

控制与决策

Control and Decision

一种具有情感和记忆机制的迷宫机器人认知模型

张晓平, 李凯, 王力, 闫佳庆, 何忠贺

引用本文:

张晓平, 李凯, 王力, 闫佳庆, 何忠贺. 一种具有情感和记忆机制的迷宫机器人认知模型[J]. *控制与决策*, 2023, 38(10): 2850–2858.

在线阅读 View online: <https://doi.org/10.13195/j.kzyjc.2022.0734>

您可能感兴趣的其他文章

Articles you may be interested in

[一种结合内在动机理论的移动机器人环境认知模型](#)

An environment cognition model combined with intrinsic motivation for mobile robots
控制与决策. 2021, 36(9): 2211–2217 <https://doi.org/10.13195/j.kzyjc.2019.1744>

[基于接触状态感知发育的机器人柔性装配方法](#)

Flexible assembly method based on contact state perception development
控制与决策. 2021, 36(4): 876–884 <https://doi.org/10.13195/j.kzyjc.2019.1079>

[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>

[基于生物启发神经网络和DMPC的多机器人协同搜索算法](#)

Multi-robot cooperative search algorithm based on bio-inspired neural network and DMPC
控制与决策. 2021, 36(11): 2699–2706 <https://doi.org/10.13195/j.kzyjc.2020.0959>

[机器人信息增益RRT环境探索算法](#)

Robot RRT based on information gain for environment exploration
控制与决策. 2021, 36(11): 2683–2689 <https://doi.org/10.13195/j.kzyjc.2020.1007>

一种具有情感和记忆机制的迷宫机器人认知模型

张晓平[†], 李 凯, 王 力, 闫佳庆, 何忠贺

(北方工业大学 电气与控制工程学院, 北京 100144)

摘要: 情感作为人类的高级认知, 在环境学习和环境理解方面具有重要意义. 将情感引入机器人搜索任务, 同时结合记忆机理, 提出一种具有情感和记忆机制的认知模型, 由内部状态、感受器、环境状态系统、情感系统、动态知识库、行为决策系统以及执行器 7 部分组成. 情感系统包含情感生成、情感状态以及情感记忆 3 个模块, 其中, 情感记忆用于提供内部奖励. 记忆功能在动态知识库中实现. 基于强化学习理论框架, 将情感内部奖励与记忆进行融合, 形成新的奖励机制, 并设计相关认知学习算法. 以需要“能量补给”的迷宫机器人搜索任务对所提出认知模型进行验证, 结果发现, 当面对不同情境时, 机器人会产生不同的情感. 结合前期记忆, 机器人所作决策更“拟人”, 表明情感和记忆机制设计的有效性. 将所提出认知模型、无情感决策认知模型、基于 ϵ -greedy 策略的 Q 学习算法进行对比, 结果表明, 情感和记忆的引入, 能够提高机器人的学习效率, 同时学习过程更稳定.

关键词: 情感; 记忆; 内部奖励; 迷宫机器人; 认知模型

中图分类号: TP242

文献标志码: A

DOI: 10.13195/j.kzyjc.2022.0734

引用格式: 张晓平, 李凯, 王力, 等. 一种具有情感和记忆机制的迷宫机器人认知模型[J]. 控制与决策, 2023, 38(10): 2850-2858.

A cognitive model of maze robot with emotion and memory mechanism

ZHANG Xiao-ping[†], LI Kai, WANG Li, YAN Jia-qing, HE Zhong-he

(School of Electrical and Control Engineering, North China University of Technology, Beijing 100144, China)

Abstract: As a higher level of human cognition, emotion plays an important role in environment learning and understanding. In this paper, emotion is introduced into the robot search task. By combined with memory mechanism, a cognitive model is proposed, which is composed of seven parts: internal state, receptor, environmental state system, emotion system, dynamic knowledge base, behavior decision system and actuator. In further, the emotion system consists of three modules: Emotion generation, emotion state and emotion memory, where the emotion memory is used to provide internal rewards. The memory function is implemented in the dynamic knowledge base. Based on the theoretical framework of reinforcement learning, the learning algorithm of the cognitive model is designed by integrating emotional internal reward and memory as a new reward mechanism. The maze robot search task requiring “energy replenishment” is tested, and results show that when faced with different situations, the robot can generate different emotions. Combined with the previous memory, the robot’s decision-making seems more “anthropomorphic”, which firstly proves the effectiveness of the design of the emotion and memory mechanism. The cognitive model, the cognitive model without emotion and the Q -learning algorithm based on the ϵ -greedy strategy are compared, and results show that with the mechanism of emotion and memory, the robot can learn faster, and at the same time, its learning process is more stable.

Keywords: emotion; memory; internal reward; maze robot; cognitive model

0 引言

众多研究表明情感与认知间存在紧密联系, 如认知科学家 Loew^[1] 指出认知感受是一种情感体验, 并通过在认知任务中观察效价 (valence) 和唤醒度

(arousal) 的生理与行为表征变化对此进行了证实; 脑科学家 Ahumada-Méndez 等^[2] 通过系统性分析脑电图 (EEG) 中情感状态和认知控制的神经标记物, 发现情感调节存在于大脑反应的认知控制中; 认知心理

收稿日期: 2022-05-01; 录用日期: 2022-07-17.

基金项目: 国家自然科学基金项目 (61903006); 北京市自然科学基金项目 (4202022, 4204096, 4212035); 北京市长城学者项目 (CIT&TCD 20190304); 北京市教委科研项目 (KM202010009006, 201910009008).

责任编辑: 冯俊娥.

[†]通讯作者. E-mail: zhangxiaoping369@163.com.

学家Culot等^[3]通过开展情感-认知相关实验,发现情感状态影响大脑中认知控制的投入以及相关的元认知经验。随着神经科学和核医学影像技术的发展,不少科学家利用电生理技术和核磁共振成像(fMRI)研究行为过程中认知和情感在大脑中的表现^[4],揭示出二者所涉及的核心脑区单元紧密相关,如杏仁核、前扣带皮质、海马体、前额叶等^[5-8]。

当今,情感对人类学习的重要影响^[9-11]已被证实,时刻影响着人们的认知、决策、交流等。因此,将情感对人类的有益影响迁移至机器人系统,研究机器人基于情感机制的智能是认知机器人实现的重要途径之一。

2000年,MIT的Picard^[12]提出情感计算,开启了机器人情感研究的热潮。至今,机器人情感方面的研究主要集中于以下两方面:一方面是进行情感的识别和表达,如研究人脸表情、语音文字、身体姿态的识别和表达。另一方面则是进行具有情感机制的认知机器人研究,该方面研究目的是将机器人的情感、认知与自主系统整合,令机器人能够产生和调节情感,并作出基于情感的决策。如2003年,Gadanh^[13]结合强化学习提出一种ALEC(asynchronous learning by emotion and cognition)情感认知架构,该架构中的认知系统能够生成从环境学习的规则,与EBII(emotion-based II)模型方法在移动机器人生存实验下比较,采用该认知架构机器人的碰撞率减少了40%,活动距离增加20%。2011年,Sequeira等^[14]受情感评价理论的启发,提出了一种IMRL(intrinsically motivated reinforcement learning)框架,将情感评价理论中的4个评价维度(新颖性、动机、支配和效价)转化为数字特征,作为强化学习的奖励,提高了智能体在4组觅食实验(探索、平衡、猎物季节和不同奖励场景)中的环境适应性。2015年,李宏光等^[15]将性格特征、心境状态以及情感状态进行融合,构建了多层情感认知模型,并在此基础上提出一种多目标情感认知决策方法,实现了智能体集感性 with 理性于一身。针对多机器人协调任务分配问题,2021年,Zhong等^[16]提出了一种基于意愿可解释和计算的情感模型以及度量机器人差异的个性模型CASE,通过情感传染模型计算机器人间的情感交互作用,利用意愿度量情感因素对任务分配的影响,提高了多机器人在追捕任务的性能。

基于情感和认知生理学、心理学、脑科学等学科基础,本文以迷宫机器人为对象,引入情感和认知,同时结合记忆机理,提出一种具有情感和记忆机制的认

知模型。首先,设计情感系统中的情感生成系统,为迷宫机器人建立人工情感模型;然后,基于强化学习理论框架,将情感系统中产生的内部奖励以及动态知识库产生的记忆融合为新的奖励机制,提出认知模型学习算法;最后,通过需要“能量补给”的迷宫搜索任务验证所提出认知模型,在该认知模型指导下,机器人具有更自主、高效、稳定的搜索过程。

1 具有情感和记忆机制的认知模型

1.1 认知模型结构

脑科学、认知神经科学、认知心理学等不同学科对大脑结构及其情感和认知原理展开了相关研究,结果表明,大脑中情感和认知功能区域主要集中于丘脑、边缘系统以及大脑皮层。丘脑作为感觉传导中枢,主要负责视觉、听觉和嗅觉等外部感觉信息以及身体内部感觉信息的传导^[17];边缘系统作为情感中心,主要包括海马体、杏仁核、扣带回^[18]等,海马体负责情感的记忆和学习,杏仁核负责情感产生、调节以及识别^[19],而扣带回前下部与情感的加工有关,后上部与认知功能有关。这些结构与下丘脑以及植物神经系统联系,参与调节本能反应和情感行为;大脑皮层主要参与理解事件、决策目标、管理行为时序等人脑活动^[20]。

基于上述交叉学科研究基础,提出一种认知模型,包含感受器、内部状态、环境状态系统、情感系统、行为决策系统、动态知识库以及执行器7部分,模型结构如图1所示。

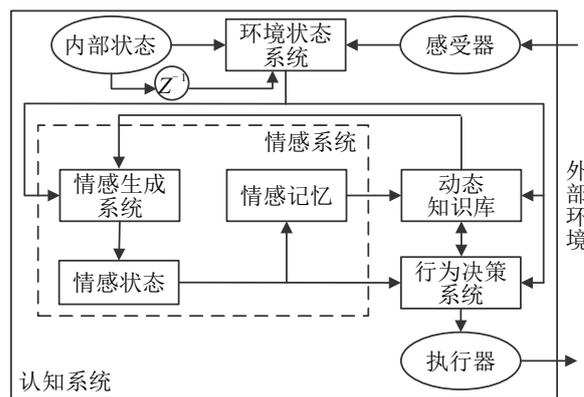


图 1 认知模型结构

模型各部分含义具体如下。

1) 感受器: 感受外部环境信息,表示为3元组: $PER_ORG = \langle S, A, Ga \rangle$ 。其中: $S = \{S_i | i = 1, 2, \dots, n_s\}$ 为可感受的离散状态集合; $A = \{A_i | i = 1, 2, \dots, n_s\}$ 为离散状态所对应的可选动作子集集合; $Ga = \{Ga_i | i = 1, 2, \dots, n_s\}$ 为离散状态所对应最大内部能量状态补给集合, n_s 为可感知的离散状态个数。

2) 内部状态: 机器人身体内部状态信息, 如机器人内部能量、机器人耐久度. 针对需要“能量补给”的迷宫搜索任务, 本文选取的机器人内部状态为内部能量状态 $P(t)$, $P = \{P(t)|t = 0, 1, \dots, n_t\}$ 为机器人身体内部能量状态集合, n_t 为机器人任务生存时间, $t = 0$ 为机器人处于任务开始时刻; $t = n_t$ 为机器人身体内部能量状态为0或完成任务时刻.

3) 环境状态系统: 机器人外部环境信息和身体内部状态中枢站, 表示为 $(\text{PER_ORG}, P, G)$, 其中 $G = \{G(t)|t = 0, 1, \dots, n_t\}$ 为机器人从离散状态获得的内部能量状态增益集合, 内部能量状态增益定义如下:

$$G(t) = \begin{cases} P(t) - P(t-1), & \text{Ga}(t) \neq 0; \\ G(t-1), & \text{Ga}(t) = 0. \end{cases} \quad (1)$$

其中 $\text{Ga}(t)$ 为 t 时刻所处离散状态所对应最大内部状态补给 Ga_i .

4) 情感系统: 机器人情感状态和情感记忆产生中心, 表示为3元组: $\text{EMO_SYS} = \langle E, R_{\text{emo}}, R_{\text{mem}} \rangle$. 其中: $E = \{E(t)|t = 0, 1, \dots, n_t\}$ 为情感生成系统产生的情感状态集合; $R_{\text{emo}} = \{R_{\text{emo}}(t)|t = 0, 1, \dots, n_t\}$ 为情感记忆产生的逆向情感奖励集合; $R_{\text{mem}} = \{R_{\text{mem}}(T)|T = 1, 2, \dots, n_T\}$ 为情感记忆产生的周期情感奖励集合, n_T 为机器人任务生存时间内部能量状态补给周期, $T = 1$ 为进行第1次内部能量状态补给, $T = n_T$ 表示机器人内部能量状态为0前或完成任务最大能量补给周期.

5) 行为决策系统: 根据环境状态系统和情感系统输出, 结合动态知识库, 实现机器人行为决策, 表示为2元组: $\langle \pi, a \rangle$. 其中: $\pi = \{\pi_j|j = 1, 2, \dots, n_j\}$ 为行为决策集合, n_j 为迷宫机器人的行为个数; $a = \{a_m|m = 1, 2, \dots, n_m\}$ 为动作选择集合, n_m 为迷宫机器人的动作个数. 针对需要“能量补给”的迷宫搜索任务, 行为分为搜索和能量补给, 动作为迷宫节点处 $\{\text{north}, \text{south}, \text{east}, \text{west}\}$ 的方向选择.

6) 动态知识库: 机器人与迷宫环境交互所学知识, 知识元素表示为5元组: $\text{DYN_KNO} = \langle A', L, D, \text{STA_ACT}, \text{STA_PWO} \rangle$. 其中: $A' = \{A'_i|i = 1, 2, \dots, n_s\}$ 为离散状态对应能量补给行为最佳动作集合; $L = \{L(T)|T = 1, 2, \dots, n_T\}$ 为路径搜索状态集合; $D = \{D(t)|t = 0, 1, \dots, n_t\}$ 为环境搜索状态集合; $\text{STA_ACT} = \langle (Y, U), (Y', U') \rangle = \{(Y_k, U_k), (Y'_c, U'_c)|k = 1, 2, \dots, n_k, c = 1, 2, \dots, n_c\}$ 为状态-动作记忆集合, (Y, U) 逐次记录周期内所遇状态和动作选择序列, n_k 为机器人周期内所遇离散状态总次数, (Y', U') 记录周期内最后一遍所遇状态和该状态的动作选择

序列, 原理如文献[21], n_c 为机器人周期内所遇离散状态个数; $\text{STA_PWO} = \{(Y'_z, B_z)|z = 1, 2, \dots, n_z\}$ 为状态-能量记忆集合, 记录周期内所遇离散状态以及对应返回能量补给点所需内部能量状态 B_z , n_z 为搜索周期内所遇离散状态个数, STA_ACT 和 STA_POW 体现了模型的记忆功能.

7) 执行器: 机器人动作执行机构, 本文选取迷宫机器人为两轮机器人, 动作执行机构表示为2元组: $\langle V_1, V_2 \rangle$. 其中: $V_1 = \{V_{1m}|m = 1, 2, \dots, n_m\}$ 为左轮转矩集合, $V_2 = \{V_{2m}|m = 1, 2, \dots, n_m\}$ 为右轮转矩集合.

1.2 情感生成系统

研究具有情感机制的机器人, 首先, 要生成人工情感, 需要对情感进行建模. 目前, 关于情感建模的研究存在许多理论和方法. 就情感模型而言, 根据其描述方式可将其大致分为离散情感模型、维度情感模型^[22]以及其他情感模型. 离散情感模型: 如Ekman提出的6种基本情感, 愤怒、厌恶、恐惧、高兴、悲伤和惊讶^[23]; 维度情感模型: 如一维情感模型、VA二维情感模型、PAD情感模型^[24]等; 其他情感模型: 如基于认知评价理论的OCC情感模型^[25]、基于马尔可夫的HMM情感模型^[12]、基于个性的多层情感模型^[26]等. 本文基于一维情感模型理论, 为情感生成系统设计了一种情感模型, 可产生高兴、恐惧、愤怒3种情感, 形式如下:

$$E(t) = \frac{G(t)}{k_1} \left[\arctan(P(t) - k_2) + \frac{P(t) - B(t)}{k_3} \right] \cdot e^{-D(t)}, \quad (2)$$

$$E(t) = G(t) \cdot e^{\frac{2P(t) - B(t) - k_4}{k_5} - D(t)} - k_6. \quad (3)$$

其中: $B(t)$ 为 t 时刻所处离散状态对应返回能量补给点所需内部能量状态值 B_z , 情感强度 $|E(t)|$ 与机器人获得的内部状态增益 $G(t)$ 成正相关. 当 $G(t) > 0$ 时, 式(2)产生高兴或恐惧两种情感, $E(t) > 0$ 时为积极的高兴情感, $k_7 < E(t) < 0$ 时为恐惧情感; 当 $G(t) < 0$ 时, 式(3)产生的 $E(t) < k_7$ 为愤怒情感. $k_1 \sim k_7$ 为情感模型参数.

1.3 奖励机制

在日常生活中, 奖励通常是稀疏的, 执行者在一项任务过程中, 往往需要经历一系列的尝试, 直至完成任务才会获得奖励, 过程中则无奖励体现. 但是人类大脑拥有反思机制, 能够建立时序情景与目标事物的相关联系, 基于记忆机理获得奖励. 因此, 本文在考虑情感认知的同时, 结合记忆认知, 基于强化学习理

论框架,将情感内部奖励与记忆融合,提出一种奖励机制,该奖励机制由环境奖励、逆向情感奖励和周期情感奖励组成.其中,逆向情感奖励和周期情感奖励属于内部奖励.

图2为迷宫机器人一次搜索奖励获得过程(示意图).机器人搜索过程消耗内部能量,根据节点可选动作子集 $A_i \in \{\text{north, south, east, west}\}$ 进行动作 a_m 选择,假设节点1为能量补给点,节点X为未搜索点,如图2所示,机器人搜索轨迹为: $1 \rightarrow 2 \rightarrow 3 \rightarrow 4 \rightarrow 5 \rightarrow 6$,若到达节点6时,机器人必须返回节点1补充内部能量状态以维持下次搜索,返回轨迹为: $6 \rightarrow 5 \rightarrow 4 \rightarrow 3 \rightarrow 2 \rightarrow 1$,机器人从能量补给点出发搜索完再返回至能量补给点.此时,迷宫机器人完成一个周期的搜索.此过程中的奖励机制产生奖励 $R = \langle R_{\text{env}}, R_{\text{emo}}, R_{\text{mem}} \rangle$,各奖励如表1所示.

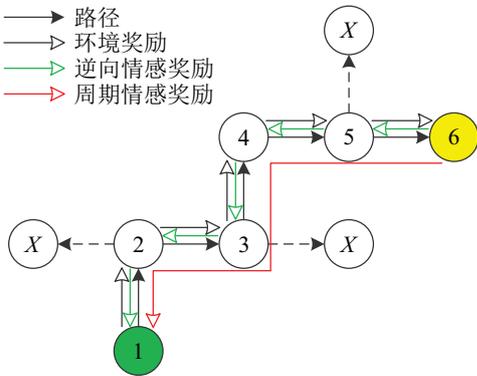


图 2 奖励机制示意图

表 1 奖励集合

奖励类型	奖励集合
R_{env}	$R_{\text{env}1 \rightarrow 2}, R_{\text{env}2 \rightarrow 3}, R_{\text{env}3 \rightarrow 4}, R_{\text{env}4 \rightarrow 5}, R_{\text{env}5 \rightarrow 6}$ $R_{\text{env}6 \rightarrow 5}, R_{\text{env}5 \rightarrow 4}, R_{\text{env}4 \rightarrow 3}, R_{\text{env}3 \rightarrow 2}, R_{\text{env}2 \rightarrow 1}$
R_{emo}	$R_{\text{emo}01 \leftarrow 2}, R_{\text{emo}02 \leftarrow 3}, R_{\text{emo}03 \leftarrow 4}, R_{\text{emo}04 \leftarrow 5}, R_{\text{emo}05 \leftarrow 6}$
R_{mem}	$R_{\text{mem}6 \rightarrow 5 \rightarrow 4 \rightarrow 3 \rightarrow 2 \rightarrow 1}$

表1中: $1 \rightarrow 2$ 过程至节点2时, $R_{\text{env}1 \rightarrow 2}$ 为节点1获得向节点2动作的外部环境奖励; $R_{\text{emo}01 \leftarrow 2}$ 为节点2获得向节点1动作的奖励,即逆向情感奖励. $6 \rightarrow 5 \rightarrow 4 \rightarrow 3 \rightarrow 2 \rightarrow 1$ 过程至节点1时, $R_{\text{mem}6 \rightarrow 5 \rightarrow 4 \rightarrow 3 \rightarrow 2 \rightarrow 1}$ 为节点6、节点5、节点4、节点3、节点2逐级获得趋向节点1动作的奖励,即周期情感奖励.

环境奖励按下式设置:

$$R_{\text{env}}(t) = \begin{cases} 100, & \text{能量补给点;} \\ 0, & \text{普通节点;} \\ -5, & \text{死路节点;} \\ -100, & \text{陷阱点.} \end{cases} \quad (4)$$

其中:能量补给点为机器人在迷宫中补充内部能量的节点;死路节点为迷宫中仅能选择“返回”动作节

点;陷阱点为机器人在此节点额外损失内部能量状态节点;普通节点则是内部状态增益 $Ga(t) = 0$,且非死路节点的节点.

按下式更新Q值:

$$Q(s, a) = (1 - \alpha)Q(s, a) + \alpha[R_{\text{env}} + \max Q(s', a')]. \quad (5)$$

其中: s 为当前状态, a 为当前状态选择的动作, α 为学习率, $\max Q(s', a')$ 为当前状态选择动作后下一状态最大收益.

逆向情感奖励按下式设置:

$$R_{\text{emo}}(t) = \begin{cases} E(t), & E(t) > 0; \\ \frac{1}{|E(t)|}, & E(t) < 0. \end{cases} \quad (6)$$

按下式更新Q值:

$$Q(s, a'') = R_{\text{emo}}, \quad (7)$$

其中 a'' 为进入该节点状态时的逆方向.

周期情感奖励按下式设置:

$$R_{\text{mem}}(T) = \begin{cases} E(T), & T > 0; \\ 0, & T = 0. \end{cases} \quad (8)$$

其中 $E(T)$ 为第 T 次周期完成时刻产生的情感状态.

按下式更新Q值,此奖励仅当周期完成时获得:

$$Q(s, a) = (1 - \alpha)Q(s, a) + \alpha[R_{\text{mem}} + \max Q(s', a')]. \quad (9)$$

2 认知模型学习算法

动物和人在未知环境搜索时,通过抵达目标获得“奖励”,进而掌握相关“知识”.对于迷宫机器人搜索任务,若考虑“能量补给”,即要求机器人探索过程中若发生能量低于一定值时,相比于继续探索,应以能量补给为首要任务,这更符合生物环境探索实际情境.此时,相比于路径自主规划^[27],要求迷宫机器人应具有更高层级的智能,对此,本文在迷宫机器人认知模型和算法设计过程中,引入情感和记忆,基于新的奖励机制,提出认知模型的迷宫机器人自主搜索算法.

2.1 认知模型学习步骤

需要“能量补给”的迷宫机器人,其搜索任务分为2个阶段:第1阶段为寻找提供内部能量的能量补给点,第2阶段为能量补给点信息获取后的搜索,学习步骤如图3所示.

step 1: 开始第1阶段任务,数据初始化:初始化内部状态和动态知识库.

step 2: 根据状态选择动作,将“状态-动作”写入 STA_ACT;更新至下一状态;由式(4)获得 $R_{\text{env}}(t)$,由

式(5)更新 Q 值.

step 3: 判断是否找到能量补给点,若找到则执行 step 4;判断是否满足 $P(t) > 0$,若满足则转向 step 2,否则结束.

step 4: 开始第2阶段任务:根据式(2)和(3)更新 $E(T)$;由式(8)获得 $R_{mem}(T)$,由式(9)和 (Y', U') 更新 Q 值;由记忆模块2更新 $L(T)$.

step 5: 判断是否满足 $D(t) = 1$,若满足则结束.

step 6: 将 STA_ACT 和 STA_PWO 重置清空.

step 7: 根据式(2)和(3)更新 $E(t)$;由记忆模块1更新 $B(t)$;判断状态是否在 STA_PWO ,若不在则转向 step 9.

step 8: 由式(6)获得 $R_{emo}(t)$,由式(7)更新 Q 值.

step 9: 根据状态选择动作,将“状态-动作”写入

STA_ACT ;更新至下一状态;由式(4)获得 $R_{env}(t)$,由式(5)更新 Q 值.

step 10: 判断是否继续搜索,若满足则转向 step 7,判断条件如下式所示:

$$\pi(t) : \text{搜索} \begin{cases} E(t) > 0; \\ k_7 < E(t) < 0, P(t) - B(t) > k_8; \\ E(t) < k_7, P(t) - B(t) = 0. \end{cases} \quad (10)$$

step 11: 根据状态选择最大价值动作,将“状态-动作”写入 STA_ACT ;更新至下一状态;由式(4)获得 $R_{env}(t)$,由式(5)更新 Q 值.

step 12: 判断是否回到能量补给点,若是则转向 step 4,否则转向 step 11.

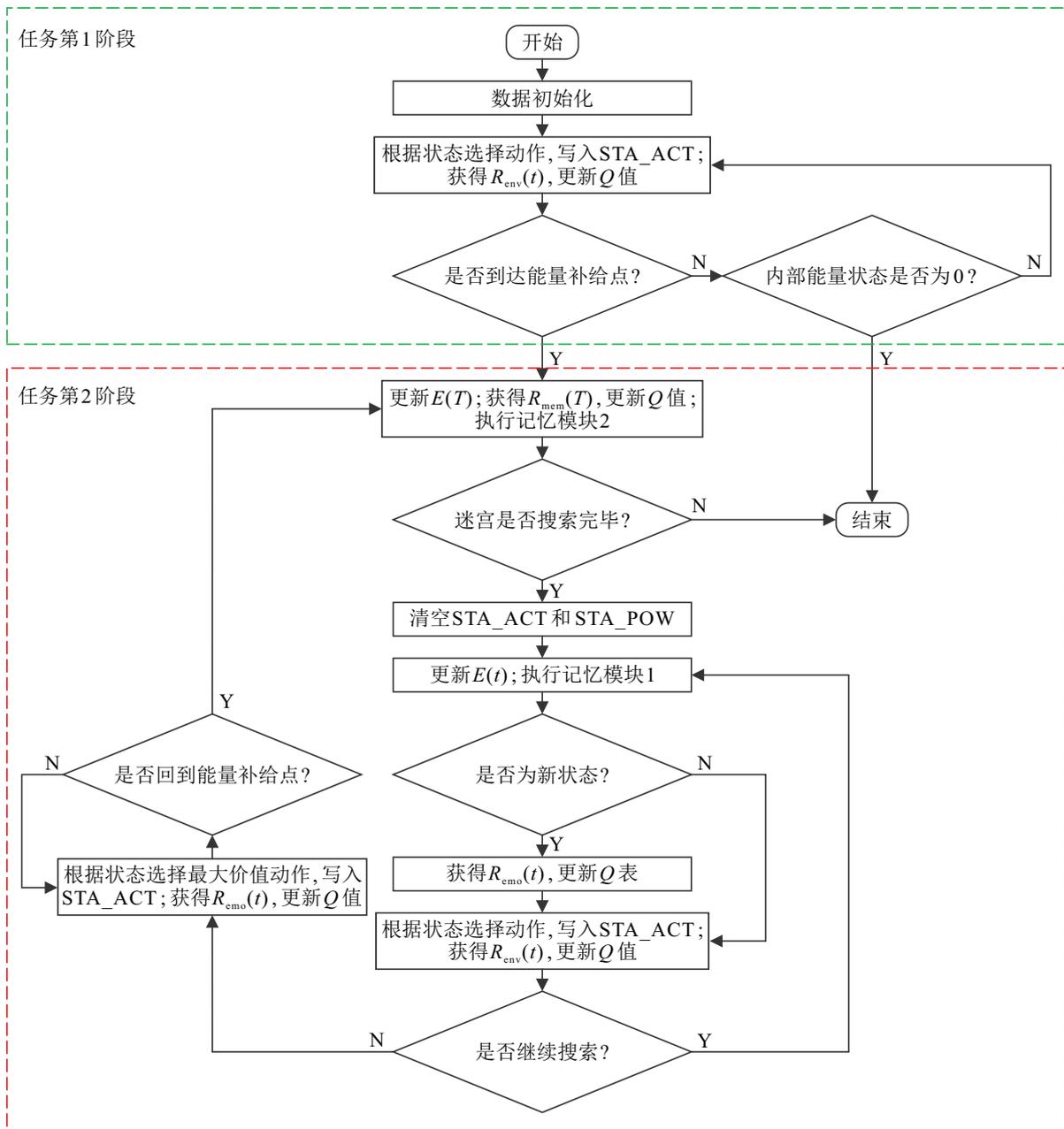


图3 认知模型学习步骤

2.2 记忆模块 1

通过反思记忆的方式,反馈机器人返回能量补给点所需内部能量状态值 $B(t)$,记忆模块 1 步骤如下.

step 1: 判断当前状态是否在 STA_PWO 中, 若不在则转向 step 3.

step 2: 调取 STA_PWO 中当前状态对应的 B_z , 令 $B(t) = B_z$.

step 3: 令 $B(t) = B(t - 1) + 1$, 将“状态-能量”写入 STA_PWO.

2.3 记忆模块 2

通过反思记忆的方式,分析已搜索路径和未搜索路径情况得到 $L(T)$, 记忆模块 2 步骤如下.

step 1: 判断 (Y, U) 的第 k 个状态是否在前 $k - 1$ 个状态中, 若在则转向 step 3.

step 2: 判断第 k 个状态的动子集 A_i 是否搜索完毕, 若未搜索完毕则将 (Y, U) 中第 $k - 1$ 个状态对应动作选择标记重置.

step 3: 判断是否满足 $k > 1$, 若满足则令 $k = k - 1$, 并转向 step 1, 否则结束.

3 实验结果与分析

3.1 实验参数设置

本实验平台为 Intel Core i7-4710MQ CPU@2.5 GHz, 16 GB 内存, 程序运行环境为 Python 3.6, 仿真运行环境为 V-REP. 迷宫机器人情感模型参数设置为 $k_1 = 5$ 、 $k_2 = 10$ 、 $k_3 = 10$ 、 $k_4 = 10$ 、 $k_5 = 2.5$ 、 $k_6 = 2$ 、 $k_7 = -2$, 搜索行为判断参数设置为 $k_8 = 4$. 需要“能量补给”的迷宫机器人搜索任务具体设置为: 机器人初

始能量值为 10, 由第 1 阶段开始执行任务, 过程中, 每经过 1 个节点进行动作选择, 会消耗 1 点能量, 每次到达“能量补给点”能量将补充至 19, 若遇到陷阱点, 则机器人额外损失 2 点能量. 首次搜索到能量补给点, 开始第 2 阶段任务, 在此过程中, 机器人不断补充能量进行搜索, 直至迷宫搜索完毕. 在 V-REP 中搭建路径宽度为 45 cm 的迷宫环境, 并使用其自集成的 dr 20 移动机器人. 迷宫环境如图 4 所示, 其中: 共设有 21 个节点, 机器人起始点设置于节点 3 与节点 4 间, 能量补给点为节点 21, 陷阱点为节点 5, 其他节点影响机器人的内部能量变化以及环境奖励如表 2 所示.

3.2 迷宫自主搜索过程

在具有情感和记忆机制的认知模型以及算法指导下, 迷宫机器人一次完整搜索过程如图 5 所示. 搜索过程中内部能量状态 $P(t)$ 以及情感状态 $E(t)$ 变化曲线如图 6 所示. 其中: 当 $E(t) < -2$ 时, 机器人处于愤怒状态; 当 $-2 < E(t) < 0$ 时, 机器人处于恐惧状态; 当 $E(t) > 0$ 时, 机器人处于高兴状态. 当 $t = 0$ s 时, 机器人出发开始第 1 阶段任务, 由于机器人对迷宫环境信息未知, 不具任何情感, 初次遇到各节点时随机选择动作; 当 $t = 53$ s 时, 机器人搜索到能量补给节点 21, 内部能量状态 $P(t)$ 补充至 19, 同时如图 6 所示, 产生高兴情感 $E(T)$. 能量补充后, 开始第 2 阶段任务, 搜索过程中, 随着内部能量状态 $P(t)$ 减少以及返回能量补给点“距离” $B(t)$ 增加, 机器人高兴情感强度 $|E(T)|$ 逐渐衰减; 当 $t = 128$ s 时, 机器人遇到陷阱点, 如图 6 所示, 额外损失 2 点能量, 情感由高兴转变为愤怒. 而后机器人的行为也会更倾向于“冒险”, 即内部能量状态 $P(t)$ 仅够返回能量补给点才进行“能量补给”决策; 当 $t = 275$ s 时, 如图 6 所示, 机器人内部能量降为 0 时, 刚好返回能量补给点补充能量, 情感由愤怒恢复为高兴, 同时机器人进行“反思”分析路径搜索状况 $L(T)$, 并出发再次搜索; 当 $t = 324$ s 时, 机器人对未搜索完路径进行搜索; 当 $t = 435$ s 时, 由机器人通过记忆机制“回想”返回能量补给点所需内部能量状态值 $B(t)$ 与体内剩余内部能量状态 $P(t)$ 接近, 机器人情感状态由高兴变为恐惧; 当 $t = 479$ s 时, 机器人恐惧情感强度增大, 如图 5(g) 所示, 机器人提前做出“能量补给”决策. 如图 6 所示, 在 $t = 523$ s 时, 内部能量状态剩余 2 点时, 返回至能量补给点. 直至 $t = 757$ s 时, 机器人完成迷宫搜索任务并返回能量补给点.

上述迷宫搜索过程, 机器人能够产生情感和记忆, 并自主地作出提前返回补充能量决策、分析路径的搜索情况, 验证了所提出认知模型和算法的有效性.

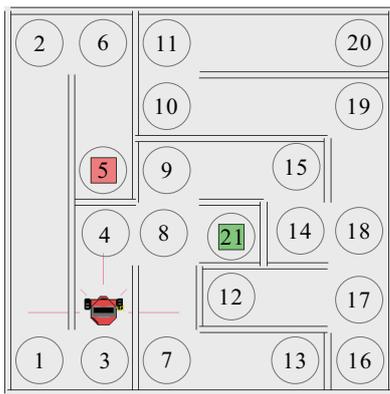


图 4 迷宫环境

表 2 节点信息

节点类型	节点序号	内部能量状态变化	环境奖励
普通节点	1, 2, 3, 4, 6, 7, 8, 9, 10, 11, 14, 15, 17, 18, 19	-1	0
死路节点	12, 13, 16, 20	-1	-5
陷阱点	5	-3	-100
能量补给点	21	→ 19	100

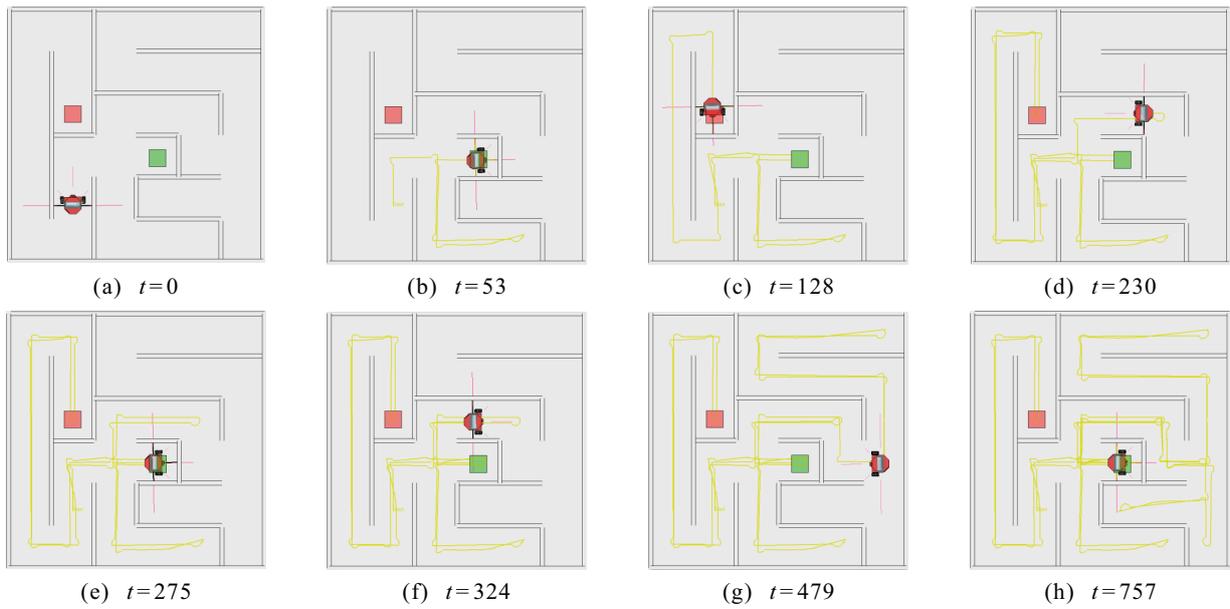


图5 搜索过程

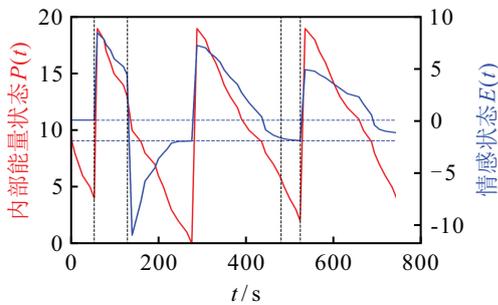


图6 内部能量状态、情感状态变化

搜索过程中搜索率变化曲线如图7所示。0~53s期间,搜索率随着机器人开始搜索逐渐上升;当 $t = 53s$ 时,机器人处于能量补充状态,会消耗一定时间,此阶段搜索率无变化;随后重新出发继续进行搜索,可见搜索率增大。由图7可见,迷宫机器人搜索率曲线在部分时间段内保持不变,除补充能量耗时因素外,以275~324s为例,是机器人此期间所走路线属于已搜索区域,因此搜索率无变化。

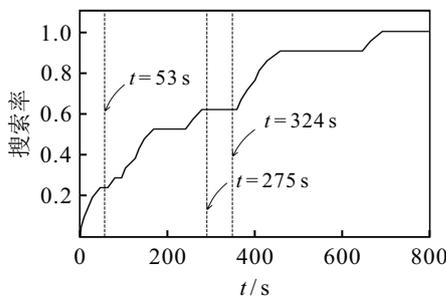


图7 搜索率曲线

3.3 迷宫搜索对比

将所提出认知模型、无情感决策认知模型以及基于 ϵ -greedy策略的Q学习算法,在相同迷宫环境中进行比较,实验结果如下。

3种学习算法参数调优后搜索率随时间变化曲线如图8所示。其中:绿色折线对应所提出认知模型;蓝色折线对应无情感决策认知模型,是所提出认知模型去除情感决策但拥有记忆机制的模型,主要用于分析情感决策对认知模型的影响,其行为决策机制为:当机器人内部能量仅够返回能量补给点时($P(t) = B(t)$),进行能量补给决策,其余涉及到的相关参数与所提出认知模型一致;红色折线对应基于 ϵ -greedy策略的Q学习算法,贪婪值设置为0.3。学习初期0~125s,初期阶段迷宫区域均处于未搜索状态,此阶段机器人记忆和情感对搜索影响小,因此,3种方法对应机器人迷宫搜索范围变化无差距。125s后的学习中后期,迷宫部分区域已搜索,所提出认知模型和无情感决策认知模型均具有记忆机制,能够区分迷宫区域搜索情况,因此,搜索效率超越基于 ϵ -greedy策略的Q学习算法。而在227~346s期间,机器人内部能量接近返回“能量补给点”所需最小能量值,在本文具有情感机制的认知模型下,机器人更加“拟人化”和“个性化”,表现出“恐惧”反应,即作出“能量补给”

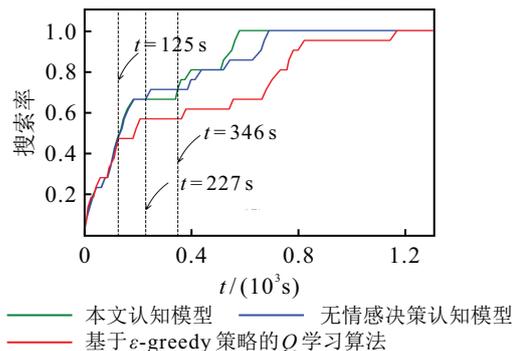


图8 搜索率对比曲线

行为决策,以应对后续搜索过程中可能面临“能量不足”无法返回的风险,因此,搜索率无上升.对3种方法进行对比可以看出,所提出模型记忆机制的引入能够通过减少机器人对已搜索区域的重复搜索提高迷宫搜索速度;而情感机制的引入则在增加机器人“拟人化”程度的同时在学习后期提高搜索速度.

统计10次学习过程中3种学习方法完成搜索所需时间,结果如图9所示.由图9可见,所提出认知模型(平均729 s)与无情感决策认知模型(平均830 s)均低于基于 ϵ -greedy策略的Q学习算法(平均1338 s).其中:具有情感决策的认知模型搜索时间相比无情感决策认知模型缩短12%,相比基于 ϵ -greedy策略的Q学习算法缩短45%.此外,情感决策加入后的认知模型在不同次实验中搜索时间变化波动(方差)更小,表明其稳定性更优.可见,具有情感和记忆机制的认知模型在学习过程中兼顾快速性和稳定性.

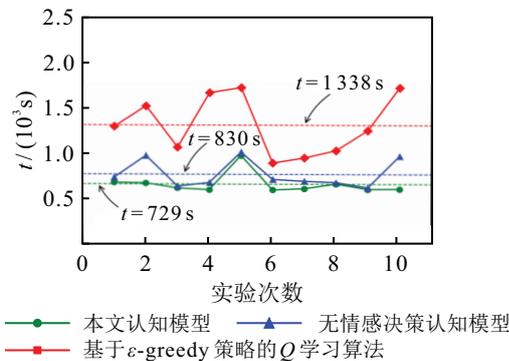


图9 搜索时间对比实验

4 结论

本文基于情感和认知生理学、脑科学等相关学科知识,以迷宫机器人为对象,建立了一种具有情感和记忆机制的认知模型.其中,情感作为内在奖励,记忆通过反思记忆的方式,引入至迷宫机器人学习过程中.以需“能量补给”的迷宫搜索任务对所提出模型进行验证,实验过程中,机器人表现出对不同情境的情感变化、决策以及分析路径搜索情况能力,表明模型及其相关算法的有效性.将所提出认知模型、无情感决策认知模型以及基于 ϵ -greedy策略的Q学习算法进行对比,实验验证所提出算法具有更快的学习效率和更好的学习稳定性.所提出认知模型情感机制的引入使得机器人具有多种人工情感,更加拟人化,在任务执行过程中会自主考虑“能量补给”,更符合实际场景全自主机器人应用.

参考文献(References)

[1] Loev S. Epistemic feelings are affective experiences[J]. *Emotion Review*, 2022, 14(3): 206-216.

- [2] Ahumada-Méndez F, Lucero B, Avenanti A, et al. Affective modulation of cognitive control: A systematic review of EEG studies[J]. *Physiology & Behavior*, 2022, 249: 113743.
- [3] Culot C, Gevers W. Happy is easy: The influence of affective states on cognitive control and metacognitive reports[J]. *Cognition and Emotion*, 2021, 35(6): 1195-1202.
- [4] Kraljević N, Schaare H L, Eickhoff S B, et al. Behavioral, anatomical and heritable convergence of affect and cognition in superior frontal cortex[J]. *NeuroImage*, 2021, 243: 118561.
- [5] 夏文然, 陈思, 王亮. 条件化恐惧记忆消退的神经振荡机制[J]. *生物化学与生物物理进展*, 2019, 46(3): 276-286.
(Xia W R, Chen S, Wang L. Neural oscillations underlying the extinction of conditioned fear memory[J]. *Progress in Biochemistry and Biophysics*, 2019, 46(3): 276-286.)
- [6] Palomero-Gallagher N, Hoffstaedter F, Mohlberg H, et al. Human pregenual anterior cingulate cortex: Structural, functional, and connective heterogeneity[J]. *Cerebral Cortex*, 2018, 29(6): 2552-2574.
- [7] Quiñones-Laracuente K, Vega-Medina A, Quirk G J. Time-dependent recruitment of prefrontal circuits for retrieval of fear memory[J]. *Frontiers in Behavioral Neuroscience*, 2021, 15: 665116.
- [8] Madan C R, Fujiwara E, Caplan J B, et al. Emotional arousal impairs association-memory: Roles of amygdala and hippocampus[J]. *NeuroImage*, 2017, 156: 14-28.
- [9] Shao K Q, Pekrun R, Nicholson L J. Emotions in classroom language learning: What can we learn from achievement emotion research?[J]. *System*, 2019, 86: 102121.
- [10] 郎悦茹, 龚少英, 曹阳, 等. 网络学习中师生交互与大学生学习投入的关系: 自主动机与学业情绪的序列中介作用[J]. *心理发展与教育*, 2022, 38(4): 530-537.
(Lang Y R, Gong S Y, Cao Y, et al. The relationship between teacher-student interaction and college students' learning engagement in online learning: Mediation of autonomous motivation and academic emotions[J]. *Psychological Development and Education*, 2022, 38(4): 530-537.)
- [11] Endres T, Weyreter S, Renkl A, et al. When and why does emotional design foster learning? Evidence for situational interest as a mediator of increased persistence[J]. *Journal of Computer Assisted Learning*, 2020, 36(4): 514-525.
- [12] Picard R W. *Affective computing*[M]. Cambridge: MIT Press, 2000: 1-304.

- [13] Gadanho S C. Learning behavior-selection by emotions and cognition in a multi-goal robot task[J]. *Journal of Machine Learning Research*, 2003, 4(3): 385-412.
- [14] Sequeira P, Melo F S, Paiva A. Emotion-based intrinsic motivation for reinforcement learning agents[C]. *International Conference on Affective Computing and Intelligent Interaction*. Heidelberg, 2011: 326-336.
- [15] 李宏光, 代凯峰, 刘振娟, 等. 一类面向多目标决策的分层情感认知模型[J]. *控制与决策*, 2015, 30(12): 2129-2136.
(Li H G, Dai K F, Liu Z J, et al. A kind of layered affective cognition model for multi-objective decision-making[J]. *Control and Decision*, 2015, 30(12): 2129-2136.)
- [16] Zhong Q B, Fang B F, Guo X P, et al. Task allocation for affective robots based on willingness[J]. *IEEE Access*, 2021, 9: 80028-80042.
- [17] Ren S C, Wang Y L, Yue F G, et al. The paraventricular thalamus is a critical thalamic area for wakefulness[J]. *Science*, 2018, 362(6413): 429-434.
- [18] Rolls E T. The cingulate cortex and limbic systems for emotion, action, and memory[J]. *Brain Structure & Function*, 2019, 224(9): 3001-3018.
- [19] Staniloiu A, Markowitsch H J. Episodic memory is emotionally laden memory, requiring amygdala involvement[J]. *The Behavioral and Brain Sciences*, 2020, 42: e299.
- [20] Battaglia-Mayer A. A brief history of the encoding of hand position by the cerebral cortex: Implications for motor control and cognition[J]. *Cerebral Cortex*, 2018, 29(2): 716-731.
- [21] Zhang X P, Liu Y H, Hu D L, et al. A maze robot autonomous navigation method based on curiosity and reinforcement learning[C]. *The 7th International Workshop on Advanced Computational Intelligence and Intelligent Informatics*. Beijing, 2021: 1-6.
- [22] Harmon-Jones E, Harmon-Jones C, Summerell E. On the importance of both dimensional and discrete models of emotion[J]. *Behavioral Sciences*, 2017, 7(4): 66.
- [23] Dalglish T, Power M J. *Handbook of cognition and emotion*[M]. Chichester: John Wiley & Sons, Ltd, 1999: 1-866.
- [24] Mehrabian A. Pleasure-arousal-dominance: A general framework for describing and measuring individual differences in Temperament[J]. *Current Psychology*, 1996, 14(4): 261-292.
- [25] Ortony A. *The cognitive structure of emotions*[M]. Cambridge: Cambridge University Press, 1990: 1-226.
- [26] Kshirsagar S. A multilayer personality model[C]. *Proceedings of the 2nd International Symposium on Smart Graphics*. New York, 2002: 107-115.
- [27] 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.
(Ruan X G, Zhang J H, Huang J, et al. An environment cognition model combined with intrinsic motivation for mobile robots[J]. *Control and Decision*, 2021, 36(9): 2211-2217.)

作者简介

张晓平(1991—), 女, 副教授, 博士, 从事人工智能、机器学习、智能机器人与认知机器人等研究, E-mail: zhangxiaoping369@163.com;

李凯(1996—), 男, 硕士生, 从事机器学习、认知机器人、情感机器人等研究, E-mail: 17674764382@163.com;

王力(1978—), 男, 教授, 博士生导师, 从事智能交通控制、道路交通工程等研究, E-mail: wangli939@ncut.edu.cn;

闫佳庆(1985—), 男, 副教授, 博士, 从事计算神经科学、神经控制工程等研究, E-mail: yjq@ncut.edu.cn;

何忠贺(1982—), 男, 副教授, 博士, 从事城市道路交通网络建模与控制、混杂动态系统切换控制与应用等研究, E-mail: zhonghehe@ncut.edu.cn.