theta$ 为机器人在当前时刻需要转动的角度, $\Delta t$ 为采样时间间隔, $(\theta_t, x_t, y_t)$ 和 $(\theta_{t+1}, x_{t+1}, y_{t+1})$ 分别为 $t$ 时刻和 $t+1$ 时刻机器人朝向及位置.机器人半径 $r = 0.2\text{ m}$ ,底盘上有两个驱动轮 $W_L$ 和 $W_R$ ,通过两轮差速为自身提供转向,并假设机器人在环境中始终以 $v = 0.2\text{ m/s}$ 的速度向前移动.在机器人周围均匀分布6个声呐传感器,用来检测自身与环境中障碍物的距离,其有效测量距离为 $0.5\text{ m} \sim 5\text{ m}$ .图2展示了机器人靠近障碍物时的场景,其中 $d_0$ 为机器人距离最近障碍物的距离.

实验的仿真环境为 $5\text{ m} \times 5\text{ m}$ 的方形环境,如图3所示.其中,绿色圆圈代表机器人的起点,红色星形为终点,黑色形状为实验环境中的障碍物,模拟了移

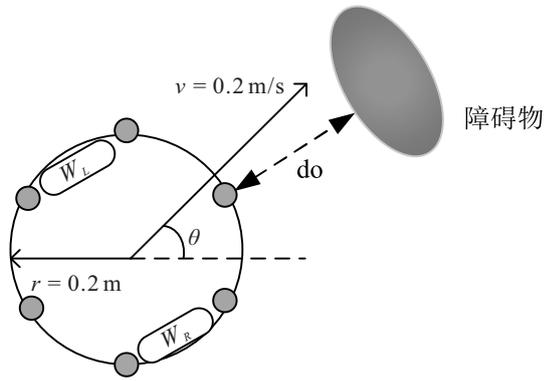


图2 机器人运动示意

动机器人导航问题中的密集障碍物场景. 假定机器人在任意时刻, 可执行的动作均为  $k(\pi/6)$ ,  $k \in [0, 12]$ ,  $k \in \mathbb{Z}$ , 则空间角度被划分成了离散且等距的12份, 即  $n_a = 12$ . 在设计好奇心函数时, 假设某种状态行为执行25次后, 将不再产生好奇心, 同时仅有4000步内的状态行为会对好奇心产生影响, 设定  $c_1 = 0.3$ ,  $c_2 = 0.002$ . 取向函数中:  $m_1 = -1$ ,  $m_2 = 10$ ,  $m_3 = -12$ . 为避免机器人与障碍物发生碰撞, 当  $d_o(t) > 2r$  时, 取向函数主要由当前位置与终点的距离影响, 当  $d_o(t) \leq 2r$  时, 取向函数主要由当前位置与最近障碍物距离影响. 在此参数下机器人的取向函数值如图4所示, 可以看出, 当机器人靠近终点时取向函数值较高, 而靠近障碍物时取向函数值较低. 为产生较好的环境认知效果, 经多次前置实验, 设定学

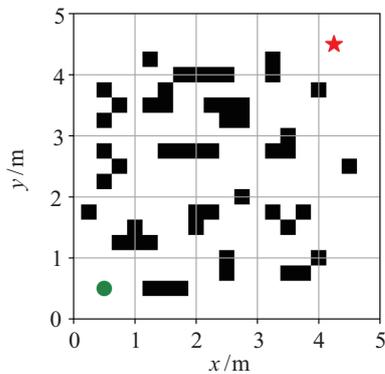


图3 实验环境

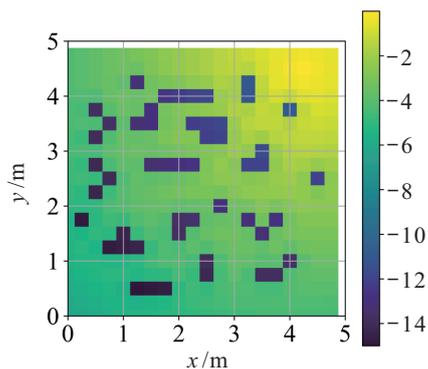


图4 实验环境取向值

习轮次阈值  $T_{sh} = 20$ , 取向性比例系数  $\eta = 1.6$ . 其他相关参数具体设置为:  $\Delta t = 1$  s,  $d_{sh} = 0.2$  m.

本文提出的CM-LA模型可使机器人在实验场景中不断学习环境信息, 最终对障碍物实现避让, 完成自主导航.

## 2.2 实验结果及分析

### 2.2.1 自主环境认知实验

首先令机器人在图3环境中进行探索, 实现对环境信息的自主认知. 实验共进行20轮, 每轮探索4000步. 在每轮探索结束后将机器人重置到起点位置. 图5分别为机器人在第1轮、第5轮、第15轮及第20轮自主认知的过程.

从机器人在环境的自主认知过程中可以看出, 在学习初期, 机器人对环境信息并不了解, 因此其探索范围倾向于整个状态空间. 而随着知识的不断积累, 机器人在任意状态空间的取向性趋于某一种动作, 其探索范围逐渐向目标点靠拢, 最终在第20轮学习时机器人较快地接近目标点附近, 实验结果展现出机器人对环境的认知过程是渐近的, 符合生物对环境的认知过程.

信息熵作为系统复杂程度的度量, 同时也是系统自组织能力的体现. 其值越低, 说明智能体对环境越加了解, 系统的自组织能力越强. 同时, 熵值下降速度可以反映出智能体对环境的认知程度. 为了进一步体现本文方法使得机器人在环境中的自主学习程度不断提升, 可以根据状态空间中取向性大小对熵值进行描述, 计算方法为

$$E(t) = - \sum_{i=1}^{n_s} \sum_{j=1}^{n_a} o_{ij}(t) \times \log_2 o_{ij}(t). \quad (8)$$

为验证本文提出的模型使机器人对环境具有较快的学习速率, 分别对比了文献[6]基于操作条件反射的方法(OCLM)及文献[14]基于内在动机机制的方法(IM-n). 3种方法均以学习自动机为数学模型, 并在自主学习过程中依据操作条件反射原理对任意状态空间的可执行动作概率进行更新. 不同之处在于: 在自主学习过程中, OCLM仅依据外在动机对可执行动作概率进行更新. IM-n引入了内在动机中的好奇心, 但仅考虑动作访问次数对其大小的影响, 并未加入访问时间间隔因素. 本文模型不仅同时考虑内在动机及外在动机对机器人环境认知的影响, 且将好奇心函数定义为与动作访问次数及访问时间间隔相关的函数, 能更全面、充分地反应好奇心在环境认知中的作用. 同时, 在进行路径规划任务中选择  $d_0(t) > d_{sh}$  下取向性最大的动作, 使得本模型较其他两种方

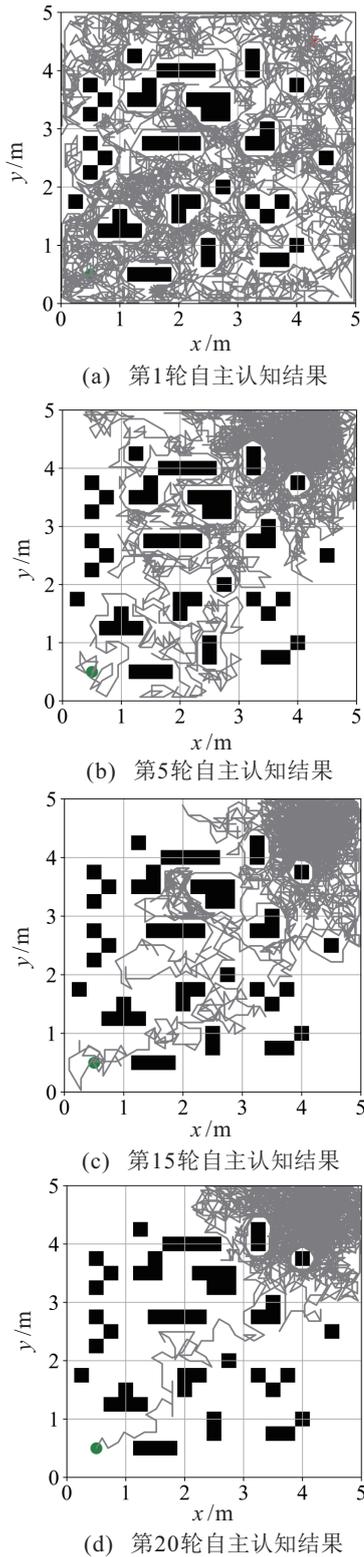


图5 自主环境认知结果

法对可变环境具有快速的自适应性。

相同实验环境及参数设置的情况下,分别对比3种方法的自主学习速率,实验结果如图6所示.初始时刻机器人对环境没有任何先验信息,三者熵值相同.随着自主学习过程的不断进行,熵值以不同速度下降.可以看出,本文提出的方法较另外两种方法,可以快速提升机器人对环境的自组织能力。

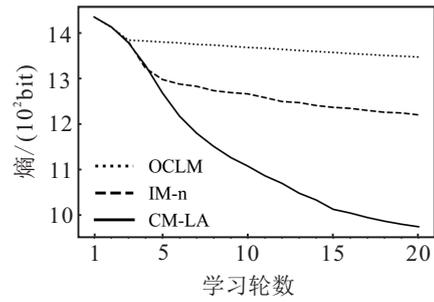


图6 熵值变化曲线

为验证本文方法在环境探索中具有更强的主动性,对状态空间中访问次数为0的行为个数进行统计,实验结果如图7所示.可以看出,初始阶段3种方法均未对环境进行探索,而随着探索的不断进行,本文方法快速地使访问次数为零的动作个数减小,说明机器人对环境中状态动作探索较为完全,对环境的认知程度较高。

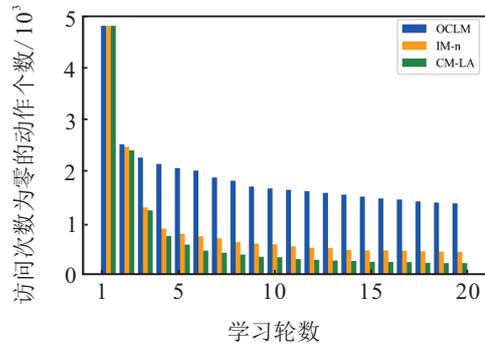


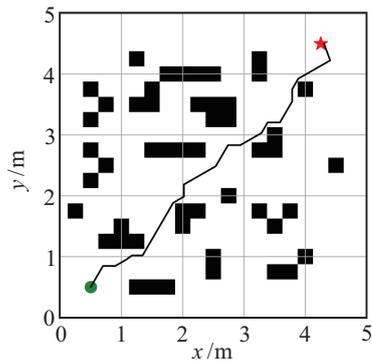
图7 未访问的状态动作统计结果

### 2.2.2 静态环境路径规划实验

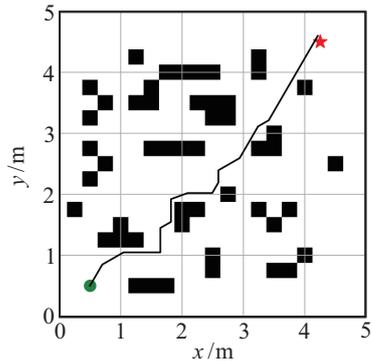
自主学习过程同时也是机器人环境认知的过程.在完成对环境的认知后,机器人便可在在此基础上完成路径规划.在图3环境下令机器人开始运动,当到达终点位置时,则认为路径规划任务结束.图8为利用本文方法,机器人自主规划的不同路径,分别用31步和32步。

### 2.2.3 新增障碍物路径规划实验

在机器人进行实际路径规划任务时,其规划好的路径上可能存在新的障碍物,需要机器人利用对环境的已有认知及传感器信息对新增障碍物进行实时避让.当机器人处于任意状态空间时,在动作集合中选择  $d_0(t) > d_{sh}$  下取向性最大的动作.本组实验在机器人已经规划好的路径上随机增添两个新障碍物,如图9中黄色方块所示.当机器人移动至新增障碍物附近时,选择新的动作以完成对障碍物的避让,如图9中红色圆圈处.可以看出,新动作并非盲目选取,而是依赖于自主学习过程中对状态空间较为完全的认知之上的.同时,新增障碍物的选取并非具有特定性,而

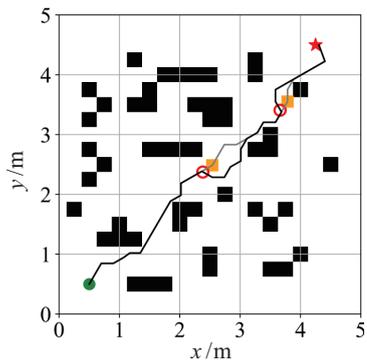


(a) 静态路径规划结果1

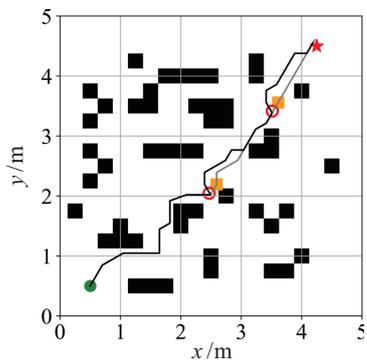


(b) 静态路径规划结果2

图8 静态环境路径规划任务



(a) 新增障碍物路径规划结果1



(b) 新增障碍物路径规划结果2

—原始路径 —当前路径  
 ■新增障碍物 ○重新选择动作处

图9 新增障碍物路径规划任务

是在其路径上任意选取,体现出机器人具有利用已有环境认知对新增障碍物实时避障的能力.图9分别为机器人自主学习并规划好路径后,两次成功避让新增障碍的过程,体现了机器人对环境具有较强的自适应能

力.

在对比CM-LA,IM-n及OCLM三种方法的路径规划效率时,首先在图3环境中让机器人自主学习,之后均在该环境中进行定点导航任务以检验学习过程中习得的环境信息.分别比较3种方法由起点到达终点所用路径长度的平均值、标准差、碰撞概率及是否能够实现实时障碍物避障.为使实验结果更加可靠,共进行100次实验,得到的实验结果均为进行100次后的平均值.

由表1及图10的对比结果可以看出,本文提出的方法较另两种方法搜寻到的平均路径较短且长度较为集中,偏离程度较小,体现出本文算法的高效性及稳定性.其路径规划效率较强,正是在自主学习过程中对状态空间中可执行动作的探索较为完全,环境认知能力较高.同时,在碰撞概率方面较另两种算法更低,能够根据已有认知对新增障碍物进行避让,体现出本方法对环境的自适应能力.

表1 3种方法路径规划效率对比

方法名称	平均步长	标准差	平均碰撞概率	是否能够实现实时避障
OCLM	34.57	1.98	0.76	否
IM-n	33.92	1.96	0.62	否
CM-LA	31.70	1.68	0.11	是

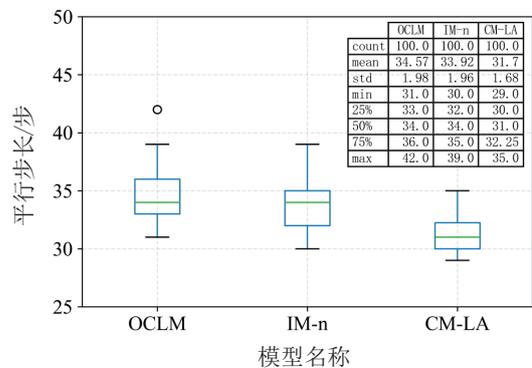


图10 重复实验路径规划结果箱线图

### 3 结论

本文提出了一种结合内在动机理论的移动机器人环境认知模型CM-LA.将内在动机与外在动机相结合,并将好奇心函数定义为与动作访问次数及访问时间间隔相关的函数,提高了机器人环境认知能力.自主探索是机器人进行环境认知的过程,路径规划则是其认知结果的体现.同时,设计了一种基于已有认知的自主避障方法,可以在路径规划时实时避让环境中存在的新障碍物,体现出该方法对环境具有较强的自适应能力.通过对比实验可以看出,本文方法可快速提升机器人对环境的认知程度,也表明了将内在动机与外在动机共同应用在机器人环境认知问题

中的可行性. 下一步的工作是结合心理学研究, 分析并讨论能够产生每种动机的几种因素之间的内在联系, 使智能体的行为更具仿生性.

### 参考文献(References)

- [1] 蔡自兴, 邹小兵. 移动机器人环境认知理论与技术的研究[J]. 机器人, 2004, 26(1): 87-91.  
(Cai Z X, Zou X B. Research on environment cognition theory and methodology for mobile robots[J]. Robot, 2004, 26(1): 87-91.)
- [2] Ryan R M, Deci E L. Intrinsic and extrinsic motivations: Classic definitions and new directions[J]. Contemporary Educational Psychology, 2000, 25(1): 54-67.
- [3] Soong L E, Pauline O, Chun C K. Solving the optimal path planning of a mobile robot using improved Q-learning[J]. Robotics & Autonomous Systems, 2019, 115: 143-161.
- [4] Duguleana M, Mogan G. Neural networks based reinforcement learning for mobile robots obstacle avoidance[J]. Expert Systems with Application, 2016, 62(15): 104-115.
- [5] Skinner B F. The behavior of organisms: An experimental analysis[M]. New York: D Appleton-Century Company, 1938, 186(1): 43-50.
- [6] 阮晓钢, 黄静, 范青武, 等. 一种基于操作条件反射原理的学习模型[J]. 控制与决策, 2014, 29(6): 1016-1020.  
(Ruan X G, Huang J, Fan Q W, et al. Artificial sensorimotor system with operant conditioning function[J]. Control and Decision, 2014, 29(6): 1016-1020.)
- [7] 黄静, 阮晓钢, 于乃功, 等. 具有操作条件反射机能的人工感觉运动系统[J]. 控制理论与应用, 2015, 32(5): 674-681.  
(Huang J, Ruan X G, Yu N G, et al. A learning model based on operant conditioning principles[J]. Control Theory & Applications, 2015, 32(5): 674-681.)
- [8] Baranes A, Oudeyer P Y. Active learning of inverse models with intrinsically motivated goal exploration in robots[J]. Robotics and Autonomous Systems, 2013, 61(1): 49-73.
- [9] Pierre-Yves O. What is intrinsic motivation? A typology of computational approaches[J]. Frontiers in Neurobotics, 2007, 1: 6.
- [10] Santucci V G, Baldassarre G, Mirolli M. Intrinsic motivation signals for driving the acquisition of multiple tasks: A simulated robotic study[C]. Proceedings of the 12th International Conference on Cognitive Modelling (ICCM). 2013: 1-6.
- [11] Oudeyer P Y, Kaplan F. Intelligent adaptive curiosity: A source of self-development[C]. Proceedings of the International Workshop on Epigenetic Robotics. Lund: Lund University Cognitive Studies, 2004: 127-130.
- [12] Baranes A, Oudeyer P Y. R-IAC: Robust intrinsically motivated exploration and active learning[J]. IEEE Transactions on Autonomous Mental Development, 2009, 1(3): 155-169.
- [13] Hester T, Stone P. Intrinsically motivated model learning for developing curious robots[J]. Artificial Intelligence, 2017, 247: 170-186.
- [14] 张晓平, 阮晓钢, 肖尧, 等. 基于内发动机机制的移动机器人自主路径规划方法[J]. 控制与决策, 2018, 33(9): 1605-1611.  
(Zhang X P, Ruan X G, Xiao Y, et al. Mobile robot autonomous path planning method based on intrinsic motivation mechanism[J]. Control and Decision, 2018, 33(9): 1605-1611.)
- [15] Berlyne D E. Curiosity and learning[J]. Motivation & Emotion, 1978, 2(2): 97-175.
- [16] 郜园园, 阮晓钢, 宋洪军. 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[J]. 控制与决策, 2013, 28(6): 930-934.  
(Gao Y Y, Ruan X G, Song H J. Operant conditioning learning automatic and its application on robot balance control[J]. Control and Decision, 2013, 28(6): 930-934.)

### 作者简介

阮晓钢(1958—), 男, 教授, 博士生导师, 从事人工智能与机器人等研究, E-mail: adrxg@bjut.edu.cn;

张家辉(1996—), 男, 硕士生, 从事认知机器人的研究, E-mail: zjh1004108554@126.com;

黄静(1979—), 女, 副教授, 从事认知机器人、机器学习及工业大数据应用等研究, E-mail: huangjing@bjut.edu.cn;

柴洁(1993—), 女, 博士生, 从事认知学习和认知导航的研究, E-mail: chajie@emails.bjut.edu.cn;

武悦(1995—), 男, 硕士生, 从事环境认知建模与类脑计算的研究, E-mail: wuy50271@gmail.com.

(责任编辑: 孙艺红)