

文章编号: 1001-0920(2012)12-1822-06

一种基于混合学习策略的移动机器人路径规划方法

郜园园, 阮晓钢, 宋洪军, 于建均

(北京工业大学 电子信息与控制工程学院, 北京 100124)

摘要: 针对未知环境下移动机器人路径规划问题, 以操作条件反射学习机制为基础, 根据模糊推理系统和学习自动机的原理, 提出一种应用于移动机器人导航的混合学习策略. 运用仿生的自组织学习方法, 通过不断与外界未知环境交互从而使机器人具有自学习和自适应的功能. 仿真结果表明, 该方法能使机器人学会避障和目标导航任务, 与传统的人工势场法相比, 能有效地克服局部极小和振荡情况.

关键词: 模糊推理系统; 学习自动机; 操作条件反射; 混合学习策略; 路径规划

中图分类号: TP242

文献标志码: A

Path planning method for mobile robot based on a hybrid learning approach

GAO Yuan-yuan, RUAN Xiao-gang, SONG Hong-jun, YU Jian-jun

(College of Electronic Information and Control Engineering, Beijing University of Technology, Beijing 100124, China.)

Correspondent: GAO Yuan-yuan, E-mail: yuangao84@163.com)

Abstract: Aim to solve the path planning problem of mobile robot in the unknown environment, a hybrid learning approach is proposed for the robot navigation based on the operant conditioning theory and the principle of fuzzy inference system and learning automata. The robot is endowed with the capabilities of self-learning and self-adapting with unknown environment by using a bionic self-organizing method. Simulation results show that, compared with the method of artificial potential field, the proposed method can make the robot learn the ability of obstacle avoidance and goal seeking without being stuck in local minima and oscillation.

Key words: fuzzy inference system; learning automata; operant conditioning; hybrid learning approach; path planning

1 引言

路径规划是当前移动机器人研究的热点之一, 它要求机器人能够自主地决定路径, 避开障碍物, 实现任务目标. 近年来, 随着移动机器人应用范围的扩大, 对于规划技术的要求也越来越高, 单一的规划方法有时不能较好地解决某些规划问题, 因而一些新的智能技术逐渐被引入路径规划中, 促使了各种方法的融合发展, 尤其是将模糊推理系统(FIS)与其他学习方法相结合的混合学习策略已成为研究路径规划方法的重点. Cang等^[1]提出将一种结合监督学习与强化学习的模糊控制器用于移动机器人避障, 但是该控制器有243条规则, 使得该控制器很难满足实时性的要求. Meng等^[2-3]将FIS与非监督学习相结合提出了一种神经模糊控制器, 并利用行为评价学习方法进行参数和

结构的学习, 可动态地自主生成、删除和调整模糊规则, 将其应用于移动机器人的路径规划, 实现了避障和沿墙走的行为. Boubertakh等^[4]结合FIS与改进的Q学习方法, 在专家知识基础上只用8条模糊规则便成功完成了移动机器人避障和目标导航任务.

上述混合学习方法虽然有诸多优点, 但均未涉及仿生学的控制机理. 若将模糊控制本身所具有的鲁棒性与基于生理学的感知-动作行为相结合, 不断通过与外界未知环境交互, 则可使机器人具有自学习和自适应的功能, 为移动机器人在复杂环境中的避障导航提出新的思路. 学习有限状态自动机(LFSA)能有效地模拟生理学上的感知-动作映射, 并且能很好地体现生物的操作条件反射学习机制, 其在机器人学中的应用研究已逐渐增多^[5-11]. Wang等^[9]运用确定性有限

收稿日期: 2011-07-07; 修回日期: 2011-11-07.

基金项目: 国家863计划项目(2007AA04Z226); 国家自然科学基金项目(61075110); 北京市自然科学基金项目(4102011); 北京市教委重点基金项目(KZ201210005001).

作者简介: 郜园园(1984-), 女, 博士生, 从事智能机器人、机器学习的研究; 阮晓钢(1958-), 男, 教授, 博士生导师, 从事机器人、模式识别等研究.

状态自动机, 结合 Q 学习, 帮助机器人进行行为选择和最优控制. 蔡建美等^[12]初步探讨了 FIS 与概率自动机的结合, 构造了一种仿生自主学习系统, 并将其应用于机器人姿态平衡控制, 体现了生物仿生的特性. 可见, 模糊逻辑法与学习自动机的结合能有效地解决移动机器人研究中的一些问题, 并具有仿生学的机理.

本文在学习自动机基础上构建了一种操作条件反射学习自动机 (OCLA), 用来模拟动物的操作条件反射机理, 并将其用于优化 FIS 的模糊规则. FIS 产生输出作用于移动机器人, 使其通过学习来达到最终感知到行为的最佳映射. 此外, 引入有效的专家知识, 提高了系统的灵活性和可扩展性. 该方法将生物学习、模糊控制与学习自动机相结合, 实现了移动机器人在未知环境下的自主避障与导航行为. 通过仿真实验分析, 该策略不仅能使机器人学会避障和目标导航, 而且与传统的人工势场法相比能有效地避免陷入局部极小和摆脱左右振荡.

2 移动机器人模糊控制器

本文设计的模糊控制器的总体框架如图 1 所示. 首先利用人类专家知识初步建立一个简单的模糊控制器; 然后模拟生物操作条件反射, 利用 OCLA 进一步调节模糊规则的参数, 通过不断与外界未知环境交互从而使机器人具有自学习和自适应的能力.

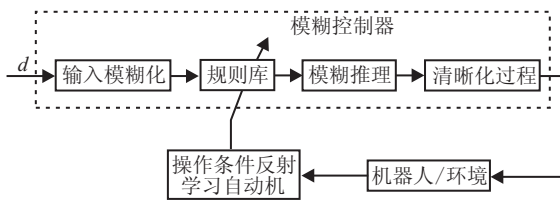


图 1 系统结构

2.1 机器人模型

仿真中采用一个圆形对称移动机器人结构模型, 其半径为 $R = 20 \text{ cm}$. 机器人有左右两个轮加前方一个万向轮, 其圆周由 24 个均匀分布的超声传感器组成. 对于每一个传感器 $s_i, i = 1, 2, \dots, 24$, 其覆盖角度为 15° , 在其范围内得到与障碍物的距离信息为 l_i . 考虑到移动机器人的对称结构和传感器均匀分布特征, 针对避障和导航设计了一个简单易行的模糊控制器, 本文简化了输入维数. 本文采用机器人前方的 18 个传感器, 每 6 个 1 组共分为 3 组作为模糊控制器的输入, 3 个方向的距离分别表示为: 左侧距离 d_L , 前方距离 d_F 和右侧距离 d_R .

$$\begin{cases} d_R = R + \min_{i=1,2,\dots,6} (l_i), \\ d_F = R + \min_{i=7,8,\dots,12} (l_i), \\ d_L = R + \min_{i=13,14,\dots,18} (l_i). \end{cases} \quad (1)$$

采用两个坐标系: 一个是用 XOY 表示的世界坐标系; 一个是用 xoy 表示的机器人的坐标系, 其中 o 是机器人的中心, y 轴直接通过机器人中心 (当机器人的后面两个轮子与 y 轴同直线时则会径直向前走). 机器人的控制变量是线速度 v 和机器人运动的转角 $\Delta\phi$. 在机器人行为控制结构中, 提出决定机器人合适的动作 v 和 $\Delta\phi$ 即可实现避障和目标导向的目的.

2.2 模糊控制器

模糊推理系统如图 2 所示, 输出 y_1 为机器人的运行速度 v, y_2 为机器人的运动转角 $\Delta\phi$. 首先对输入 $d = (d_L, d_F, d_R)$ 进行模糊化操作, 采用高斯型隶属度函数来估计每一个输入, 分别表示为 N 和 F .

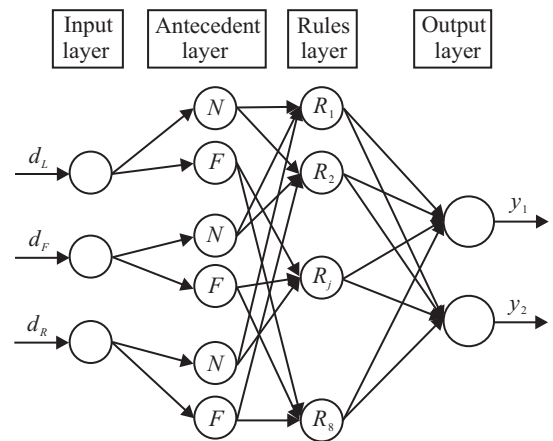


图 2 模糊推理系统

隶属度函数可描述为

$$N : \mu_k^N(t) = \begin{cases} 1, & d_k \leq d_m; \\ e^{-(1/2)((d_k(t)-d_m)/\sigma^N)^2}, & \text{else;} \end{cases} \quad (2)$$

$$F : \mu_k^F(t) = \begin{cases} e^{-(1/2)((d_k(t)-d_s)/\sigma^F)^2}, & d_k \leq d_s; \\ 1, & \text{else.} \end{cases} \quad (3)$$

其中: d_m 为机器人与障碍物可允许的最小距离, d_s 为机器人运动的安全距离, 当 $d > d_s$ 时机器人可以最大速度安全行走. N 和 F 的中心及标准差分别为 d_m 和 d_s, σ^N 和 σ^F .

根据专家知识针对避障与目标导向任务建立了模糊规则集^[4]. 这种策略可以大大减小机器人搜索空间, 使其快速地对环境做出反应. 专家知识表述如下:

1) 如果机器人在 3 个方向上距离障碍物均比较远, 则机器人直接以最大的速度向着目标导航;

2) 如果目标不在机器人前方, 但前方有障碍物存在, 则机器人向着目标方向沿距离最近的障碍物向左或向右运动, 即沿左或右障碍物行走;

3) 如果机器人前方既有目标又有障碍物存在, 则机器人试着向目标沿着距离最近的障碍物向左或向右运动, 即沿左或右障碍物行走.

为使机器人自主地决定导航模式, 设定两个参

数 p 和 q , 其具体描述如下:

$$p = \begin{cases} 0, & d_L < d_{\max}; \\ 1, & \text{else}; \end{cases} \quad (4)$$

$$q = \begin{cases} 0, & 90^\circ < \theta < 270^\circ; \\ 1, & \text{else}. \end{cases} \quad (5)$$

其中: d_{\max} 是传感器可以探测到与障碍物的最大距离, θ 是机器人与目标点的夹角. 如果 $p = 1$ ($p = 0$), 则选择沿右(左)侧墙行为; 如果 $q = 1$, 则激活目标导向行为; 如果 $q = 0$, 则选择沿墙走的行为.

采用高斯隶属度函数, 获得 3 个方向 (d_L, d_F, d_R) 的距离模糊输入信息. 模糊规则结论的符号语言为: PB (正大), PM (正中), PS (正小), ZR (零), NS (负小), NM (负中), NB (负大). 这些值由下一步所设计的操作条件反射学习获得. 由于机器人几何结构上的对称性, 可以得出 $NB = -PB$, $NM = -PM$, $NS = -PS$. 简化参数个数, 引入上述专家知识, 产生了 8 条模糊规则, 描述如下.

R1: If (d_L, d_F, d_R) is (N, N, N) Then v_1 is ZR and $\Delta\Phi_1$ is $(2 \times p - 1) \times PB$;

R2: If (d_L, d_F, d_R) is (N, N, F) Then v_2 is ZR and $\Delta\Phi_2$ is NB;

R3: If (d_L, d_F, d_R) is (N, F, N) Then v_3 is $C_v \times V_{\max}$ and $\Delta\Phi_3$ is ZR;

R4: If (d_L, d_F, d_R) is (N, F, F) Then v_4 is V_{\max} and $\Delta\Phi_4$ is $(2 \times p - 1) \times NS$;

R5: If (d_L, d_F, d_R) is (F, N, N) Then v_5 is ZR and $\Delta\Phi_5$ is PB;

R6: If (d_L, d_F, d_R) is (F, N, F) Then v_6 is ZR and $\Delta\Phi_6$ is $(2 \times p - 1) \times NB$;

R7: If (d_L, d_F, d_R) is (F, F, N) Then v_7 is V_{\max} and $\Delta\Phi_7$ is $(2 \times p - 1) \times NS$;

R8: If (d_L, d_F, d_R) is (F, F, F) Then v_8 is V_{\max} and $\Delta\Phi_8$ is $q \times \theta + (1 - q) \times (2 \times p - 1) \times PM$.

其中: V_{\max} 为机器人运行的最大速度; v_i 和 $\Delta\Phi_i$ 为模糊规则结论; C_v 为速度的衰减率.

最后, 由模糊推理系统输出的运动转角 $\Delta\Phi$ 和运行速度 v 来决定机器人下一步的动作, 本文采用了 Sugeno 模糊推理方法得出精确化后的输出

$$\Delta\Phi = \frac{\sum_{i=1}^8 \alpha_i \Delta\Phi_i}{\sum_{i=1}^8 \alpha_i}, \quad (6)$$

$$v = \frac{\sum_{i=1}^8 \alpha_i v_i}{\sum_{i=1}^8 \alpha_i}, \quad (7)$$

其中 α_i 是通过代数乘积方法所得到的每一条规则的真实值.

3 操作条件反射学习自动机

在本文所设计的仿生混合学习策略中, 通过上一节所设计的模糊控制器可以获得机器人的动作输出. 学习自动机的引入是为了进一步调节模糊控制器的规则参数, 具体调节每一条规则的 $\Delta\Phi_i$, 即调节模糊规则结论 PB, PM, PS, ZR, NS, NM 和 NB 的值. 这里基于操作条件反射学习机制设计了一种新的学习自动机, 从而使得模糊控制器具有自调节的能力.

3.1 操作条件反射基础

“操作条件反射”(OC)的概念由 Skinner^[13]首次提出, 由此创立了 Skinner 操作条件反射理论. 其核心内容为: 某一操作行为一旦受到其结果的强化, 则该行为发生的概率便会增加. 操作条件反射这一概念的特点在于, 它强调行为结果对行为的影响. 行为的结果决定有机体行为的加强或减弱以及是否重复某种行为, 从而决定有机体行为模式的形成. 这一理论阐明了有机体许多复杂行为得以形成的原因, 它把对有机体行为的解释, 从经典条件反射论的“刺激-反应”模式, 深入到了操作条件反射论的“反应-刺激-反应”加强或减弱模式.

自 20 世纪 90 年代中期开始, 许多学者致力于研究 Skinner 操作条件反射理论, 期望机器能复制动物的行为学习方式, 并建立了相关的模型^[14-18]. 以概率自动机为平台, 蔡建美等^[12]用其来模拟操作条件反射机制, 设计了相应的仿生系统, 并成功实现了两轮机器人的平衡控制.

本文在此基础上设计了一种操作条件反射学习自动机模拟动物学习, 并将其用于对模糊推理系统的规则进行进一步细化调节, 使移动机器人能够适应复杂的环境变化, 具有自主的学习能力.

3.2 OCLA 的结构设计

定义 OCLA 是一个五元组模型, 即

$$OCLA = \langle S, O, Z, \psi, \delta \rangle.$$

其中:

1) OCLA 的内部状态集合 $S = \{s_i | i = 0, 1, \dots, n_S\}$, s_i 为 OCLA 第 i 个状态符号.

2) OCLA 的内部操作集合 $O = \{o_k | k = 1, 2, \dots, n_O\}$, o_k 为 OCLA 第 k 个操作符号.

3) OCLA 的输出符号集合 $Z = \{z_m | m = 0, 1, \dots, n_Z\}$, z_m 为 OCLA 第 m 个输出符号.

4) OCLA 的状态取向函数 $\psi: S \rightarrow [h, f]$, h 定义为取向性最差的取向函数值, f 为取向性最好的取向函数值(这里的取向性是从生物学意义上定义的, 环境决定生物进化的方向, 即生物的取向性), h 和 f 的值可根据所处理的具体对象来取值. 对于任一

$s_i (s_i \in S)$, $\psi_i = \psi(s_i)$ 是 OCLA 关于状态 s_i 的期望值, 如果 $\psi_i < 0$, 则称 s_i 是负取向状态; 如果 $\psi_i = 0$, 则称 s_i 是零取向状态; 如果 $\psi_i > 0$, 则称 s_i 是正取向状态.

5) OCLA 的操作条件反射学习律 δ . 设 OCLA 在 t 时刻的状态为 $s(t) = s_a \in S$, 依集合 R 中随机“条件-操作”规则选中的操作为 $o(t) = o_c \in O$, 实施操作后观测到 $t+1$ 时刻的状态 $s(t+1) = s_d \in S$, 则基于操作条件反射原理, 操作集合 R 中随机“条件-操作”规则 $p_{ak} (k = 1, 2, \dots, n_O)$ 的激发概率依

$$\delta : \begin{cases} p_{ak}(t+1) = p_{abk}(t) - \xi(\vec{\psi}_{ac}) \cdot p_{ak}(t), \\ p_{ak}(t+1) = \max \min(p_{ak}(t+1), 0, 1), k \neq c; \\ p_{ac}(t+1) = 1 - \sum_{k \neq c} p_{ak}(t+1), k = c \end{cases}$$

进行调节. 其中 $\psi_{ac} = \psi(s_d) - \psi(s_a)$, 是 OCLA 在状态处于 $s_a (s_a \in S)$ 的条件下实施操作 $o_c (o_c \in O)$ 后状态转移为 $s_d (s_d \in S)$ 后取向函数值的变化量, 可用此变化量来判断该操作的好坏. $\xi(\psi_{ac}) = \xi(\psi(s_d) - \psi(s_a)) = \lambda * \psi_{ac} / r$ 是单调增函数, $\xi(x) = 0$ 当且仅当 $\psi_{ac} = 0$, r 是操作规则总数, λ 是学习率, 即每次迭代学习的速率. $p_{ac}(t) (a \in \{0, 1, \dots, n_S\}, c \in \{1, 2, \dots, n_O\})$ 是 OCLA 状态处于 $s_a (s_a \in S)$ 时实施操作 $o_c (o_c \in O)$ 的概率, 当 $\psi_{abc} > 0$ 时, 实施操作 o_c 并转移状态为 s_d 后的取向函数值变大, 相应地 $p_{ac}(t+1) > p_{ac}(t)$, 即下一时刻选择操作 o_c 的概率增大, 反之亦然. $\max \min(p_{ak}(t+1), 0, 1)$ 可保证 $p_{ak}(t+1) \in [0, 1]$, 且 $\sum_{k=1}^{n_O} p_{ak}(t) = 1$, 即表示在同一输入同一状态下采取不同操作的概率和为 1, 当 $t \rightarrow \infty$ 时, 若 $p_{ac}(t) \rightarrow 1$, 则说明操作 $o_c (o_c \in O)$ 在状态处于 $s_a (s_a \in S)$ 的条件下行为最优. 一般情况下给定学习迭代次数 T_f 或最优行为选择概率最大阈值 p_ε , 当学习达到迭代次数或者当某一状态处于 $s_a (s_a \in S)$ 条件下实施操作 $o_c (o_c \in O)$ 的概率 $p_{abc}(t) \geq p_\varepsilon$ 时, 停止学习; $p_\varepsilon \in [0.7, 1]$, 由实际的系统环境具体设定, 本文设 $p_\varepsilon = 0.9$.

3.3 路径规划中的 OCLA

将 OCLA 应用于上一节所设计的模糊控制器中, 并对模糊规则进行调节. OCLA 的内部状态数与模糊控制器的规则数相同, 取 $n_S = 8$, 内部操作集合为每一个规则下的结论个数, 取 $n_O = 5$. 输出是按照操作条件反射学习率所选择的模糊规则的结论. 状态取向函数简单设计如式 (8) 所示, P_{ik} 表示第 i 个规则第 k 条结论被选择的概率. 根据具体实验设定参数的值, 路径规划仿真实验中设定 $\lambda = 0.3$, $r = 5$, 从而可以得到 $\xi(\psi)$.

$$\psi = \begin{cases} -1, \min(d_R, d_F, d_L) < d_m; \\ 0, \text{else.} \end{cases} \quad (8)$$

如果概率选择的操作是 $o(t) = o_c \in O$, 对于每一条规则应用 OCLA 表述如下:

If Rule is i Then $A_i = A_{ic}$, 按照 OCLA 操作条件反射学习率 δ 改变 P_{ic} ; or $A_i = A_{ik}$, 按照 OCLA 操作条件反射学习率 δ 改变 P_{ik} . 这里 $i = 1, 2, \dots, n_S$, $k = 1, 2, \dots, n_O$, 且 $k \neq c$.

对于每一条规则下的 5 个可选结论都有对应的动作选择概率值 P_{ij} , A_{ij} 表示第 i 条规则下可供概率选择的结论 j , $j = 1, 2, \dots, n_O$, 在本文路径规划中 A_{ij} 具体的是机器人运动转角 $\Delta\phi$ 的模糊规则结论. 机器人在当前状态下按概率选择模糊规则结论, 通过取向函数进行概率更新, 以改变下一次该行为动作选择的概率来模拟生物在未知环境下动作选择的不确定性和取向性.

4 基于仿生混合学习策略的路径规划

4.1 实验条件

本文研究的目的是使移动机器人像人或动物一样通过学习具有避障和目标导向的能力, 同时还能够克服局部极小和左右振荡的情况. 本文实验中使用的机器人模型如第 2 节中所述, 这里假设超声传感器的可测范围为 $0 \sim 250$ cm. 仿真中所用的参数见表 1, 表 2 为 $\Delta\phi$ 模糊规则结论. 每一个对应 5 个模糊规则结论, 表 2 中只给出了最小值和最大值, 例如 NS 是 $-5, -10, -15, -20, -25$, 即每一条规则的内部操作集合为每一个规则下的结论个数, 即 $n_O = 5$.

表 1 实验参数

d_m/cm	d_z/cm	σ^N	σ^F	$V_{\max}/(\text{m/s})$	C_v
30	90	0.25	0.25	1	0.1

表 2 $\Delta\phi$ 模糊规则结论

	NS	NM	NB	ZR	PB
min	-25	-45	-65	0	45
max	-5	-25	-45	0	65

为方便进行实验研究, 本文建立了移动机器人基于超声传感器导航的仿真研究环境, 障碍物的类型包括圆形、方形和隔墙型障碍, 其大小和位置可以改变, 并且对于移动机器人系统而言是完全未知的. 移动机器人系统仅能通过超声传感器探测不同方向的障碍物距离, 但可以感知当前的位置信息. 移动机器人在未知环境中进行学习, 每一轮学习的运行时间为 500 步, 每一步时间 $T_s = 1$ s. 机器人与环境如图 3 所示, 其中实体部分表示未知的障碍物, “—”表示机器人运行的轨迹, 与轨迹一端相连的圆形表示移动机器人.

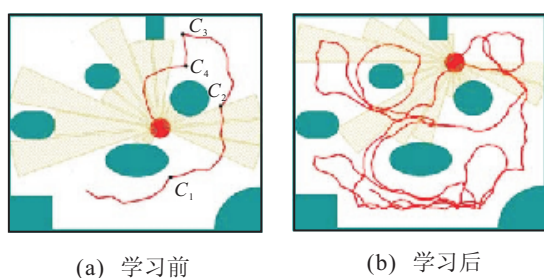


图 3 机器人自由巡航轨迹图

4.2 结果与分析

实验 1 自由巡航实验.

移动机器人从起始点开始出发,直到与障碍物发生碰撞,即与障碍物距离小于最小可允许距离 $d_m = 30\text{ cm}$. 每次碰撞后,机器人在碰撞处重新进行动作选择,直到避开障碍物继续前行. 机器人每一次动作通过操作条件反射自动机的取向函数评价后,对所选规则的概率值进行更新,以使好的行为动作选择的概率越来越大. 图 3(a) 为第 1 轮学习前 100 步时的运行轨迹,图中 C_1, C_2, C_3 和 C_4 分别是机器人与障碍物的碰撞点,很明显机器人刚开始还没有学会避碰行为. 随着学习的进行,图 3(b) 为第 3 轮的学习,机器人在这一轮已学会避碰行为,运动曲线也较为光滑.

当模糊推理系统中各规则的结论值存在任意一个选择概率大于 0.9 时,停止学习,认为当前结论达到近似最优. 学习完成后的结论值大于 0.9 的值在表 3 中列出,这一部分学习的结果与文献 [4] 中的结果相同,证明了该方法的有效性.

表 3 学习完成后的 $\Delta\phi$ 值

R1	R2	R3	R4	R5	R6	R7	R8
PB	NB	ZR	NS	PB	PB	NS	NM
45	-65	0	-10	50	55	-10	-25

实验 2 目标导航实验.

当机器人经过 3 轮学习完成后已习得避障的能力,在此基础上进行目标导航实验. 图 4 为机器人基于本文方法进行目标导航的运行轨迹图. 从图 4 可以看出,机器人能够自主判断出目标点的方向并实施行为动作,以期以较优的路径到达目标点位置. 图 4(a) 所示机器人周围远离障碍物, $p = 1$ 直接启动目标导向模式进行动作; 图 4(b) 和图 4(c) 是在图 4(a) 的环境中加入障碍物,通过与环境交互 p 值变化 1-0-1 进行模式切换来选择动作,并最终到达目标点 ($p = 1$ 目标导向模式, $p = 0$ 沿障碍物行走模式); 图 4(d) 中 p 值的变化过程为 1-0-1-0-1; 图 4(e) 是在图 4(d) 中又加入了随机的障碍物,机器人仍能另辟新径到达目标点,表现出了很好的鲁棒性与适应性,导航过程 p 值变化为 1-0-1-0-1-0-1.

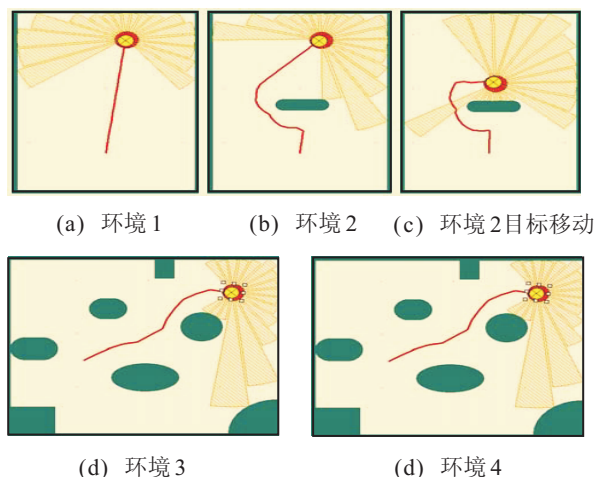


图 4 未知环境下目标导航

实验 3 克服局部极小与振荡对比实验.

仿真中布置了凹型障碍物环境和窄的通道环境与传统的人工势场法进行相关对比实验研究. 对于传统的人工势场法,在一些复杂障碍物条件下往往存在参数优化困难,无法对环境变化实现自适应;而学习的方法则可以有效地利用与环境的交互克服上述问题. 特别地,对于凹型障碍物等复杂环境,传统的人工势场法会出现局部极值,即引力与斥力相等,一些其他的方法也容易出现循环振荡.

图 5 给出了在凹型障碍物环境下和窄的通道环境下的对比实验,实验设置窄道的宽为 80 cm. 图 5(a) 为采用人工势场法的移动机器人导航轨迹,图 5(b) 为采用本文方法的移动机器人运行轨迹. 图中第 1 横排是凹形环境下的对比图. 第 2 横排为窄的通道环境下的实验效果对比图. 可以看出,本文方法在凹形环境中能够实时地识别环境,自主地改变行为选择策略, p 值变化过程是 1-0-1, 从目标导航行为变为沿障碍物走的行为,并且在通过障碍物后又切换为目标导航行为,克服了陷入局部极小的情况,成功绕开了障碍物趋向目标点. 机器人采用传统的人工势场法在通过窄道环境时出现了左右波动,而本文导航方法则在识别环境后持续选择规则 R3 进行导航,路径比较平坦,没有出现振荡,鲁棒性较好.

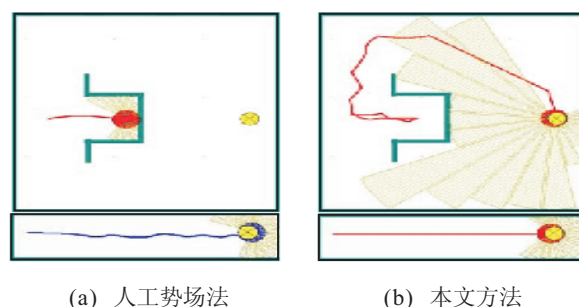


图 5 克服局部极小与振荡对比实验

5 结 论

本文对移动机器人路径规划方法进行了研究,有效地把生物学习、模糊控制与学习自动机相结合,实现了移动机器人在未知复杂环境下的自主避障与导航行为.采用仿生的混合学习策略来使机器人学习路径规划行为,通过自动获取知识、积累经验和不断更新从而降低了由对象和系统造成的不确定性影响.实验结果表明,本文方法能够使机器人学会避障、目标导航,解决了特殊环境下的局部极值和左右振荡问题,并通过与环境交互实现了自主学习,从而提高了机器人的反应式行为的灵活性和对环境的适应性.因为所研究的是一种拟人的学习机制,是以提高机器人的避障能力为主而效果作为其次,故在下一步研究中还需进一步地优化.

参考文献(References)

- [1] Cang Y, Yung N H C, Danwei Wang. A fuzzy controller with supervised learning assisted reinforcement learning algorithm for obstacle avoidance[J]. *IEEE Trans on Systems, Man, and Cybernetics, Part B: Cybernetics*, 2003, 33(1): 17-27.
- [2] Meng Joo Er, Chang Deng. Obstacle avoidance of a mobile robot using hybrid learning approach[J]. *IEEE Trans on Industrial Electronics*, 2005, 52(3): 898-905.
- [3] Meng Joo Er, Yi Zhou. Automatic generation of fuzzy inference systems via unsupervised learning[J]. *Neural Networks*, 2008, 21(10): 1556-1566.
- [4] Boubertakh H, Tadjine M, Glorennec P Y. A new mobile robot navigation method using fuzzy logic and a modified Q-learning algorithm[J]. *J of Intelligent and Fuzzy Systems*, 2010, 21(1/2): 113-119.
- [5] Narendra K S, Thathchar M A L. Learning automat: An introduction[M]. Upper Saddle River: Prentice-Hall, 1989.
- [6] Pierce D, Kuipers B. Learning to explore and build maps[C]. *Proc of the National Conf on Artificial Intelligence*. Seattle: MIT Press, 1994: 264-1271.
- [7] Dean T, Anqluin D, Basye K, et al. Inferring finite automata with stochastic output functions and an application to map learning[J]. *Machine Learning*, 1995, 18(1): 81-108.
- [8] El-Osery A, Jamshidi M. A stochastic learning automaton based autonomous control of robotic agents[C]. *IEEE Int Conf on Systems, Man and Cybernetics*. Yasmine Hammamet: IEEE Press, 2002: 100-105.
- [9] Wang X, Ray A, Lee P, et al. Optimal control of robot behaviour using language measure[J]. *Int J of Vehicle Autonomous Systems*, 2004, 2(3/4): 147-167.
- [10] Cattinelli I, Goldwurm M, Borghese N. Interacting with an artificial partner: Modeling the role of emotional aspects[J]. *Biological Cybernetics*, 2008, 99(6): 473-489.
- [11] 闫保英, 赵合计. 机器人足球中角色决策的有限状态自动机模型[J]. *山东轻工业学院学报: 自然科学版*, 2008, 22(2): 98-101.
(Yan B Y, Zhao H J. The finite state automation model of the decision-making of robocup based on role[J]. *J of Shandong Institute of Light Industry: Natural Science Edition*, 2008, 22(2): 98-101.)
- [12] 蔡建美, 阮晓钢. OCPA仿生自主学习系统及在机器人姿态平衡控制上的应用[J]. *模式识别与人工智能*, 2011, 24(1): 138-146.
(Cai J X, Ruan X G. OCPA bionic autonomous learning system and its application to robot poster balance control[J]. *Pattern Recognition and Artificial Intelligence*, 2011, 24(1): 138-146.)
- [13] Skinner B F. *The behavior of organisms*[M]. New York: Appleton-Century-Crofts, 1938.
- [14] Touretzky D S, Saksida L M. Operant conditioning in skinnerbots[J]. *Adaptive Behavior*, 1997, 5(3/4): 219-247.
- [15] Saksida L M, Raymond S M, Touretzky D S. Shaping robot behavior using principles from instrumental conditioning[J]. *Robotics and Autonomous Systems*, 1998, 22(3/4): 231-249.
- [16] Touretzky D S, Daw N D, Tira-Thompson E J. Combining configural and TD learning on a robot[C]. *The 2nd Int Conf on Development and Learning*. Cambridge, 2002: 47-52.
- [17] Gaudiano P, Chang C. Adaptive obstacle avoidance with a neural network for operant conditioning: Experiments with real robots[C]. *IEEE Int Symposium on Computational Intelligence in Robotics and Automation*. Monterey: IEEE Press, 1997: 13-18.
- [18] Itoh K, Miwa H, Matsumoto M, et al. Behavior model of humanoid robots based on operant conditioning[C]. *The 5th IEEE-RAS Int Conf on Humanoid Robots*. Tsukuba: IEEE Press, 2005: 220-225.