

文章编号: 1001-0920(2009)04-0532-05

进化强化学习及其在机器人路径跟踪中的应用

段 勇¹, 崔宝侠¹, 徐心和²

(1. 沈阳工业大学 信息科学与工程学院, 沈阳 110178; 2. 东北大学 信息科学与工程学院, 沈阳 110004)

摘 要: 研究了一种基于自适应启发评价(AHC)强化学习的移动机器人路径跟踪控制方法. AHC 的评价单元(ACE)采用多层前向神经网络来实现,将 TD()算法和梯度下降法相结合来更新神经网络的权重. AHC 的动作选择单元(ASE)由遗传算法优化的模糊推理系统(FIS)构成. ACE 网络的输出构成二次强化信号,用于指导 ASE 的学习. 最后将所提出的算法应用于移动机器人的行为学习,较好地解决了机器人的复杂路径跟踪问题.

关键词: 强化学习; 自适应启发评价; 遗传算法; 路径跟踪

中图分类号: TP242

文献标识码: A

Evolutionary reinforcement learning and its application in robot path tracking

DUAN Yong¹, CUI Baoshia¹, XU Xinde²

(1. School of Information Science and Engineering, Shenyang University of Technology, Shenyang 110178, China;

2. College of Information Science and Engineering, Northeastern University, Shenyang 110004, China.

Correspondent: DUAN Yong, E-mail: duanyong0607@126.com)

Abstract: The control policy of robot path-tracking based on adaptive heuristic critic(AHC) reinforcement learning is researched. The adaptive critic element(ACE) of AHC is composed of a multi-layer feedforward network. TD() algorithm and gradient descent algorithm are integrated, which is used to update the weights of network. The output of the ACE generates the secondary reinforcement signal which can direct the learning of the action select element(ASE). ASE can be implemented by the fuzzy inference system(FIS) which is optimized by using the genetic algorithms. Finally, the method is used for learning the robot behavior. The experiment shows that the scheme can effectively solve the problem of the robot path-tracking.

Key words: Reinforcement learning; AHC; Genetic algorithm; Path tracking

1 引 言

路径跟踪是移动机器人控制的基本问题之一,其目的是使机器人按照已知路径或局部感知路径运动,通常用于路径规划后的执行阶段.文献[1]研究了一种由前馈和反馈两个单元构成的路径跟踪控制器,前馈部分由路径参考点决定,反馈部分为设计的优化控制器.文献[2]使用模糊小波神经网络设计了自适应控制策略,用于解决路径跟踪问题.[3]研究了基于模糊逻辑的路径跟踪方法,其模糊规则来自于人类的驾驶行为.[4]研究了基于数学模型的路径跟踪方法,并使用神经网络进行位置误差补偿.

由于移动机器人动力学模型的复杂性和工作环

境的不确定性,且环境知识获取困难,采用传统的基于模型和专家知识的控制方法往往难以取得良好的路径跟踪控制效果.强化学习(RL)具有不依赖于环境模型、不需要先验知识以及鲁棒性等优点,为移动机器人路径跟踪的研究提供了新的思路.

强化学习是指 agent 从环境状态到动作映射的学习,以使动作从环境中获得的累积强化信号(回报)最大.自适应启发评价(AHC)方法是一种重要的强化学习算法,AHC 学习系统通常由自适应评价单元(ACE)和动作选择单元(ASE)组成. Sutton 在 Barto 等人研究工作的基础上提出了 AHC 算法,他采用两个神经网络实现评价单元和动作选择单元构

收稿日期: 2008-04-03; 修回日期: 2008-08-12.

基金项目: 国家自然科学基金项目(60475036).

作者简介: 段勇(1978—),男,沈阳人,讲师,博士,从事智能机器人、机器学习的研究;徐心和(1940—),男,河北山海关人,教授,博士生导师,从事计算机博弈、智能机器人等研究.

成具有 AHC 结构的控制系统. 此后, 一些学者对 AHC 学习算法作了进一步的研究, 将其扩展到解决连续状态空间的强化学习问题, 并扩大了 AHC 的应用领域^[5-7].

本文使用前向神经网络实现 AHC 强化学习算法的 ACE 单元. 神经网络的权值学习使用梯度下降算法, 通过瞬时差分(TD)预测方法得到差分梯度的迭代形式. 该方法不但可以有效地解决强化学习的时间信度分配问题, 而且提高了神经网络的学习速度. 为了有效地解决强化学习的维数灾难问题, 使用模糊推理系统(FIS)构成 AHC 的动作选择单元 ASE, 通过 FIS 来实现状态空间到动作空间的映射. 然而, 由于工作环境的复杂性以及专家知识的不足, FIS 的模糊规则难以设计. 这里使用遗传算法(GA)来学习模糊规则, 根据 ACE 网络的输出确定 GA 的适应度函数, 从而通过进化取得最优的模糊规则. 最后将本文算法应用于移动机器人路径跟踪的控制中, 使机器人在无教师指导的情况下, 通过自身与环境的交互来学习行为能力, 从而完成路径跟踪的任务.

2 AHC 学习原理及结构

AHC 算法的原理是学习系统从环境中获取状态信号和原始强化信号, 然后根据状态信号和原始强化信号以某种搜索策略选择动作, 再由执行单元执行该动作并作用于环境. 学习单元用来更新执行单元和评价单元的内容, 而评价单元将原始强化信号转化为更利于学习系统使用的二次强化信号.

原始强化信号可能含有噪声, 也可能是延时的, 而且外部原始强化信号通常是根据经验定义的, 必然会存在不精确或不完善. 为了提高学习系统的性能, 需要构造一个二次强化信号, 相对于原始强化信号, 具有立即性、完善性、独立性和可靠性等优点^[5-7]. 此外, 学习系统在外部的强化信号不精确或延时出现时仍然能够指导学习, 因此二次强化信号更利于系统学习.

AHC 系统的基本构成如图 1 所示. ACE(预测单元)的作用是根据外部的原始强化信号 r_t 及当前状态信号 s_t 来对候选的动作进行评价, 其输出为 $V(s_t)$, 从而构成内部的二次强化信号 \hat{r}_t . 在执行某一选择动作时, 系统转移到新状态, ACE 单元的输出可用来判断策略的优劣. ASE(动作单元)主要用来生成输出动作, 通过执行改动作, 使环境状态发生转变, 并同时获得来自于环境的原始外部强化信号 r_t . 动作的选择策略由系统环境状态 s_t 和二次强化信号 \hat{r}_t 决定.

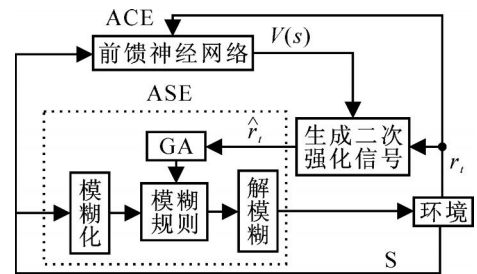


图 1 AHC 强化学习系统结构

2.1 ACE 网络学习算法

为了解决信度分配问题, ACE 采用 3 层前向神经网络来实现, 根据增量式多步瞬时差分的误差反向传播算法来推导神经网络各层的权值. 设网络有 n 个输入节点, 隐层有 m 个神经元, 输出层有 1 个神经元. 网络的输出神经元同时与输入节点和隐层神经元连接. s_1, s_2, \dots, s_n 为神经网络输入, y_1, y_2, \dots, y_m 为隐层神经元输出, $V(s)$ 为网络输出. a_{ij} 为输入节点和隐层神经元的连接权, b_i 为输入节点与输出神经元的连接权, c_j 为隐层神经元与输出神经元的连接权. 下面提出基于多步瞬时差分的神经网络学习算法.

令 $\sum_{i=1}^n a_{ij}(t) s_i(t) = 1$, 则隐层神经元输出为

$$y_j(t) = f \left[\sum_{i=1}^n a_{ij}(t) s_i(t) \right] = f(x_1).$$

其中 $f(\cdot)$ 为 sigmoid 函数, 定义为

$$f(x) = \left[\frac{1 - \exp(-x)}{1 + \exp(-x)} \right] = \frac{2}{1 + \exp(-x)}, \quad (1)$$

和 用于调整 sigmoid 函数的动态范围和斜率.

令

$$\sum_{i=1}^n b_i(t) s_i(t) + \sum_{j=1}^m c_j(t) y_j(t) = 2,$$

则有

$$V(s_t) = f \left[\sum_{i=1}^n b_i(t) s_i(t) + \sum_{j=1}^m c_j(t) y_j(t) \right] = f(x_2). \quad (2)$$

基于 TD 方法构成二次强化信号如下:

$$\hat{r}_t = r_t + \gamma [V(s_{t+1}) - V(s_t)], \quad (3)$$

其中 γ (0 < γ < 1) 为强化学习的折扣率, 它决定将来的强化信号对当前学习的作用.

下面讨论如何根据误差反向传播算法来学习神经网络权值. 首先使用 TD 算法定义神经网络的误差代价函数, 即

$$E = \frac{1}{2} [r_t + \gamma [V(s_{t+1}) - V(s_t)] - V(s_t)]^2 = \frac{1}{2} \hat{r}_t. \quad (4)$$

按照梯度下降算法, 连接权增量可计算为

$$w(t) = - \frac{\partial E}{\partial w(t)} = \gamma_t \frac{\partial V(s_t)}{\partial w(t)}, \quad (5)$$

其中 γ_t 为学习率。

根据增量式多步瞬时差分 TD()^[5] 算法,在非监督学习的条件下,需要根据相继的预测值之差来整定权值。在学习过程中,考虑到较近的预测值对权值的调整贡献较大,而较远的预测贡献较小,将 TD() 算法引入梯度下降算法,采用指数加权的形式计算梯度。

$$w(t) = \gamma_t \sum_{k=1}^t \gamma^{t-k} \nabla_w V(s_k). \quad (6)$$

其中: $\nabla_w V(s)$ 是对 w 的偏导数; γ 为 TD() 算法的多步折扣因子,采用指数加权的形式,对以前出现的预测调整赋予权重。

令

$$\begin{aligned} w(t) &= \gamma_t \sum_{k=1}^t \gamma^{t-k} \nabla_w V(s_k) = \\ &\nabla_w V(s_t) + \gamma \sum_{k=1}^{t-1} \gamma^{t-1-k} \nabla_w V(s_k) = \\ &\nabla_w V(s_t) + \gamma (t-1), \end{aligned}$$

所以对网络输出层的权值调整为

$$\begin{aligned} \nabla_{c_j(t)} V(s_t) &= \frac{\partial V(s_t)}{\partial c_j(t)} = f'(z) y_j(t), \\ c_j(t) &= \gamma_t [\nabla_{c_j(t)} V(s_t) + \gamma (t-1)], \\ c_j(t) &= c_j(t-1) + \gamma c_j(t). \end{aligned} \quad (7)$$

输入节点和隐层神经元的连接权通过误差反向传播可以得到如下修正公式:

$$\begin{aligned} \nabla_{a_{ij(t)}} V(s_t) &= \frac{\partial V(s_t)}{\partial a_{ij(t)}} = \frac{\partial V(s_t)}{\partial y_j(t)} \frac{\partial y_j(t)}{\partial a_{ij(t)}} = \\ &f'(z) c_j(t) f'(x) s_i(t), \\ a_{ij}(t) &= \gamma_t [\nabla_{a_{ij(t)}} V(s_t) + \gamma (t-1)], \\ a_{ij}(t) &= a_{ij}(t-1) + \gamma a_{ij}(t). \end{aligned} \quad (8)$$

为输入节点与输出神经元的连接权的整定公式表示为

$$\begin{aligned} \nabla_{b_i(t)} V(s_t) &= \frac{\partial V(s_t)}{\partial b_i(t)} = f'(z) s_i(t), \\ b_i(t) &= \gamma_t [\nabla_{b_i(t)} V(s_t) + \gamma (t-1)], \\ b_i(t) &= b_i(t-1) + \gamma b_i(t). \end{aligned} \quad (9)$$

以上为 ACE 单元神经网络的误差反向传播算法推导过程,根据式(3)得到 AHC 学习系统的二次强化函数。根据式(7)~(9)可分别计算网络各层的节点权值,从而完成神经网络的学习。

2.2 ASE 学习算法

ASE 由 FIS 构成, FIS 的输入为环境状态 s_t , 输出为强化学习算法的动作变量 a_t , 模糊规则通过 GA 进化得到。GA 种群中的每个染色体个体表示一

个模糊规则库。GA 每一代种群中的优秀个体构成本次学习的获胜模糊规则库,通过模糊推理得到输出动作,该动作作用于环境使状态转移到 s_{t+1} , 并从环境得到原始外部强化信号 r_t 。

Goldberg 提出一种变长度染色体编码的遗传算法(MGA)^[8],该方法适合解决在进化过程中染色体长度改变的问题。下面举例来说明基于 MGA 的染色体编码方式。将 FIS 的模糊规则库进行变长度编码,每个模糊规则库表示成 MGA 种群中的一个染色体,其中的每条规则为染色体中的基因。FIS 的模糊规则库中模糊规则 k 可以表示为

R_k : If x_1 is small and x_2 is large then y is large, 其中 x_1, x_2 和 y 均被模糊化为语言变量 {small, mean, large}, 编号为 {1, 2, 3}。将输入/输出变量 $\{x_1, x_2, y\}$ 也编号为 {1, 2, 3}。基于 MGA 的编码思想,每条模糊规则可以使用一系列二元组来表示,二元组的第 1 位表示输入/输出变量的编号,第 2 位表示相应的语言变量编号,例如 (1, 1) 表示 x_1 is small。所以以上的模糊规则 R_k 的变长度染色体编码可以表示为 {(1, 1), (2, 3), (3, 3)}, 其中的二元组排列顺序可以任意改变。

MGA 用来优化模糊规则,即在给定模糊规则前提的条件下,通过进化来得到与前提部分最匹配的结论部分。为了加快进化过程,将整个种群分为若干个子种群,每一个子种群用来进化一条模糊规则,子种群中的染色体具有相同的前提部分和各不相同的结论部分。这样在完成进化过程后,选择每个子种群中的最优染色体个体就可以得到最优的模糊规则库。在进化过程中,根据输入变量激活相应的子种群,然后按照贪婪策略来选择子种群中的个体作为当前的模糊规则。MGA 由于编码长度可变,其遗传操作使用切断算子和拼接算子进行遗传操作^[8]。

遗传算法优化 FIS 的目的是确定从状态空间到动作空间的最佳映射,ACE 网络的输出 $V(s_t)$ 是 ASE 所选择的动作的评价,因而遗传算法的适应度函数由二次强化信号构成(见式 3)。定义适应度函数为

$$\text{fitness}(t) = k / |\gamma(t)|. \quad (10)$$

这样定义表示越小的二次强化信号对应的 ASE 输出动作具有越高的适应值。这是因为从式(3)可以知道,二次强化信号 γ_t 表示对相邻时刻输出动作的评估误差,显然该误差越小越好,即 ASE 的动作越优,其评估误差也越小。在极限的情况下, $|\gamma(t)| = 0$ 可能出现,此时式(10)无意义。因此进行如下定义:当 $|\gamma(t)| < \epsilon$ (ϵ 为足够小的正数)时,取适应值 $\text{fitness}(t)$ 为之前进化过程中出现过的最大适应值。

由于适应度是由选择的多个染色体所表示的模糊规则共同作用的结果,应将适应度分配给模糊推理所激活的每条模糊规则.子种群 i 中染色体 k 的适应度可以计算为

$$f_k^i = \text{fitness}^i \left(\prod_{l=1}^n \mu_l^i \right), \quad (11)$$

其中 μ_k^i 为模糊规则的模糊隶属度.

综上所述,ACE单元和ASE单元的交互学习算法可以描述如下:首先,随机生成MGA的初始种群和ACE神经网络的连接权;之后,将环境状态 s_t 同时输入ACE网络和ASE的FIS,ACE网络通过前向计算得到输出值 $V(s_t)$;选择MGA种群的个体作为模糊规则,并通过模糊推理得到输出动作 a_t ,该动作使状态转移到 s_{t+1} ,并得到原始外部强化信号 r_t ;将状态 s_{t+1} 输入ACE网络得到输出 $V(s_{t+1})$,根据式(3)得到二次强化信号 r_t ,该值体现对动作 a_t 的评价;之后,由式(4)~(9)通过误差反向传播来更新网络连接权,同时由 r_t 得到个体的适应度;重复该过程得到MGA每个染色体适应度,根据适应度进行遗传操作,进化生成下一代种群.在学习结束时选择每个子种群中适应值最高的染色体个体作为FIS的模糊规则.

3 基于AHC学习的机器人路径跟踪

智能移动机器人是一类能够通过传感器感知环境和自身状态,实现自主导航,从而完成一定作业功能的系统^[9,10].对于机器人路径跟踪问题,主要是建立基于机器人系统状态与参考路径之间的关系^[11].此关系给出了参考路径上的目标点(虚拟目标点),然后使机器人跟踪该目标点.图2显示了控制机器人按照路径 $S_1 S_2$ 运动的示意图.其中: (x_R, y_R) 为机器人的当前位姿, P 点为机器人到路径的最近点, d 为机器人质心到路径的距离, θ 为路径曲线在 P 点的切线角, Q 点为相对于机器人跟踪的路径参考点, α 为路径曲线在 Q 点的切线角, L 为机器人质心到路径参考点的距离, β 为 PQ 的曲线长度, γ 为机器人方向角与路径参考点切线角之差, $\gamma = \theta - \beta$.

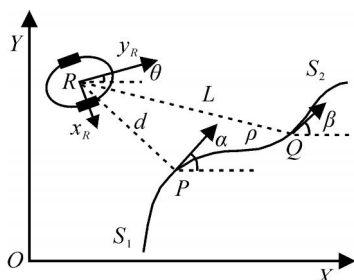


图2 移动机器人路径跟踪示意图

下面讨论使用提出的AHC强化学习方法来实

现机器人路径跟踪行为的学习.由于路径跟踪根据对参考目标点的跟踪来减小机器人与路径的距离误差和角度误差,选择机器人当前位置与虚拟目标点的距离误差 L , 角度误差 γ 以及机器人与路径曲线的距离 d 作为强化学习系统状态矢量.自主机器人采用左右轮独立驱动方式,将机器人运动的控制变量线速度 v_c 和角速度 ω_c 作为强化学习的动作变量.当 PQ 的曲线长度 L 过大时,意味着机器人到路径参考点的前瞻距离较大,此时可以实现机器人的平滑控制并有较快的跟踪速度,但其路径跟踪的精度较差.如果 PQ 的曲线长度 L 过小,则可以减小跟踪误差,但根据速度较慢甚至会出现控制的不稳定.因此将曲线长度 L 作为强化学习的另外一个强化学习的动作变量,使其能根据环境状态信息来自适应地选择路径的参考点.综上所述,可以得到AHC强化学习系统的状态空间到动作空间的映射: (L, d, γ) $\rightarrow (v_c, \omega_c)$.

对于ASE中的FIS单元,将状态变量和动作变量分别在论域上进行模糊划分成模糊集合并用模糊语言变量表示,状态(输入)和动作(输出)变量均使用三角形隶属度函数.状态变量 L 和 d 分别划分为3,5,3个模糊集合,动作变量 v_c , ω_c 和 γ 分别划分为3,4,3个模糊集合,由状态变量和动作变量构成的模糊规则表示为MGA染色体.开始时种群中染色体的个数为零,即初始的模糊规则库没有任何模糊规则.当机器人进行行为学习时,获取相应的状态变量,进而可以得到状态分量对应的模糊规则前提部分,然后为种群中添加表示模糊规则的染色体,并建立相应的子种群.随着机器人对环境的探索,新的染色体被不断添加.每个子种群的规模为 M , 它们有相同的前提部分,但结论部分各不相同.在每代的进化中,随着遗传操作产生优秀个体的加入,遗弃适应度较低的个体,以保持种群规模不变.在AHC学习结束后,选择MGA中具有最优适应值的染色体作为路径跟踪行为的控制器.

4 实验结果及分析

根据以上分析,进行机器人路径跟踪实验.对于路径跟踪行为,控制目的是使距离误差 d 和角度误差 γ 尽可能小,所以定义强化信号函数如下:

$$r_t = \begin{cases} C, & |d_{t_1}| < |d_{t_0}| \text{ and } |\gamma_{t_1}| < |\gamma_{t_0}|; \\ d, & |d_{t_1}| < |d_{t_0}|; \\ 0, & \text{otherwise.} \end{cases} \quad (12)$$

其中: C 表示正的常数; d 为常数, $d > 0$ 表示回报比例系数; d_{t_0} , γ_{t_0} , d_{t_1} 和 γ_{t_1} 分别表示当前时刻以及

下一时刻的距离误差和角度误差. 这里没有对强化信号函数进行详尽的、完善的定义, 但由于 AHC 强化学习系统可以将外部原始强化信号转化为更完善的、包含更多信息量的二次强化信号, 在具体实现算法的过程中, 相关参数定义如下: 学习率为 0.3, MGA 的切断概率 p_c 为 0.85, 变异概率 p_m 为 0.05.

为了减少移动机器人路径跟踪的学习时间和机器人探索环境的消耗, 在仿真环境中进行 AHC 强化学习, 然后使用 Pioneer II 移动机器人在真实环境测试学习结果. 将机器人置于含有变化曲率路径的环境中进行学习, 当机器人完成固定数目的学习次数后回到起始点重新开始新一轮的学习, 然后在改变机器人的起始点继续进行学习, 直到学习结束.

图 3 显示了种群在进化过程中历代最大个体适应度和最小个体适应度的变化曲线. 可见当进化到一定代数后, 算法逐渐收敛, 个体的最大适应值到达稳定, 此时能够得到优化的模糊规则库. 在学习结束后, 使用 ASE 单元中优化后的 FIS 作为机器人路径跟踪控制器. 使用学习的控制器控制机器人进行路径跟踪, 路径曲线方程为 $x(t) = 1000 \sin(t)$, $y(t) = 1200 \cos(t)$.

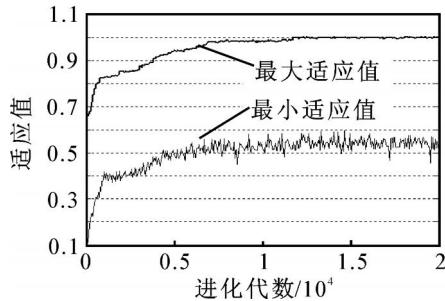


图3 适应值进化曲线

图 4 为在跟踪该曲线过程中距离误差和角度误差的规一化变化曲线. 为了验证算法的适应性, 使用控制器控制机器人完成另一变曲率曲线路径的跟踪, 其方程为

$$x(t) = 1200 \sin(t), y(t) = 1000 \cos(t/2).$$

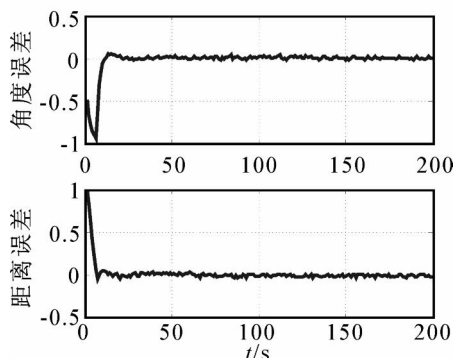


图4 误差变化曲线

图 5 显示了跟踪轨迹. 图中虚点线为待跟踪的曲线, 实线为机器人运行轨迹. 从以上实验结果可见, 本文的方法可以较好地实现机器人路径跟踪行为的学习, 能够较好地完成对变化曲率曲线的跟踪任务, 并且能够以较快的速度到达稳定. 此外, 在 Pioneer II 机器人的导航实验中, 使用研究结果对规划的路径进行了跟踪, 较好地实现了机器人对复杂路径的跟踪.

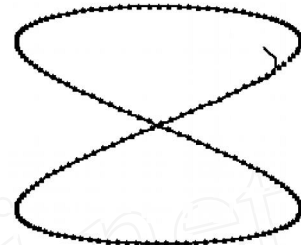


图5 路径跟踪轨迹

5 结 论

本文对 AHC 强化学习方法进行了研究, AHC 的 ACE 单元使用一个多层前馈神经网络来实现, 将 TD() 算法和梯度下降法相结合来更新神经网络的权值. AHC 的 ASE 单元通过 FIS 来实现状态空间到动作空间的映射, 利用 MGA 算法来优化 FIS 的模糊规则. 模糊规则编码成 MGA 的染色体, 通过进化学习来设计优化的 FIS. 与基本的 AHC 强化学习相比, 该方法有效地提高了 ACE 单元的学习速度, 同时 ASE 单元通过解模糊可以输出连续的动作, 从而提高了状态空间到动作空间映射的精度. 此外, FIS 使学习系统易于引入有效的专家知识, 可对学习结果进行必要的补充和完善, 因而提高了系统的灵活性和可扩展性. 使用 AHC 方法来使机器人学习路径跟踪行为, 学习得到的模糊规则库作为机器人路径跟踪的控制器, 实验结果表明该方法能够较好地解决移动机器人的路径跟踪问题. 基于强化学习的控制方法不依赖于精确的环境模型和先验知识, 通过与环境的交互完成非监督学习, 从而提高了机器人反应式行为的灵活性和对环境的适应性.

参考文献(References)

- [1] Zhu Z X, Chen J, Yoshida Toyofumi, et al. Path tracking control of autonomous agricultural mobile robots[J]. J of Zhejiang University: Science A, 2007, 8(10): 1596-1603.
- [2] Oh J S, Park J B, Choi Y H. Stable path tracking control of a mobile robot using a wavelet based fuzzy neural network[J]. Int J of Control, Automation and Systems, 2005, 3(4): 552-563.

(下转第 541 页)

学反应机理,优化控制锌粉的添加量,保证第 2 段出口钴离子浓度符合工艺要求,并且降低锌粉的消耗量.

参考文献(References)

- [1] Boyanov B S, Konareva V V, Kolev N K. Purification of zinc sulfate solutions from cobalt and nickel through activated cementation[J]. Hydrometallurgy, 2004, 73(2): 163-168.
- [2] 桂卫华, 李勇刚, 阳春华, 等. 基于改进聚类算法的分布式 SVM 及其应用[J]. 控制与决策, 2004, 19(8): 852-856.
(Gui W H, Li Y G, Yang C H, et al. Distributed SVM based on improved clustering algorithm and its application[J]. Control and Decision, 2004, 19(8): 852-856.)
- [3] Wu C H, Ho J M, Lee D T. Travel-time prediction with support vector regression [J]. IEEE Trans on Intelligent Transportation Systems, 2004, 5(4): 276-281.
- [4] Davy M, Desobry F, Gretton A, et al. An online support vector machine for abnormal events detection [J]. Signal Processing, 2006, 86(8): 2009-2025.
- [5] Lau K W, Wu Q H. Online training of support vector classifier[J]. Pattern Recognition, 2003, 36(8): 1913-1920.
- [6] 张英, 苏宏业, 褚健. 基于 ISVM 的软测量建模及其在 PX 生产中的应用研究[J]. 控制与决策, 2005, 20(10): 1102-1106.
(Zhang Y, Su H Y, Chu J. Soft sensor modeling based on ISVM and its application in PX fractionation [J]. Control and Decision, 2005, 20(10): 1102-1106.)
- [7] Cauwenberghs G, Poggio T. Incremental and decremental support vector machine learning [J]. Machine Learning, 2001, 44(13): 409-415.
- [8] Ma J, Theiler J, Perkins S. Accurate online support vector regression [J]. Neural Computation, 2003, 15(11): 2683-2704.
- [9] Omitaomu O A, Jeong M K, Badiru A B, et al. Online support vector regression approach for the monitoring of motor shaft misalignment and feedwater flow rate [J]. IEEE Trans on Systems, Man and Cybernetics — Part C: Applications and Reviews, 2007, 37(5): 962-970.
- [10] 刁翔, 李奇. 基于加权支持向量回归的在线训练算法及应用[J]. 系统仿真学报, 2007, 19(17): 3970-3973.
(Diao X, Li Q. On-line training algorithm and its application based on weighted SVR [J]. J of System Simulation, 2007, 19(17): 3970-3973.)
- [11] 汪辉, 皮道映, 孙优贤. 支持向量机在线训练算法及其应用[J]. 浙江大学学报, 2004, 38(12): 1642-1649.
(Wang H, Pi D Y, Sun Y X. On-line support vector machine training algorithm and its application [J]. J of Zhejiang University, 2004, 38(12): 1642-1649.)
- [12] 刘志宏, 唐朝波, 张多默, 等. 铋盐除钴净化工艺研究 [J]. 中南工业大学学报, 2000, 31(3): 225-227.
(Liu Z H, Tang C B, Zhang D M, et al. Study on the purifying processes for zinc leach solutions with antimony trioxides [J]. J of Central South University of Technology, 2000, 31(3): 225-227.)
- [3] Antonelli Gianluca, Chiaverini Stefano, Fusco Giuseppe. A fuzzy-logic-based approach for mobile robot path tracking [J]. IEEE Trans on Fuzzy Systems, 2007, 15(2): 211-221.
- [4] Chaitanya S K, Sarkar P K. A neural network algorithm for the error optimization in the path tracking control of a mobile robot [C]. IEEE Int Conf on Neural Networks. 2006: 2501-2507.
- [5] Sutton R S, Barto A G. Reinforcement learning: An introduction [M]. Cambridge: MIT Press, 1998.
- [6] Su S F, Hsieh S H. Embedding fuzzy mechanisms and knowledge in Box-Type reinforcement learning controllers [J]. IEEE Trans on Systems, Man and Cybernetics — Part B, 2002, 32(5): 645-653.
- [7] Zeybek Z. Role of adaptive heuristic criticism in cascade temperature control of an industrial tubular furnace [J]. Applied Thermal Engineering, 2006, 26(2/3): 152-160.
- [8] Hoffmann F, Pfister G. Evolutionary design of a fuzzy knowledge base for a mobile robot [J]. Int J of Approximate Reasoning, 1997, 17(4): 447-469.
- [9] 石鸿雁, 孙昌志, 陈冬阳, 等. 动态环境下自主移动机器人的导航复杂性 [J]. 沈阳工业大学学报, 2006, 28(5): 534-537.
(Shi H Y, Sun C Z, Chen D Y, et al. Research on navigation complexity for autonomous mobile robots under dynamic environments [J]. J of Shenyang University of Technology, 2006, 28(5): 534-537.)
- [10] 许巍丽, 孙茂相. 全方位移动机器人鲁棒控制 [J]. 沈阳工业大学学报, 2007, 29(3): 312-316.
(Xu W L, Sun M X. Robust control of omni-direction wheeled mobile robot [J]. J of Shenyang University of Technology, 2007, 29(3): 312-316.)
- [11] Koh K C, Cho H S. A smooth path tracking algorithm for wheeled mobile robots with dynamic constraints [J]. J of Intelligent and Robotic Systems, 1999, 24(4): 367-385.

(上接第 536 页)