

# 控制与决策

Control and Decision

移动机器人动态避障的调节发育学习

王东署, 赵红燕

引用本文:

王东署, 赵红燕. 移动机器人动态避障的调节发育学习[J]. 控制与决策, 2023, 38(11): 3112–3120.

在线阅读 View online: <https://doi.org/10.13195/j.kzyjc.2021.1726>

---

## 您可能感兴趣的其他文章

### Articles you may be interested in

#### 一种基于免疫机理的确定性移动机器人路径规划算法

A path planning algorithm of deterministic mobile robot based on immune mechanism

控制与决策. 2021, 36(10): 2418–2426 <https://doi.org/10.13195/j.kzyjc.2020.0059>

#### 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>

#### 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>

#### 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>

#### 基于免疫优化的平面Acrobot线性自抗扰鲁棒镇定

Robust stabilization of planar Acrobot using linear active disturbance rejection control with immune optimization

控制与决策. 2020, 35(12): 3053–3058 <https://doi.org/10.13195/j.kzyjc.2019.0289>

# 移动机器人动态避障的调节发育学习

王东署<sup>†</sup>, 赵红燕

(郑州大学 电气与信息工程学院, 郑州 450001)

**摘要:** 在环境认知的动态避障过程中,除了预期不确定性事件,移动机器人还可能会遇到非预期不确定性事件.如何高效、灵活地应对非预期不确定性事件是移动机器人动态避障中面临的一个重要挑战.目前关于这方面的研究相对较少,且基于这些研究的移动机器人普遍缺乏自主学习能力,难以快速、灵活地应对突变的外部环境.鉴于此,首先,设计一个新的碰撞危险度指标,该指标不仅考虑障碍物的距离,同时也考虑障碍物速度对移动机器人运动的影响.模拟人脑中乙酰胆碱和去甲肾上腺素在应对环境不确定性时的反应机理,通过碰撞危险度指标引导移动机器人的注意力网络在关注预期刺激的背侧注意力网络和关注新刺激的腹侧注意力网络之间切换,使得机器人灵活应对环境中的不确定性事件;然后,设计新的神经元学习率,以增强调节发育网络隐含层神经元的学习能力,提高机器人应对突变环境的快速响应能力;接着,修改突触权值更新规则,以提高移动机器人行为决策的准确性;最后,通过在两种不同场景下的仿真实验以及物理环境中的实验,验证所提出的应对环境中非预期不确定性事件的移动机器人调节发育学习方法的可行性.

**关键词:** 预期不确定性; 非预期不确定性; 碰撞危险度; 调节发育网络; 注意力网络

中图分类号: TP242 文献标志码: A

DOI: 10.13195/j.kzyjc.2021.1726

引用格式: 王东署,赵红燕.移动机器人动态避障的调节发育学习[J].控制与决策,2023,38(11):3112-3120.

## Motivated developmental learning of mobile robots in dynamic collision-avoidance

WANG Dong-shu<sup>†</sup>, ZHAO Hong-yan

(School of Electrical and Information Engineering, Zhengzhou University, Zhengzhou 450001, China)

**Abstract:** In dynamic collision-avoidance of the environmental cognition, except the expected uncertainty, mobile robots may also encounter unexpected uncertainty. Studying how to deal with the unexpected uncertainty efficiently and flexibly is an important challenge for mobile robots. At present, there are relatively few studies on this aspect, and the mobile robots based on these studies generally lack the ability of autonomous learning, and it is difficult to quickly and flexibly respond to the abrupt external environment. In this paper, a novel collision risk index is designed, which not only considers the influence of the distance of the obstacle, but also that of the speed of the obstacle on the motion of the mobile robots. Simulating the reaction mechanism of the acetylcholine and norepinephrine in human brain in response to environmental uncertainty, through the collision risk index, the attention network of the mobile robot is guided to switch between the dorsal attention network which focuses on the expected stimulus, and the ventral attention network which focuses on new stimulus, and make the robot flexible response to the uncertainty in the environment. At the same time, a new neuronal learning rate is designed to enhance the learning ability of neurons in the hidden layer of the motivated developmental network and improve the robot's ability to respond quickly to the abrupt environment. In addition, the synaptic weight updating rule is modified to improve the accuracy of the mobile robot's behavioral decision. Simulation results in two different scenarios, as well as the physical experiment, verify the feasibility of the proposed motivated development learning method in response to unexpected uncertainty in the environment.

**Keywords:** expected uncertainty; unexpected uncertainty; collision risk; motivated developmental network; attention network

收稿日期: 2021-10-08; 录用日期: 2022-06-24.

基金项目: 国家自然科学基金项目(62173309); 河南省自然科学基金项目(202300410483).

责任编辑: 高会军.

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

\*本文附带电子附录文件,可登录本刊官网该文“资源附件”区自行下载阅览.

## 0 引言

作为机器人与人工智能技术相结合的完美载体,移动机器人已广泛应用于人类的生产和日常生活中,如工业<sup>[1]</sup>、农业<sup>[2]</sup>、军事<sup>[3]</sup>、医疗<sup>[4]</sup>、太空探索<sup>[5]</sup>和家政服务<sup>[6]</sup>等,高效、灵活的环境认知是移动机器人完成各项任务的重要前提。在环境认知过程中,移动机器人一般不具有环境的先验知识,运动过程中可能会遇到各种障碍物,甚至各种突发情况。如何有效调整机器人的运动行为,使其能够快速灵活地应对变化的环境,特别是突发的不确定的事件或场景,即非预期不确定性事件,是移动机器人行为决策中必须解决的一个重要问题。

当环境中发生非预期不确定性事件时,研究人员通过计算不确定性事件的概率<sup>[7]</sup>、预测非预期不确定行为<sup>[8]</sup>以及采用神经网络<sup>[9]</sup>等方法来确定机器人的行为。但是,概率预测方法需要大量的统计数据,有些数据在实际现场中很难获得,难以保证行为预测结果的准确性;基于神经网络的机器人行为决策需要大量的训练样本,学习效率低,当样本类别不确定或样本数量较少时,决策效果并不理想。

近年来,随着神经科学的发展,研究人员开始研究生物体在面对环境中突发不确定事件时的神经调节机制,尝试将生物神经调节系统的工作机理应用于移动机器人的行为决策中。Krichmar<sup>[10]</sup>和Dasgupta等<sup>[11]</sup>研究表明,基于生物神经调节系统的构型可使得智能体具有自主认知和经验学习的能力,使它们能够更好地探索环境以执行任务,并在环境变化时迅速作出合适的决策。

研究表明<sup>[12-13]</sup>,神经调节系统中的去甲肾上腺素、血清素(5-羟色胺)、多巴胺和乙酰胆碱在处理风险、奖励、新颖等环境信号方面发挥着重要作用。故研究人员尝试将人脑的神经调节机制应用于移动机器人的行为决策中,包括处理环境中的突发不确定性事件。Krichmar<sup>[14]</sup>用突然出现的光束模拟环境中的突发不确定性事件,通过前额叶皮质到神经调节系统的自上而下的调节信号指导机器人的行为决策。Khamassi等<sup>[15]</sup>提出了一种新的计算模型,通过研究外侧前额叶与前扣带皮层间的相互作用,以应对环境中的突发不确定性。本文曾对此问题进行了初步探索<sup>[16]</sup>,除多巴胺和5-羟色胺外,进一步研究了乙酰胆碱和去甲肾上腺素在移动机器人应对环境不确定性事件时的调节机理,但是前期研究只考虑了障碍物距离对移动机器人行为决策的影响,网络参数调节方法较为简单,导致机器人在应对突发不确定性事件时

灵活性有限。

考虑到神经调节系统在注意力、情绪、目标引导的行为和决策等方面的重要作用,在前期工作的基础上,本文设计了一个新的调节发育网络模型,以解决移动机器人在面对环境中的突发不确定性障碍物时的行为决策问题。除多巴胺和血清素外,本文进一步研究了乙酰胆碱和去甲肾上腺素在移动机器人应对环境中的突发不确定性事件的行为调节机制。此外,还设计了碰撞危险度指标,该指标不仅考虑了障碍物距离,同时考虑了障碍物速度对机器人运动行为的影响。利用碰撞危险度,通过调节机器人的注意力网络在关注预期刺激的背侧注意力网络和关注非预期刺激的腹侧注意力网络之间的切换,使得移动机器人具有灵活、快速的行为调节能力。最后通过2个仿真场景以及真实的物理场景,验证了所提出调节发育网络模型在处理机器人应对环境中非预期不确定性事件时的行为决策能力。

本文主要内容如下:1)与传统方法一般只考虑障碍物距离对机器人运动的影响不同,本文同时考虑了障碍物速度和距离对机器人运动行为的影响;2)为使机器人能够更快速地适应变化的环境,设计了调节发育网络隐含层神经元新的学习率;3)为降低输入噪声对机器人动作决策的影响,在top-*k*竞争机制中采用了多获胜神经元策略;4)前期工作只研究了如何修改机器人的运动方向,本文进一步研究了如何调整机器人的运动速度,使得机器人更好地应对环境中的非预期不确定性事件。

## 1 调节发育学习的实现

### 1.1 神经调节系统的生理学基础

Aston-Jones等<sup>[17]</sup>、Briand等<sup>[18]</sup>以及Krichmar<sup>[19]</sup>研究指出,神经调节系统有两种活动模式:基音模式和相位模式。基音模式下,动物更有可能考虑其周围的干扰物,此时神经调节活动与环境干扰和动物自身的不良表现相关;相位模式下,由于相位信号具有增强信号检测和驱动决策过程的认知功能,此时动物的注意力更加集中且其决策更加果断。

此外,蓝斑的阶段性的激活对“网络重置”<sup>[20]</sup>至关重要,会导致整个大脑的神经活动大规模重新配置,从而引起生物体运动行为和认知学习的变化,这与将注意力导向关注预期刺激的背侧注意网络和关注新刺激的腹侧注意网络之间的转换有关。

Corbetta等<sup>[21]</sup>和Yu等<sup>[22]</sup>的研究表明,乙酰胆碱和去甲肾上腺素分别对环境中的预期不确定事件和非预期不确定事件作出响应。特别地,乙酰胆碱-去甲

肾上腺素系统能够调节腹侧被盖区和中缝核与眶额叶皮层和内侧前额叶皮层神经元的连接,并抑制眶额叶皮层和内侧前额叶皮层与腹侧被盖区和中缝核神经元的连接.

此外,Seo等<sup>[23]</sup>研究指出,当小鼠暴露于不同程度的危险环境中时,其体内的5-羟色胺系统功能会发生相应的变化,会从抑制运动转变为促进运动.如当小鼠面临低度或中度危险时,5-羟色胺系统会促使其迅速调节当前的运动并及时停止前进,而当其面临严重危险时,5-羟色胺系统则会驱使小鼠迅速逃离危险,更好地适应环境.

### 1.2 调节发育网络结构

结合前述生理学依据,可推理如下:当生物体接收到一个外界环境刺激,乙酰胆碱/去甲肾上腺素系统判断该刺激是预期不确定性刺激还是非预期不确定性刺激,据此在关注预期不确定性的背侧注意力网络与关注非预期不确定性的腹侧注意力网络间进行切换.若是预期不确定性,则相当于生物体面临中、轻度危险,乙酰胆碱含量升高,系统进入关注预期刺激的背侧注意力网络,调节血清素含量驱使其调节运动,及时止步;否则去甲肾上腺素含量升高,系统进入关注新刺激的腹侧注意力网络,类似于生物体面对重度危险,对应于非预期不确定性情况下机器人运动行为灵活调节的类脑机理.

当机器人进入关注新刺激的腹侧注意力网络,去甲肾上腺素与乙酰胆碱立即发生反应,通过调节中缝核以及腹侧被盖区合成血清素和多巴胺的量以调节发育网络参数,迅速调节机器人的运动行为,灵活应对环境中突发的非预期不确定性事件.

基于上述分析,本文提出一种调节发育网络模型,如图1所示,用于移动机器人应对环境认知中的预期和非预期不确定性事件.

本文在前期研究的基础上<sup>[16]</sup>,并参考文献[24],扩展网络功能,如图1所示:1)增加蓝斑到基底前脑间的单向连接,以模拟去甲肾上腺素对乙酰胆碱的调节;2)增加中缝核到腹侧被盖区的抑制连接,模拟二者间的对立性能<sup>[25]</sup>;3)将基底前脑和中缝核之间、基底前脑和腹侧被盖区之间以及蓝斑和腹侧被盖区之间原来的单向连接改为双向连接;4)前期工作只研究了障碍物距离对机器人运动行为的影响,本文进一步研究障碍物速度对机器人运动行为的影响;5)前期工作只研究如何改变机器人的运动方向,本文进一步研究如何调节机器人的运动速度,更好地应对环境中的非预期不确定性事件;6)模拟人类的遗忘规

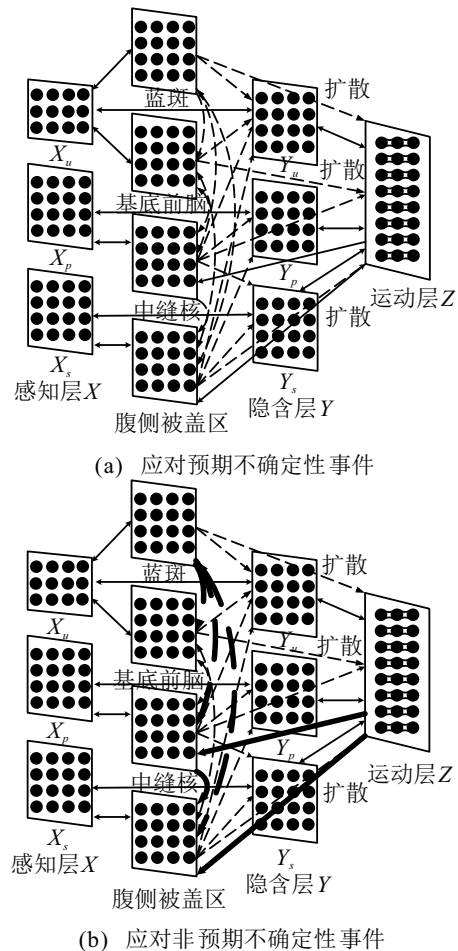


图1 增加乙酰胆碱/去甲肾上腺素后的调节发育网络模型

律,引入遗忘因子,进一步加快神经元的学习速度;7)改进中间层神经元的权值更新规则.

在如图1所示的调节发育网络中,感知层X、隐含层Y和运动层Z均被分为3个部分,即无偏区、痛觉区和甜觉区,分别用下标u、p和s表示.感知层X可表示为 $X = (X_u, X_p, X_s)$ .其中: $X_u$ 为无偏输入, $X_p$ 和 $X_s$ 分别为痛觉和甜觉输入.在任意时刻 $t$ ,感知层X可根据外部环境的状态产生一个响应向量. $X_u$ 直接来自于感知输入, $X_p$ 和 $X_s$ 作为转换开关分别用以确定机器人是否受到惩罚或奖励.

隐含层Y分为3个子层,即 $Y_u$ 、 $Y_p$ 和 $Y_s$ ,分别对应X的3个子层.若机器人、障碍物与目标间的位置关系满足一定的条件,则中缝核释放血清素用于惩罚,腹侧被盖区释放多巴胺用以奖励机器人的行为.释放的血清素和多巴胺分别对 $Y_p$ 和 $Y_s$ 层中的神经元产生非0影响,这种非0影响会进一步影响 $Y_u$ 层神经元的学习率.考虑血清素和多巴胺的影响, $Y_u$ 层神经元的学习率可描述<sup>[26]</sup>如下:

$$\omega_2(n_j) = \min \left( (1 + \alpha_{RN} + \alpha_{VTA}) \cdot \frac{1}{n_j}, 1 \right), \quad (1)$$

其中 $\alpha_{RN}$ 和 $\alpha_{VTA}$ 分别为与中缝核和腹侧被盖区相关

的变量.

运动皮层  $Z$  可用一系列神经元  $Z = (z_1, z_2, \dots, z_m)$  描述, 其中  $m$  为运动层神经元的个数. 如图1所示, 每个  $z_i$  含有3个神经元  $z_i = (z_{i_u}, z_{i_p}, z_{i_s})$ , 其中  $z_{i_u}$ 、 $z_{i_p}$  和  $z_{i_s}$  分别为无偏、痛觉和甜觉运动神经元. 因此, 某个动作  $i$  是否被激活不仅取决于  $z_{i_u}$  的响应, 同时也取决于  $z_{i_p}$  和  $z_{i_s}$  的响应. 运动神经元激活前的能量值可计算如下:

$$z_i = z_{i_u} \gamma (1 - \alpha z_{i_p} + \beta z_{i_s}), \quad (2)$$

其中  $\alpha$ 、 $\beta$  和  $\gamma$  为正常数.

最后, 第  $j$  个运动神经元被激活, 其对应的动作被执行 (在本文中是指机器人对应的运动方向), 有

$$j = \arg \max_{1 \leq i \leq m} \{z_i\}. \quad (3)$$

### 1.3 类脑调节方案设计

当有新的环境信息进入机器人“大脑”, 系统首先判断当前时刻机器人探测范围内的障碍物个数是否与前一时刻的数量相同, 若不大于前一时刻的障碍物数量, 则属于常规情况下的运动控制, 系统进入关注预期不确定性的背侧注意力网络, 如图1(a)所示. 反之, 系统进一步比较新出现的障碍物与机器人间的碰撞危险度和设定阈值的关系, 若小于设定的阈值, 则仍然按常规情况下的运动控制进行处理, 否则, 系统进入如图1(b)所示的关注新刺激的腹侧注意力网络.

#### 1) 碰撞危险度设计.

生物体的注意力切换受乙酰胆碱/去甲肾上腺素系统的调节, 同时考虑到生物体在面临重度危险时, 5-羟色胺系统能够促使其迅速逃离危险, 这种行为是通过乙酰胆碱/去甲肾上腺素调节中缝核合成5-羟色胺实现的. 模拟这种调节机理, 所设计的碰撞危险度指标, 应体现出“通过5-羟色胺对生物体行为的惩罚作用, 以调节生物体的运动行为”的原则, 而这种惩罚是通过调节机器人与障碍物间的距离实现的. 遵循这样的原则, 同时考虑到障碍物与智能体间相对速度越大, 碰撞危险度越大, 二者间相对距离越小, 则碰撞危险度越大, 所设计的碰撞危险度  $R$  应体现出障碍物速度和距离的影响. 将碰撞危险度的变化范围设置为  $0 \sim 1$  间. 当智能体在碰撞区域外时,  $R$  设置为0; 当智能体在碰撞区域内, 且其速度超出安全值或与障碍物的距离小于智能体的安全距离  $d_s$  时,  $R$  设置为1; 当智能体在碰撞区域内, 其速度没有超出安全值且与障碍物的距离大于机器人的安全距离  $d_s$  时,  $R$  设置为

$$R = \frac{2}{\pi} \cdot \arccot \left( \left( \frac{d_{\min}}{v'} - k_t T \right) (d_{\min} - d_s) \right). \quad (4)$$

其中:  $d_{\min}$  为障碍物与机器人间的最小距离,  $v'$  为障碍物速度在  $d_{\min}$  方向上的投影,  $k_t$  为调节参数,  $T$  为机器人的决策周期,  $d_s$  为机器人的安全距离.

#### 2) 学习率的调节.

首先, 模拟内侧前额叶皮层到中缝核以及眶额叶皮层到腹侧被盖区的抑制连接, 分别通过调节中缝核和腹侧被盖区合成5-羟色胺和多巴胺的量以调节发育网络中间层神经元的学习率.

考虑上述神经调节作用后, 隐含层神经元新的学习率可定义为

$$\omega_2(n_j) = \min \left( (1 + \alpha'_{RN} + \alpha'_{VTA}) \cdot \frac{1}{n_j}, 1 \right), \quad (5)$$

其中  $\alpha'_{RN}$  和  $\alpha'_{VTA}$  分别为与中缝核和腹侧被盖区相关的新变量.

式(5)中, 考虑到当非预期不确定性事件发生时, 内侧前额叶皮层到中缝核的连接应使得惩罚效应迅速增大, 以快速躲避突发的障碍物. 因此惩罚因子可表示为

$$\alpha'_{RN} = \alpha_{RN} \cdot k_1 \cdot (1 + \tan(R)), \quad (6)$$

其中  $k_1 > 1$  为常数.

然后, 考虑到在面对非预期不确定性事件时, 人脑只关注如何迅速躲避突发的障碍物, 而暂时不关心是否仍然追踪目标, 模拟这种应对方式, 在学习率中引入另一个随碰撞危险度自适应调节的调节因子, 以模拟中缝核到腹侧被盖区的抑制连接, 有

$$\alpha'_{VTA} = \alpha_{VTA} \cdot k_2 \cdot \left( 1 - \tan \left( \frac{\pi}{4} \cdot R \right) \right), \quad (7)$$

其中  $k_2 > 1$  为常数.

最后, 设计与神经元发放次数  $n$  相关的遗忘因子  $\mu(n)$ , 模拟人类学习速度与年龄的大致关系, 加快突触权值的调节速度. 遗忘因子设置<sup>[27]</sup>为

$$\mu(n) = \begin{cases} 0, & n \leq t_1; \\ c \cdot (n - t_1) / (t_2 - t_1), & t_1 < n < t_2; \\ c + (n - t_2) / r, & t_2 \leq n. \end{cases} \quad (8)$$

其中  $c$ 、 $t_1$ 、 $t_2$  和  $r$  为与神经元发放次数  $n$  有关的常数.

综合以上3个因素, 调节发育网络隐含层神经元的新学习率可表示为

$$\omega_2(n_j) = \min \left( \left( (1 + \alpha_{RN} \cdot k_1 \cdot (1 + \tan(R)) + \alpha_{VTA} \cdot k_2 \cdot \left( 1 - \tan \left( \frac{\pi}{4} \cdot R \right) \right)) \cdot \frac{1 + \mu(n_j)}{n_j}, 1 \right) \right). \quad (9)$$

利用新的学习率, 计算保持率  $\omega_1(n_j) = 1 - \omega_2(n_j)$ , 并通过叶成分分析方法<sup>[28]</sup> 计算发育网络感知层神经元到中间层神经元之间以及中间层神经元到运动皮层神经元的突触权值.

### 3) 改进的权值更新原理.

在更新中间层神经元的权值时,采用多个获胜神经元的策略<sup>[29]</sup>,具体更新规则描述如下.

按照预响应值大小对获胜的前  $k$  个神经元进行排序,相邻位次神经元之间新旧知识影响比例差值为  $1/k$ ,即预响应值最大的获胜神经元权值更新时来自输入向量的新知识的影响比例为 1,而第 2 获胜神经元权值更新时来自输入向量的新知识的影响比例为  $1 - 1/k$ ,第 3 获胜神经元权值更新时来自输入向量的新知识的影响比例为  $1 - 2/k$ ,以此类推,第  $k$  个获胜神经元权值更新时来自输入向量的新知识的影响比例为  $1 - (k - 1)/k = 1/k$ . 从第  $k + 1$  个神经元开始的所有神经元,权值更新时来自输入向量的新知识的比例全部为 0,只有来自于旧知识的更新.

结合以上分析,可设计出如下权重更新公式:

$$v_j = (1 + \mu)\omega_1(n_j)v_j + (1 - \mu)\omega_2(n_j)y_j\dot{p}, \quad (10)$$

其中  $\mu$  即为前述的调整因子  $1/k$ .

### 4) 机器人运动方向和速度的更新.

在运动皮层中,采用 top- $k$  竞争机制,即更新前  $k$  个预响应值最大的神经元(一般  $k = 1$ ,以唯一确定机器人的运动方向).运动方向调节时,需要在表示惩罚方向的对应项和表示奖励方向的对应项中分别引入相应的调节因子,并与表示无偏感觉输入对应的运动方向进行线性组合,确定最终的运动方向,如式(2)和(3)所述.

类似地,机器人的速度大小也应随碰撞危险度进行自适应调节,有

$$z(t) = \begin{cases} 0, & d_{\min} \leq d_s; \\ z(t-1) \cdot (1 + k_3 \cdot R \cdot \exp(\tau)), & \text{otherwise.} \end{cases} \quad (11)$$

## 2 仿真实验

### 2.1 实验设置

为验证调节发育网络的性能,本节设计了 2 个仿真实验.为展示智能体在应对环境中非预期不确定性事件时的行为决策,本节对比了 3 种调节发育网络的性能,即原始调节发育网络、包含两种神经递质的调节发育网络和所提出的当前调节发育网络.其中原始调节发育网络基于工作<sup>[30]</sup>,只研究了多巴胺和血清素对运动层神经元学习率的影响;包含两种神经递质的调节发育网络<sup>[26]</sup>,进一步研究了血清素和多巴胺对隐含层神经元学习率的影响;当前调节发育网络是在前期工作<sup>[16]</sup>的基础上,进一步研究了乙酰胆碱/去甲肾上腺素系统对调节发育网络参数调节

的影响.

### 2.2 输入和输出

定义如下实体,智能体(agent)、目标(target)和障碍物(obstacle),3个实体的相对位置关系如图2所示,可获得如下表达式:

$$\begin{aligned} \theta_t &= \arctan(a_x - t_x, a_y - t_y), \\ d_t &= \sqrt{(a_x - t_x)^2 + (a_y - t_y)^2}, \\ \theta_o &= \arctan(a_x - o_x, a_y - o_y), \\ d_o &= \sqrt{(a_x - o_x)^2 + (a_y - o_y)^2}, \\ x_u &= \left\{ \cos \theta_t, \sin \theta_t, \cos \theta_o, \sin \theta_o, \frac{d_t}{d_t + d_o}, \frac{d_o}{d_t + d_o} \right\}. \end{aligned} \quad (12)$$

其中:  $(a_x, a_y)$ 、 $(t_x, t_y)$  和  $(o_x, o_y)$  分别为智能体、目标和障碍物的位置坐标,  $\theta_t$  和  $\theta_o$  分别为智能体航向角与目标方向和障碍物方向间的夹角,  $d_t$  和  $d_o$  分别为智能体与目标和障碍物间的距离.

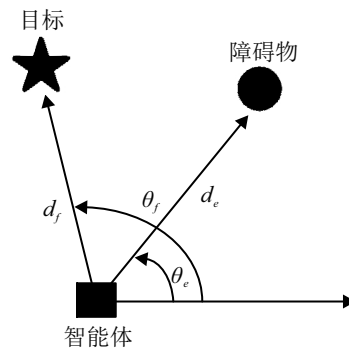


图2 智能体、目标与障碍物间的位置关系

痛觉感受器和甜觉感受器仅用一个值表示惩罚和奖励的程度.将惩罚阈值设定为 65,若  $d_e > 65$ ,没有惩罚;若  $22 < d_e \leq 65$ ,惩罚值用公式  $(65 - d_e)/43$  计算;否则惩罚值等于  $22/d_e$ .

类似地,将奖励阈值设为 45,若  $d_f < 45$ ,没有奖励;若  $45 < d_f \leq 200$ ,奖励值用公式  $(d_f - 45)/155$  计算;否则奖励值等于  $d_f/200$ .

智能体的运动方向由式(2)确定,在式(2)中,3个变量设定如下:  $\alpha = \alpha_{RN} \times \sqrt{d_o}$ ,  $\beta = \alpha_{VTA} \times \sqrt{d_t}$ ,  $\gamma = \alpha_{VTA} \times \log(d_t)$ .

### 2.3 动态环境

实验中,除智能体外,存在 3 个动态障碍物、2 个静态障碍物、1 个突然出现的动态障碍物和 1 个动态目标.障碍物、目标、突然出现的动态障碍物和智能体的初始位置如图 3 所示.

运动过程中,为观察智能体对障碍物的躲避以及追逐目标的情况,分别记录了 3 个模型下智能体与最近的障碍物和目标间的距离,如图 4 所示.

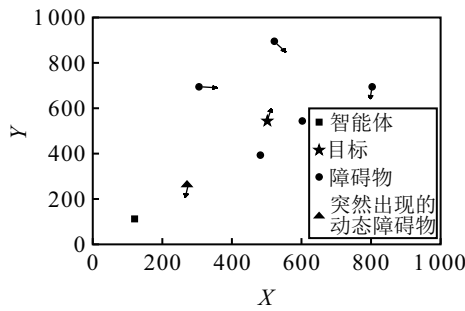
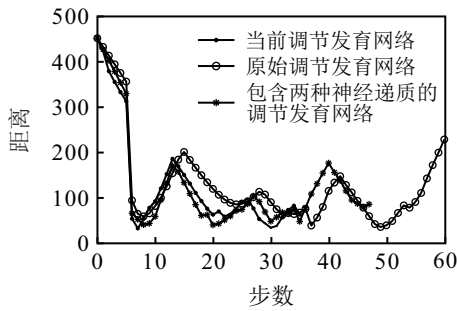
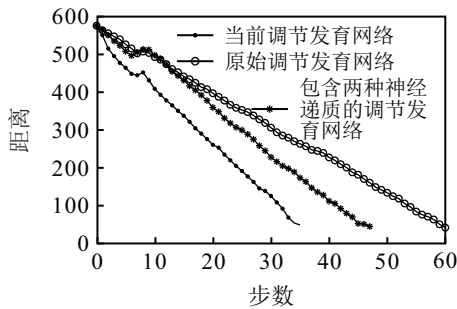


图3 动态环境结构



(a) 智能体与最近障碍物间的距离



(b) 智能体与目标间的距离

图4 动态环境下距离对比

由图4(a)可见,在第6步之前,3个模型中智能体与最近障碍物的距离没有太大差异. 由于没有新的刺激出现,乙酰胆碱/去甲肾上腺素系统不起作用,这3个模型的运动轨迹区别不大. 在第6步,动态障碍物突然出现,智能体与最近障碍物的距离突然减小,相比于其他两种模型,智能体在当前调节发育网络模型中距离突然出现的障碍物最近,但其却在几步内便避开障碍物到达最远的安全距离,可见智能体在当前调节发育网络下具有更强的学习和调节能力. 因此,智能体可迅速避开障碍物. 由图4(b)可见,在3个模型中,当前调节发育网络模型下智能体运行步数最少.

由图4可见,遇到新刺激前,由于动态障碍物的存在,智能体进入关注预期不确定性的背侧注意力网络,但是当遇到新刺激时,智能体进入关注非预期不确定性的腹侧注意力网络,神经调节系统中的去甲肾上腺素和乙酰胆碱立即发生反应,通过调节中缝核以及腹侧被盖区合成5-羟色胺和多巴胺的量以调节发育网络中间层神经元到运动皮层神经元的突触连接强度,进而调节运动皮层神经元的响应,使得智能体

方向迅速变化,速度增大,快速远离该障碍物. 经过一段时间后,突发障碍物的影响慢慢消失,智能体又进入关注预期不确定性的背侧注意力网络并继续运动直至到达目标.

### 2.4 对比实验

本节将当前调节发育网络与人工势场法、动态窗口法和Q学习进行性能对比.

在人工势场法中,引力因子设为10,斥力因子设为20,当离最近障碍物距离大于65时,斥力消失,只有引力. 在动态窗口法中,设置时间间隔为0.1s,障碍物半径为0.1m. 在Q学习中,奖励衰减值为0.9,学习率为0.5. 本次实验的环境如图5所示,存在2个静态障碍物、3个动态障碍物、1个动态的目标以及1个突然出现的动态障碍物. 当智能体运行至第14步时,智能体与目标间突然出现一个障碍物,它们的位置关系如图5所示.

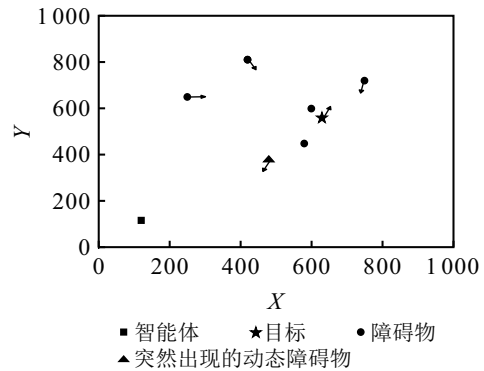
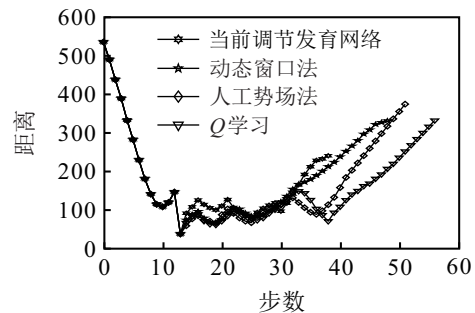
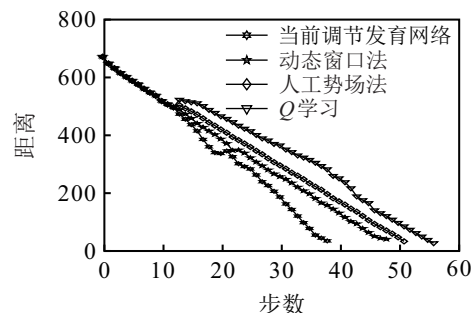


图5 对比实验时的环境结构

由于这4种方法走的路径不同,为了有效地展示



(a) 智能体与最近障碍物间的距离



(b) 智能体与目标间的距离

图6 对比实验时的距离比较

这4种方法在应对突然出现的障碍物时的躲避性能,在该障碍物突然出现前,只使用当前调节发育网络方法进行运动,使得智能体在前期时具有完全相同的运动轨迹.第14步后,分别使用当前调节发育网络和另外3种方法规划智能体的运动轨迹.同样,实验记录了智能体与最近障碍物和目标间的距离,如图6所示.

由图6可见,在动态障碍物突然出现后,由于乙酰胆碱/去甲肾上腺素系统的自适应调节,当前调节发育网络能够快速响应环境中的非预期不确定性,而另外3种方法则不能.

为了更好地比较它们的避障性能,对比了智能体的运行步数和遇到突然出现的障碍物后的避障距离,其结果如表1所示.

表1 4种算法实验结果

算法	运行步数	遇到突发障碍物后的避障距离
人工势场法	51	49.83
动态窗口法	48	65.32
Q学习	56	57.41
当前调节发育网络	38	83.56

由表1可见,当前调节发育网络模型下智能体运行步数最少,且在遇到突发障碍物后有着最大的避障距离.综上,在这4种方法中,当前调节发育网络对非预期不确定性的响应性能最好.

### 3 物理实验

为了验证算法模型在实际环境中的应用效果,在实物环境中进行了实验.使用Rikirobot移动机器人作为研究对象,另外一个Rikirobot移动机器人作为突然出现的动态障碍物.

将实物环境设置为长4m、宽3m的矩形环境,有12个静态障碍物,其初始环境如图7所示.其中机器人初始点为原点,目标点设置为(3.7, -2.6).设置机器人与障碍物的安全距离为0.65m,机器人进入关注非预期不确定性的腹侧注意力网络的距离阈值设为0.3m,没有遇到障碍物的匀速运动的速度为0.4m/s,机器人运行至距离目标0.2m以内停止运动.为了模拟突然出现的障碍物,初始时,将动态障碍物放在一个静态障碍物的后边,在运行程序使得机器人运动



图7 物理实验中初始环境结构

后,使动态障碍物以0.8m/s的速度快速移动突然接近机器人,模拟突然出现的效果.

机器人的运动如图8所示.由图8可见,机器人从初始点出发,到达图8(a)中的位置,此时动态障碍物并没有运动,但是到了图8(b),控制动态障碍物突然冲到机器人的面前,机器人迅速反应调节方向并在图8(c)避开障碍物,接着机器人继续运动,在避开了两边的静态障碍物的同时,动态障碍物继续移动,二者又在图8(d)相遇,此时机器人在图8(e)又避开了该动态障碍物,继续前进,直至在图8(f)到达目标点停止.

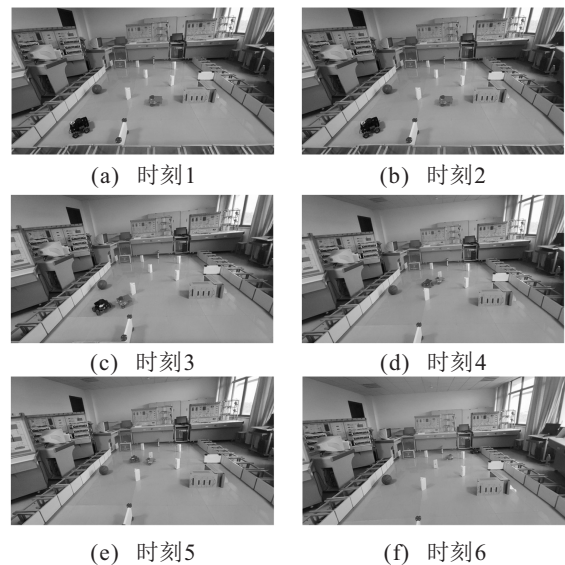


图8 机器人的实际运动过程描述

为了更清晰地显示机器人对突然出现的障碍物的避障效果,将机器人和动态障碍物同时启动,分别记录它们的轨迹点和每一步的运行时间,并将这些数据整合到一张图以观察它们的相遇情况,其结果如图9所示.

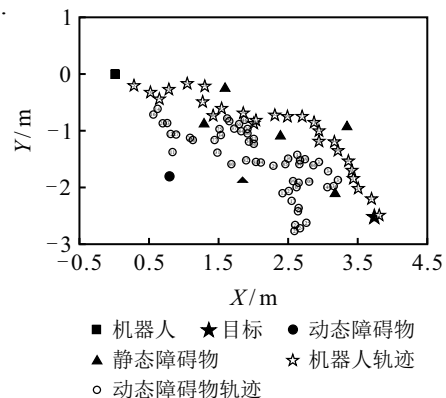


图9 机器人和障碍物的运行轨迹

由图9可见,机器人在第14s处遇见并避开了突然出现的动态障碍物,然后在第29s处遇见并避开了第1个静态障碍物,之后又在第40s处遇见了第2个静态障碍物,避开后又第55s处遇到了动态障碍物,再次避开后又避开了后边的3个静态障碍物到达了

目标。

图 10 为机器人距离最近的障碍物和目标的距离随运行步数的变化情况。

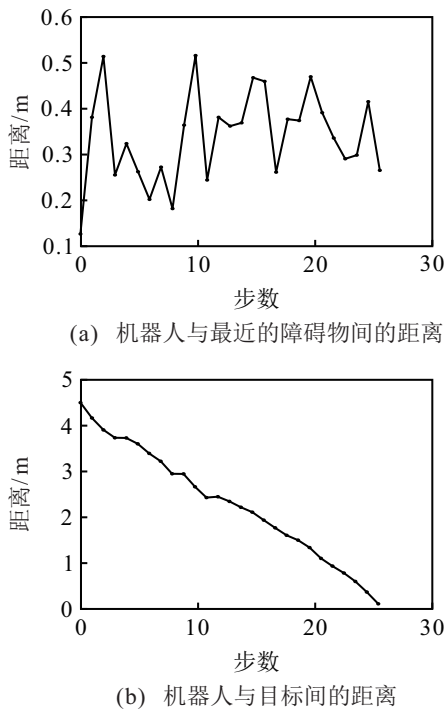


图 10 实物实验中的距离对比

由图 10(a) 可见, 机器人开始运动后在第 3 步遇见了突然出现的障碍物, 其与最近障碍物的距离减小至 0.25 m, 小于非预期不确定性阈值 0.3 m, 此时机器人进入了关注非预期不确定性的腹侧注意力网络, 快速调节运动方向和速度避开了障碍物。此后, 该障碍物继续在环境中充当动态障碍物, 因此机器人又进入了关注预期不确定性的背侧注意力网络。之后机器人又靠近了墙壁, 其与最近障碍物的距离又开始减小。因此机器人与最近障碍物的距离一直在上下波动。在接下来的运动中, 动态障碍物与其他静态障碍物共同影响机器人的运动。最后, 经过 26 步, 机器人到达目标。

由图 10(b) 可见, 当遇到突然出现的障碍物时机器人会适当地放弃靠近目标。这是因为在面对非预期不确定性事件时, 中缝核对腹侧被盖区进行抑制连接, 快速降低了奖励因子对运动方向的影响。该实验进一步验证了所提出调节发育网络模型的优秀性能。

## 4 结 论

为了应对环境中的突发不确定性事件, 本文提出了一种基于神经调节系统的调节发育网络模型。设计了新的碰撞危险度指标, 该指标不仅考虑了障碍物距离对机器人运动行为的影响, 而且考虑了障碍物速度的影响。通过该碰撞危险度以及机器人探测范围

内的障碍物个数, 实现了机器人的注意力网络在关注预期刺激的背侧注意力网络与关注非预期刺激的腹侧注意力网络间的灵活切换。同时, 设计了新的中间层神经元的学习率, 改进了神经元的权值更新公式, 提高了机器人快速应对环境中的突发不确定性事件的能力。3 个仿真实验以及实际实验验证了所提出调节发育网络模型在应对环境中非预期不确定性事件的良好性能。

接下来, 将进一步完善神经调节系统的功能, 并设计更复杂的实验环境以验证模型的性能。

## 参考文献 (References)

- [1] Asadi K, Ramshankar H, Pullagurla H, et al. Vision-based integrated mobile robotic system for real-time applications in construction[J]. *Automation in Construction*, 2018, 96: 470-482.
- [2] Rahman M M, Ishii K, Noguchi N. Optimum harvesting area of convex and concave polygon field for path planning of robot combine harvester[J]. *Intelligent Service Robotics*, 2019, 12(2): 167-179.
- [3] Ha Q P, Yen L, Balaguer C. Robotic autonomous systems for earthmoving in military applications[J]. *Automation in Construction*, 2019, 107: 1-19.
- [4] Qureshi M S, Swarnkar P, Gupta S. A supervisory on-line tuned fuzzy logic based sliding mode control for robotics: An application to surgical robots[J]. *Robotics and Autonomous Systems*, 2018, 109: 68-85.
- [5] Hady G G, Abigail C D, Sebastian H, et al. ALCIDES: A novel lunar mission concept study for the demonstration of enabling technologies in deep-space exploration and human-robots interaction[J]. *Acta Astronautica*, 2018, 151: 270-283.
- [6] Wan S H, Gu Z H, Ni Q. Cognitive computing and wireless communications on the edge for healthcare service robots[J]. *Computer Communications*, 2020, 149: 99-106.
- [7] Shriyam S, Gupta S K. Incorporation of contingency tasks in task allocation for multirobot teams[J]. *IEEE Transactions on Automation Science and Engineering*, 2020, 17(2): 809-823.
- [8] Katyal K D, Hager G D, Huang C M. Intent-aware pedestrian prediction for adaptive crowd navigation[C]. *IEEE International Conference on Robotics and Automation*. Paris, 2020: 3277-3284.
- [9] Barton A, Volna E, Kotyrba M. Control of autonomous robot behavior using data filtering through adaptive resonance theory[J]. *Vietnam Journal of Computer Science*, 2018, 5(2): 85-94.
- [10] Krichmar J L. A neurobotic platform to test the influence of neuromodulatory signaling on anxious and curious behavior[J]. *Frontiers in Neurobotics*, 2013, 7: 1-17.
- [11] Dasgupta S, Wörgötter F, Manoonpong P.

- Neuromodulatory adaptive combination of correlation-based learning in cerebellum and reward-based learning in basal Ganglia for goal-directed behavior control[J]. *Frontiers in Neural Circuits*, 2013, 8: 1-21.
- [12] Showalter I, Schwartz H M. Neuromodulated multiobjective evolutionary neurocontrollers without speciation[J]. *Evolutionary Intelligence*, 2021, 14(4): 1415-1430.
- [13] Caccavale R, Finzi A. Learning attentional regulations for structured tasks execution in robotic cognitive control[J]. *Autonomous Robots*, 2019, 43(8): 2229-2243.
- [14] Krichmar J L. A neurobotic platform to test the influence of neuromodulatory signaling on anxious and curious behavior[J]. *Frontiers in Neurobotics*, 2013, 7: 1-17.
- [15] Khamassi M, Lallée S, Enel P, et al. Robot cognitive control with a neurophysiologically inspired reinforcement learning model[J]. *Frontiers in Neurobotics*, 2011, 5: 1-14.
- [16] Wang D S, Si W J, Luo Y. A biologically inspired behavior control for the unexpected uncertainty with motivated developmental network[J]. *IEEE Transactions on Cognitive and Developmental Systems*, 2020, 12(4): 774-786.
- [17] Aston-Jones G, Cohen J D. An integrative theory of locus coeruleus-norepinephrine function: Adaptive gain and optimal performance[J]. *Annual Review of Neuroscience*, 2005, 28: 403-450.
- [18] Briand L A, Gritton H, Howe W M, et al. Modulators in concert for cognition: Modulator interactions in the prefrontal cortex[J]. *Progress in Neurobiology*, 2007, 83(2): 69-91.
- [19] Krichmar J L. The neuromodulatory system: A framework for survival and adaptive behavior in a challenging world[J]. *Adaptive Behavior*, 2008, 16(6): 385-399.
- [20] Bouret S, Sara S J. Network reset: A simplified overarching theory of locus coeruleus noradrenaline function[J]. *Trends in Neurosciences*, 2005, 28(11): 574-582.
- [21] Corbetta M, Patel G, Shulman G L. The reorienting system of the human brain: From environment to theory of mind[J]. *Neuron*, 2008, 58(3): 306-324.
- [22] Yu A J, Dayan P. Uncertainty, neuromodulation, and attention[J]. *Neuron*, 2005, 46(4): 681-692.
- [23] Seo C, Guru A, Jin M, et al. Intense threat switches dorsal raphe serotonin neurons to a paradoxical operational mode[J]. *Science*, 2019, 363(6426): 538-542.
- [24] Avery M C, Krichmar J L. Neuromodulatory systems and their interactions: A review of models, theories, and experiments[J]. *Frontiers in Neural Circuits*, 2017, 11: 1-18.
- [25] Boureau Y L, Dayan P. Opponency revisited: Competition and cooperation between dopamine and serotonin[J]. *Neuropsychopharmacology*, 2011, 36(1): 74-97.
- [26] Wang D S, Duan Y H, Weng J Y. Motivated optimal developmental learning for sequential tasks without using rigid time-discounts[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2018, 29(10): 4917-4931.
- [27] Weng J. Natural and artificial intelligence: Introduction to computational brain-mind[M]. Okemos: BMI Press, 2012: 116-117.
- [28] Weng J Y, Luciw M. Dually optimal neuronal layers: Lobe component analysis[J]. *IEEE Transactions on Autonomous Mental Development*, 2009, 1(1): 68-85.
- [29] 余慧瑾, 方勇纯. 基于改进型自主发育网络的机器人场景识别方法[J]. *自动化学报*, 2021, 47(7): 1530-1538.  
(Yu H J, Fang Y C. A robot scene recognition method based on improved autonomous developmental network[J]. *Acta Automatica Sinica*, 2021, 47(7): 1530-1538.)
- [30] Weng J Y, Paslaski S, Daly J, et al. Modulation for emergent networks: Serotonin and dopamine[J]. *Neural Networks*, 2013, 41: 225-239.

### 作者简介

王东署(1973—),男,教授,博士,从事机器人智能控制、人工智能、机器学习、类脑计算和类脑芯片等研究, E-mail: wangdongshu@zzu.edu.cn;

赵红燕(1998—),女,硕士生,从事机器人智能控制、人工智能和机器学习等研究, E-mail: 136281666@qq.com.