

基于 Markov 对策的机械臂二维路径规划

陈 魁, 刘久富, 苏青琴, 刘 蓉

(南京航空航天大学 自动化学院, 江苏 南京 210016)

摘 要: 针对机械臂应用环境状况较复杂、不确定条件较多, 文中使用基于 Markov 对策的算法对二维机械臂进行路径规划。二维机械臂路径规划是三维多关节机器人规划的基础。首先根据实际的工作环境设定机械臂的运动范围并选择经常出现的动作组合作为机械臂运动的基本行为集, 给出各种情况可能获得的报酬, 依据多智能体 Q 值学习算法更新每个关节的报酬值, 反解出对应最大报酬值的动作组合。文中仿真绘制最佳动作组合时的运动轨迹, 分别仿真绘制机械臂运动环境中无障碍与放置圆形障碍物时的二维运动轨迹, 并确定轨迹的误差。

关键词: 多关节机器人; 机械臂; 多 Agent 系统; Markov 对策; Nash 均衡

中图分类号: TP18

文献标识码: A

文章编号: 1673-629X(2012)05-0057-03

Markov Games Based Robot Arm 2D Path Planning

CHEN Kui, LIU Jiu-fu, SU Qing-qin, LIU Rong

(College of Automation, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China)

Abstract: Use the algorithm based on Markov games to do path-planning for 2D robot arm in connection with its complex application environment and more uncertain conditions. Path-planning for 2D robot arm is the basis of 3D path-planning. According to the actual work environment, set the arm's range of motion and select the common movements as the basic behavior set. Under kinds of conditions, the profits were shown. Then the profit of each joint was updated by the multi-agent Q-learning algorithm, and the formulas of movement's inverse kinematics are obtained. So the complexity of the algorithm is also reduced. It shows the best combination of trail, respectively, to draw the 2D motion trail in the case of barrier-free and a round obstacle, and then confirm the error of trail.

Key words: multi-joint robot; robot arm; multi-Agent system; Markov game; Nash equilibrium

0 引 言

机器人路径规划是指依据某个或某些优化准则(如行走路线最短、行走时间最短、工作报酬最大等), 在其运动空间内找到一条从起始状态到目标状态的最优无碰路径。路径规划技术是机器人学领域中的一个重要研究部分。

根据对环境状态的感知程度不同, 路径规划可以分为全局路径规划和局部路径规划。全局路径规划方法主要有栅格法、可视图法和自由空间法等, 局部路径规划方法主要有人工势场法、模糊逻辑控制算法、蚁群算法和粒子群算法等。全局和局部路径规划技术都只是处理自主移动机器人的路径规划问题, 缺少对多关节机器人的各个 Agent 对象之间的交互关系的描述,

因此上述规划技术均不适合码垛机器人的路径规划。

机械臂是多关节机器人, 是一个多 Agent 系统, 可以考虑应用 Markov 对策理论。Markov 对策模型是描述多 Agent 系统的一种常用数学模型, 可以清楚地描述多 Agent 学习中交互的本质。Markov 对策模型分为零和对策和非零和对策。两人零和对策可以使用极大极小 Q 学习算法求取最优解, 但是极大极小 Q 学习算法只能解决决策双方具有对抗性质的问题, 同时状态量不能过于复杂。非零和对策的适用范围更具有普遍性, 因此解决该类问题的模型都有局限性^[1-4]。文中通过机械臂各个关节之间相互预测的方法降低多 Agent 之间学习的组合难度, 减少学习状态, 提高学习速度。

收稿日期: 2011-09-24; 修回日期: 2011-12-27

基金项目: 国家自然科学基金(60674100); 南京航空航天大学基本科研业务费专项科研项目(NS2010069)

作者简介: 陈 魁(1986-), 男, 河南汝南人, 硕士研究生, 研究方向为嵌入式系统与人工智能; 刘久富, 博士, 研究方向为计算机科学与软件测试技术。

1 Markov 对策

1.1 Markov 对策

定义 1 一个 Markov 对策模型, 可以用五元组 $\langle S, \alpha, \{A_i\}_{i \in \alpha}, P, \{R_i\}_{i \in \alpha} \rangle$ 表示^[5-8]。

S : 有限状态集合。

α : 有限 Agent 集合。

A_i : 表示 Agent i 的有限行为集合。

$P: S \times A_1 \times A_2 \times \cdots \times A_{|\alpha|} \rightarrow [0, 1]$, 状态转移函数, 表示 S 集合上的状态分布。

$R_i: S \times A_1 \times A_2 \times \cdots \times A_{|\alpha|} \rightarrow \mathbb{R}$, 报酬函数, 表示 Agent i 在状态 s 下, 所有 Agent 采取行动后的即时报酬。

1.2 Markov 对策算法

多 Agent 系统的策略为 $\pi = (\pi_1, \pi_2, \cdots, \pi_n)$, 其中 π_i 表示 Agent i 执行行为 a_i 的概率。最优策略表示为 $\pi^* = (\pi_1^*, \pi_2^*, \cdots, \pi_n^*)$, 最优策略即是使每个 Agent 获得的报酬总和为最大时的行为^[9]。

定义 2 Nash 均衡: 在 n 个 Agent 的 Markov 对策中, 如果对任意的 i 有:

$$V_i(\pi_1^*, \pi_2^*, \cdots, \pi_i^*, \cdots, \pi_n^*) \geq V_i(\pi_1^*, \pi_2^*, \cdots, \pi_i, \cdots, \pi_n^*)$$

其中 $\forall \pi_i \in \Pi_i$ (1)

则称 n 元组 $(\pi_1^*, \pi_2^*, \cdots, \pi_n^*)$ 为 Nash 均衡点。其中, $V_i(\pi_1, \pi_2, \cdots, \pi_i, \cdots, \pi_n)$ 表示所有 Agent 组合策略的报酬函数。 Π_i 是 Agent i 可以执行的所有策略的组合。可以看出, 对每个 Agent 来说, Nash 均衡点是其对其他 Agent 的策略的最佳响应^[10-12]。

对 n 个 Agent 的 Markov 对策, 定义 Agent k 的 Nash 均衡 Q 值为:

$$Q_k^*(s, a_1, \cdots, a_n) = R_k(s, a_1, \cdots, a_n) + \gamma \sum_{s' \in S} p(s' | s, a_1, \cdots, a_n) V_k(s', \pi_1^*, \cdots, \pi_n^*) \quad (2)$$

该 Nash 均衡 Q 值定义于状态 s 和组合动作 (a_1, \cdots, a_n) 上, 是由状态 s 转移到状态 s' 时 Agent 遵循策略 $(\pi_1^*, \cdots, \pi_n^*)$ 选择动作 (a_1, \cdots, a_n) 所获得的折扣报酬的期望值。

Agent 更新 Q 值的规则如下:

$$Q^{t+1}(s', a'_1, \cdots, a'_n) = (1 - \alpha_t) Q^t(s', a'_1, \cdots, a'_n) + \alpha_t [R^t + \gamma V^t(s_{t+1})] \quad (3)$$

其中, α_t 为学习率, γ 为折扣因子, $Q^t(s', a'_1, \cdots, a'_n)$ 为第 t 个时间步, 环境状态为 s' , n 个 Agent 分别选择动作 a'_1, \cdots, a'_n 时, 某个 Agent 获得的报酬值。

算法 1 多智能体的 Q 学习算法:

- (1) 初始化参数。对 $\forall s \in S, \forall a \in A, Q(s, a_1, \cdots, a_n) = 1, V(s) = 1$, 指定初始状态为 s_0 。
- (2) 获得当前环境状态 s , 选择并执行动作 a' 。
- (3) 观察 $a'_1, \cdots, a'_n, R'_1, \cdots, R'_n, s^{t+1}$ 。
- (4) 按照式(3)更新 Q 值。
- (5) $t = t + 1$ 。
- (6) 如果 Q 值收敛则结束, 否则跳转到步骤(2)。

2 机械臂运动学模型

机器人运动学模型一般是基于坐标变换求得的, 最常用的是 D-H 坐标变换法。由于机械臂的机械结构限制, 可直接在二维平面坐标中建立运动学模型。

机械臂的机械结构简化为如图 1 所示的二维平面上的结构简图。机械臂的运动规划就是 A 的运动轨迹。连杆 3 在 B 点绕连杆 2 上下转动, 连杆 2 在 C 点绕连杆 1 上下转动, 设定连杆 1 不转动。

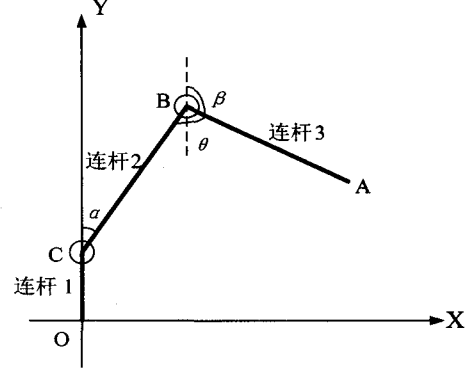


图 1 码垛机器人的结构简图

连杆 2、3 的杆长分别为 l_2, l_3 , 连杆 2 与 Y 轴正方向的夹角为 α , 连杆 3 与 X 轴正方向的夹角为 β , 连杆 2 与连杆 3 的夹角为 θ , 且 $\theta = \alpha + \pi - \beta$ 。由于机械结构的设计, 连杆 2 和 3 一直保持在一个平面内, A 点的坐标可表示为:

$$\begin{cases} x = l_2 \sin \alpha + l_3 \sin \beta \\ y = l_1 + l_2 \cos \alpha + l_3 \cos \beta \end{cases} \quad (4)$$

对机械臂的简化模型在 MATLAB 中进行运动规划仿真, 简化后的机械臂的机械结构数据如下: 连杆 1 的长度 $l_1 = 575\text{mm}$, 连杆 2 的长度 $l_2 = 1150\text{mm}$, 连杆 3 的长度 $l_3 = 1150\text{mm}$ 。初始位置时 $\alpha = -30^\circ, \beta = -45^\circ$, 目标位置时 $\alpha = 120^\circ, \beta = 105^\circ$ 。为了研究问题的方便, 在机械臂的工作空间放置圆形障碍物, 规划机械臂避障时的运动轨迹。

3 机械臂运动轨迹仿真

3.1 基本状态集合

机械臂的连杆由伺服系统驱动, 假设伺服系统驱动连杆的最小转动单位为 1° , 则连杆每次转动角度均为整数度。由连杆的运动范围知其状态空间为 $-30^\circ \leq \alpha \leq 120^\circ, -45^\circ \leq \beta \leq 105^\circ$ 。

3.2 基本动作集合

连杆 2 和 3 都有三个动作, 绕节点 C 和 B 做顺时针和逆时针转动, 或者不动。即机械臂的组合动作有 9 种, 如表 1 所示。

3.3 状态转移函数

由概率论可知, 机械臂所有动作的概率之和为 1,

即 $\sum_i a'_2 \cdot a'_3 = 1$ 。其中 a_2, a_3 分别为连杆 2 和 3 的动作, $a_2 \cdot a_3$ 为连杆 2 和 3 的组合动作。则连杆 2 和 3 的动作集及状态转移概率如表 1 所示。

表 1 连杆 2、3 的动作及状态转移概率

序号	连杆 2	连杆 3	概率
1	逆时针	逆时针	0.06
2	逆时针	不运动	0.12
3	逆时针	顺时针	0.05
4	顺时针	逆时针	0.17
5	顺时针	不运动	0.03
6	顺时针	顺时针	0.13
7	不运动	逆时针	0.18
8	不运动	不运动	0.25
9	不运动	顺时针	0.01

3.4 报酬函数

对于马尔可夫决策过程或者马尔可夫决策模型,按照评价函数得出的最后报酬是评价行为决策组合是否最优的标准。对于某个行为组合,也有相应的报酬,其报酬值如式(5):

$$R = \begin{cases} 2, & \text{接近目标位置} \\ 1, & \text{接近障碍物, 躲避} \\ -3, & \text{撞上障碍物} \end{cases} \quad (5)$$

3.5 运动轨迹仿真

设定初始位置时 $\alpha = -30^\circ, \beta = -45^\circ$, 目标位置时 $\alpha = 120^\circ, \beta = 105^\circ$ 。用坐标点表示为起始位置 $x = -1388\text{mm}, y = 2384\text{mm}$, 目标位置 $x = 2107\text{mm}, y = -298\text{mm}$ 。

(1) 无障碍物运动轨迹仿真。码垛机器人从高处抓取物体放置到低处, 其路径上没有放置障碍物, 机器人按照上述 Markov 对策准则运动, 得出最大报酬的仿真运动轨迹如图 2 所示。图中, 起始位置 $x = -1388\text{mm}, y = 2384\text{mm}$, 实际到达位置 $x = 2111\text{mm}, y = -296\text{mm}$ 。实际到达位置与目标位置的相对误差为 4mm , 在允许的误差范围内, 物体仍可以摆放到正确位置。

(2) 有障碍物运动轨迹仿真。码垛机器人从高处抓取物体放置到低处, 其路径上有一球形障碍物, 球心坐标为 $(2200\text{mm}, 750\text{mm})$, 半径为 250mm , 机器人按照上述 Markov 对策准则运动(由于码垛机器人的运动速度快, 为了能完全避开障碍物, 机器人只能在大于球形障碍物 50mm 外运动, 即算法中要躲避的球形障碍物比实际的半径大 50mm), 得出最大报酬的仿真运动轨迹如图 3 所示。图中, 起始位置 $x = -1388\text{mm}, y = 2384\text{mm}$, 实际到达位置 $x = 2092\text{mm}, y = -314\text{mm}$ 。实际到达位置与目标位置的相对误差为 23mm , 在允许的误差范围内, 物体仍可以摆放到正确位置。

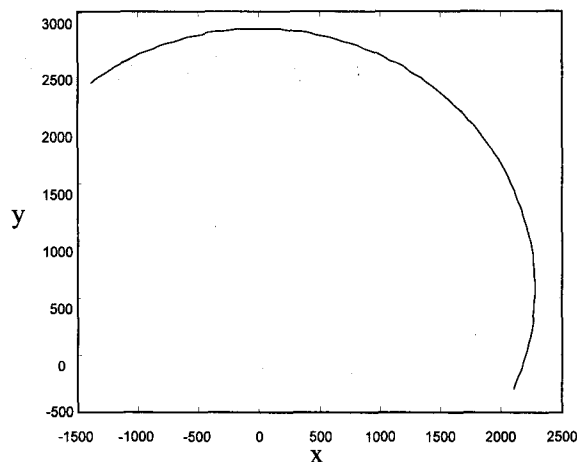


图 2 无障碍物的仿真运动轨迹

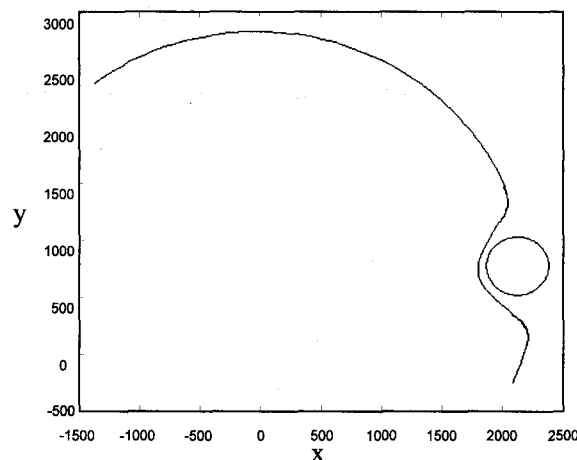


图 3 有障碍物的仿真运动轨迹

4 结束语

(1) 移动机器人的路径规划可以使用马尔可夫决策过程模型算法解决, 但是机械臂的每个关节都是一个自主运动的 Agent, 整个机械臂的运动路径是由各个关节协调完成, 是多 Agent 系统, Markov 对策模型可以很好地解决其路径规划问题。

(2) 机械臂的转动关节及连杆一直位于同一平面内, 其运动学模型是在同一个坐标系中建立的, 没有使用常用的 D-H 坐标变换法, 对绘制机械臂的运动轨迹简化了计算过程。

(3) 二维机械臂的路径规划是三维多关节机器人路径规划的基础, 仿真实验规划机械臂的路径对研究三维多关节机器人起到辅助作用。

(4) 为了避障的能力, 应该在运动路径上增加障碍物的数量, 或者改变障碍物的形状。这种情况以后要加到实验环境中, 进一步测试算法的效果。

参考文献:

- [1] 范波, 潘泉, 张洪才. 基于 Markov 对策的多智能体协

(下转第 63 页)

同时存在待定位节点请求包及信标节点应答包,全网的包传输将出现明显增长趋势。定位过程中,网络内传输的包个数基本保持在 60 以下,最多不会超过 70,并且能耗峰值的持续时间都不长,有效节约了全网能量。

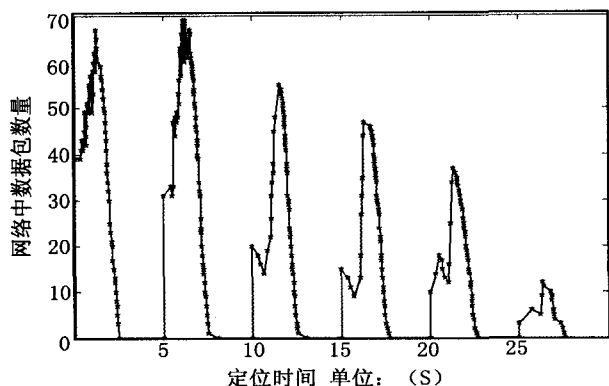


图2 各轮定位过程中的网络流量情况

4 结束语

文中以信标节点的角色变换策略为基础,提出了一种采用多种改进计算的综合协作定位方法。进行了较为完善的误差调整,拥有一定可信度。扩大了定位范围,实现了定位计算分布化,降低了网络能耗,节约了网络带宽。

本方法的定位改进策略前提是良好的网络通信环境。然而,由于无线信道传输不确定性,当网络拥塞时,将出现频繁的误差过滤,影响计算的正常进行。同时,当网内节点分布密度较大、单跳长度较短时,节点处理定位请求的时间就不应被忽略,此时易造成定位失真。这些都将成为今后研究与改进的方向。

参考文献:

- [1] Chong C Y, Kumar S. Sensor networks: evolution, opportunities and challenges[J]. *Proceedings of IEEE*, 2003, 91(8): 247-1256.
- [2] Mao G, Fidan B, Anderson B D O. Wireless Sensor Network Localization Techniques[J]. *Elsevier/ACM Computer Networks*, 2007, 51(10): 2529-2553.
- [3] 张佳,吴延海,石峰,等. 基于 DV-HOP 的无线传感器网络定位算法[J]. *计算机应用*, 2010, 30(2): 323-326.
- [4] Tian S, Zhang X M, Liu P X, et al. A RSSI-based DV-hop algorithm for wireless sensor networks[C]//*Proc. of IEEE WICOM2007*. [s. l.]: [s. n.], 2007.
- [5] Sayed A H, Tarighat A, Khajehnouri N. Network-based wireless location: challenges faced in developing techniques for accurate wireless location information[J]. *IEEE Signal Processing Magazine*, 2005, 22(4): 24-40.
- [6] 郝晓弘,李慧,粘坤. 功率控制在无线传感器网络定位中的应用[J]. *自动化仪表*, 2009, 30(8): 30-32.
- [7] Hofmann-Wellenho B, Lichtenegger H, Collins J. *GPS Theory and Practice*[M]. New York: Springer Wien, 1997.
- [8] 王福豹,史龙,任丰原. 无线传感器网络中的自身定位系统和算法[J]. *软件学报*, 2005(16): 857-868.
- [9] Bulusu B, Heidemann J, Estrin D. Density adaptive algorithms for beacon placement in wireless sensor networks[C]//*IEEE ICDCS01*. Phoenix, AZ: [s. n.], 2001.
- [10] 蒋峥,王汝传,孙力娟. 基于移动 Agent 无线传感器网络节点自定位算法[J]. *计算机技术与发展*, 2007, 17(6): 1-4.
- [11] 王书聪. 无线传感器网络分布式节点定位算法研究[J]. *计算机技术与发展*, 2008, 18(11): 62-65.
- [12] 杨永雷,朱军. 无线传感器网络中异步成簇算法的研究[J]. *计算机技术与发展*, 2010, 20(2): 145-151.
- [13] 调方法及其在 Robot Soccer 中的应用[J]. *机器人*, 2005, 27(1): 46-51.
- [14] 李晓萌,杨煜普,许晓鸣. 基于 Markov 对策和强化学习的多智能体协作研究[J]. *上海交通大学学报*, 2001, 35(2): 288-292.
- [15] 李晓萌,杨煜普,许晓鸣. 基于多级决策的多智能体自动导航车调度系统[J]. *上海交通大学学报*, 2002, 36(8): 1146-1149.
- [16] 高阳,周志华,何佳洲,等. 基于 Markov 对策的多 Agent 强化学习模型及算法研究[J]. *计算机研究与发展*, 2000, 37(3): 257-263.
- [17] Kaelbling L, Littman M, Cassandra A. Planning and acting in partially observable stochastic domains[J]. *Artificial Intelligence*, 1998, 101: 99-134.
- [18] 洪晔,王宏健,边信黔. 基于分层马尔可夫决策过程的 AUV 全局路径规划研究[J]. *系统仿真学报*, 2008, 20(9): 2361-2363.
- [19] Sharma R, Gopal M. Hybrid Markov Game Controller Design Algorithms for Nonlinear Systems[J]. *World Academy of Science, Engineering and Technology*, 2005(12): 328-332.
- [20] Sharma R, Gopal M. Markov Game Controller Design Algorithms[J]. *World Academy of Science, Engineering and Technology*, 2007(34): 585-593.
- [21] 胡佳,汪峰. 工业机器人路径规划的双目标优化策略[J]. *计算机技术与发展*, 2009, 19(5): 16-18.
- [22] Morisset B, Ghallab M. Learning how to combine sensory-motor functions into a robust behavior[J]. *Artificial Intelligence*, 2008, 172(4-5): 392-412.
- [23] Kaelbling L P. Hierarchical Task and Motion Planning in the Now[C]//*IEEE International Conference on Robotics and Automation*. [s. l.]: [s. n.], 2011.
- [24] 吕凌,曾碧. 基于评估和分工合作并行蚁群机器人路径规划[J]. *计算机技术与发展*, 2011, 21(9): 10-13.

(上接第 59 页)