

# A Multi-Robot Path Planning Method under Narrow Environments

Jie Shao, Jingru Yu

Department of Information Engineering, Zhengzhou Chenggong University of Finance and Economics, Zhengzhou Henan  
Email: [sj012328@163.com](mailto:sj012328@163.com)

Received: May 10<sup>th</sup>, 2015; accepted: May 24<sup>th</sup>, 2015; published: May 29<sup>th</sup>, 2015

Copyright © 2015 by authors and Hans Publishers Inc.  
This work is licensed under the Creative Commons Attribution International License (CC BY).  
<http://creativecommons.org/licenses/by/4.0/>



Open Access

---

## Abstract

Under narrow environments, conflict easily occurs when multi-robot path planning uses shared resources, and prioritisation is an important technology to solve this problem. This paper presents a dynamic allocation priority method based on learning classifier to improve the performance of the robot team. Firstly robots optimize their behaviors by XCS, and then high-level robot managers are introduced and trained to resolve conflicts by assigning priority. The novel approach is designed for partially observable Markov decision process environments. Simulation results show that the method presented is effective to solve the conflict in multi-robot path planning and improves the capacity of multi-robot path planning.

## Keywords

Path Planning, Multi-Robot, Learning Classifier System, Genetic Algorithm, Q Learning

---

# 狭隘环境下一种多机器人路径规划方法

邵 杰, 于景茹

郑州成功财经学院信息工程系, 河南 郑州  
Email: [sj012328@163.com](mailto:sj012328@163.com)

收稿日期: 2015年5月10日; 录用日期: 2015年5月24日; 发布日期: 2015年5月29日

## 摘要

狭隘环境下多机器人路径规划使用共享资源时, 极易产生冲突, 优先顺序化是解决共享资源冲突的一个重要技术。本文提出了一种基于学习分类器的动态分配优先权的方法, 提高机器人团队的性能。首先机器人通过XCS优化各自的行为, 然后引入和训练高水平的机器人管理者来分配优先权解决冲突。本方法适用于部分可知的Markov环境, 仿真实验结果表明本文所提方法用于解决多机器人的路径规划冲突是有效的, 提高了多机器人系统解决路径规划冲突的能力。

## 关键词

路径规划, 多机器人, 学习分类器, 遗传算法, Q学习

## 1. 引言

近年来, 多机器人路径规划已成为机器人领域中的一个热点问题。根据机器人对环境的了解程度, 可以分为两种类型: 环境信息完全已知的全局规划和环境信息完全未知或部分未知的局部路径规划[1] [2]。多机器人路径规划是为均处在同一个简单或复杂, 静态或动态, 已知或未知的工作空间中的每个机器人找到一条路径, 能避开障碍物并能保证机器人之间无碰撞, 最后以最小或较小的消耗完成自己的任务[3] [4]。由于多个机器人间存在资源竞争、且共处一个工作空间, 这就造成了机器人间的冲突, 解决机器人间的冲突已成为多机器人系统要解决的关键问题。

解决路径规划冲突问题, 已有众多学者使用众多方法做了多年的深入研究并产生了大量研究成果[5]-[7]。已有的研究解决方法大致可以分为集中式和离散式两类。集中式方法是将所有的机器人当作一个整体进行路径优化, 但在不同的环境下优化结果不可靠, 特别是对于部分可知的MDP环境或者机器人太多的环境。离散式方法是分别对单个的机器人进行路径优化, 并随时解决优化过程中出现的冲突。两种方法共同的缺陷是算法的计算时间太长, 仅能用于较小的环境, 且不能保证在较短的时间内总是获得优化或接近优化的结果。

基于上述分析, 本文提出采用基于XCS的离散启发式方法用于机器人的路径规划, 机器人的优先权由其XCS的预测回报值来确定, 而由集中式高水平的规划机器人动态分配各机器人的优先顺序, 以解决机器人间的冲突, 共同完成多机器人系统的路径规划。

## 2. 相关技术原理

### 2.1. Q学习

Q学习是由Watkins 1989年提出的基于Markov环境的一种学习算法(MDP) [8]。其原理是通过对感知的环境状态采取各种试探动作, 获得此种试探动作对此种环境状态的奖励或者惩罚信号, 并能不断调整学习策略以获得较大的奖励或较小的惩罚, 同时反馈给其它机器人, 并能够保证学习的收敛性。

单机器人的学习环境是基于Markov提出的有限四元组 $\langle S, A, f, \rho \rangle$ , 其中 $S$ 是有限环境状态集,  $A$ 为机器人有限动作集,  $f: S \times A \times S \rightarrow [0,1]$ 为状态转换概率函数。  $\rho_i: S \times A \times S \rightarrow R, i=1, \dots, n$ 为机器人的回报函数。

多机器人的学习环境是基于MR-MDP, 是一个随机的四元组 $\langle S, A, f, \rho \rangle$ 。  $S$ 为离散的环境状态集,  $A$ 为作用于机器人的动作集,  $A = A_1 \times \dots \times A_n$ ,  $n$ 为机器人数量。  $f: S \times A \times S \rightarrow [0,1]$ 为状态转换概率函数。  $\rho_i: S \times A \times S \rightarrow R, i=1, \dots, n$ 为机器人的回报函数。在多机器人系统, 状态转换是所有机器人联合行

动的结果,  $\alpha_k = [\alpha_{1,k}^T, \dots, \alpha_{n,k}^T]^T$ ,  $\alpha_k \in A$ ,  $\alpha_{i,k} \in A_i$ 。

## 2.2. 学习分类器(Accuracy-Based Learning Classifier System, XCS)

XCS [9]作为 LCS 的衍生系统是由 Wilson 提出的基于精度的学习分类器, 集成了强化学习、遗传算法和覆盖算法。XCS 是由条件、动作和预测回报三个主要的参数组成: 条件指明 XCS 可以匹配的环境输入状态; 动作表示在该 XCS 匹配的环境状态下 XCS 的建议动作; 预测回报  $P$  表示该 XCS 被匹配且执行了该 XCS 所建议的动作后系统预期从环境获得的回报。预测误差  $\varepsilon$  表示预测回报  $P$  的误差; 适应强度  $F$  表示预测回报  $P$  的精度。

Step 1: 在开始随机分类器集合选择动作概率大的动作  $P(a_t) = \frac{\sum_{cl_k \in |M|} a_t P_k \times F_k}{\sum_{cl_k \in |M|} a_t F_k}$  作用于环境, 根

据环境输入, 产生一个匹配环境输入的匹配规则集, 然后依据每个分类器的回报预测和适应强度值, 生成整个系统的预测列表和动作集。

Step 2: 分类器在执行机制过程中, 每选择一个动作  $a$  作用于环境, 分类器从环境获得一个回报  $R$ , 回报预测、分类器适应值强度更新如下:

$$P \leftarrow R + \gamma \max_a P(a) \quad (1)$$

其中  $0 \leq \gamma \leq 1$  是学习效率。

$$P_j \leftarrow P_j + \beta(P - P_j) \quad (2)$$

$$\varepsilon_j \leftarrow \varepsilon_j + \beta(|P - \varepsilon_j| - \varepsilon_j) \quad (3)$$

$$k_j = \begin{cases} 1 & \text{if } \varepsilon_j \leq \varepsilon_0 \\ \alpha(\varepsilon_j / \varepsilon_0)^{-\nu} & \text{otherwise} \end{cases} \quad (4)$$

$$k'_j = \frac{(k_j \times \text{num}_j)}{\sum_{cl_k \in |A|} (k_k \times \text{num}_k)} \quad (5)$$

其中参数  $\varepsilon_0$  ( $\varepsilon_0 > 0$ ) 控制预测误差  $\varepsilon$  的冗余度; 参数  $\alpha$  ( $0 < \alpha < 1$ ) 和  $\nu$  ( $\nu > 0$ ) 是常量, 控制当  $\varepsilon_0$  被超过时精确度  $k$  的下降速度。动作集中的绝对精确度值  $k$  被转换为相对精确度值  $k'$ , XCS 的适应度值  $F$  是依据相对精确度值进行更新:

$$F_j \leftarrow F_j + \beta(k'_j - F_j) \quad (6)$$

Step 3: 当匹配规则集中没有分类器与环境输入信息相匹配时, 启动规则发现机制的覆盖算法, 随机产生一个分类器。覆盖算法产生的新规则的条件部分与环境信息匹配, 该规则淘汰规则集中强度最小的规则, 存入规则集中。

Step 4: 未达到设定迭代次数转 Step 2。

## 3. 基于 XCS 的多机器人路径规划冲突解决方案

### 3.1. 多机器人学习训练方法

在基于 XCS 的多机器人路径规划冲突解决方案中, 机器人学习训练分为个体和团队训练两种方法。在个体训练阶段, XCS 使机器人学会避碰, 确保机器人在安全的情况下向设定目标运动。在团队训练阶段, 团队训练引入的机器人管理者决定机器人的优先权, 在任一时刻, 所有机器人在没有考虑其他机器人位置的情况下预先感知环境, 然后以机器人在管理者约束下的成功等级概率作为各机器人优先权, 各

机器人依优先权大小进行下一步的行动。

### 3.2. 机器人优先级别设计

多机器人在未到达目标位置前, 个体机器人  $R_i$  在区域  $S_i$  ( $S_i \in S, S \leq \sum S_i$ ) 内移动。如果在时刻  $t$ , 机器人  $R_i(x_i, y_i, \theta_i, v_i)^T$  对目标位置  $R_g(x_g, y_g, \theta_g, v_g)^T$  进行预测, 计算出自己预接近到目标的理想速度和方位角, 并将结果通过无线网络发布给其他机器人, 同时自己升级为领导机器人。其他机器人根据收到的信息计算出自己的位置和运动趋势, 并同时计算出目标和其他机器人与自身的相对位置, 根据其他机器人的情况, 提出自己预接近目标的理想速度和方位角。对于最终的等级方案, 领导机器人采用成功概率加权的方法。该方法首先建立如下位置估计方程:

$$\begin{cases} \left| (x_i + v_i t \cos \theta_i) - (x_g + v_g t \cos \theta_g) \right| = \varepsilon_{i,x} \\ \left| (y_i + v_i t \sin \theta_i) - (y_g + v_g t \sin \theta_g) \right| = \varepsilon_{i,y} \\ \varepsilon_{i,x}^2 + \varepsilon_{i,y}^2 = \Sigma_0^2 \end{cases} \quad (7)$$

其中:  $x_i, y_i, v_i, \theta_i$  分别为成员机器人位置、预计速度和方位角;  $x_g, y_g, v_g, \theta_g$  为机器人目标位置、速度和方向;  $\Sigma_0^2$  为对目标机器人的移动安全半径(根据机器人大小等因素设定的经验值);  $\varepsilon_{i,x}$  和  $\varepsilon_{i,y}$  为二维分量上的偏差。

将领导机器人  $x_{leader}, y_{leader}, v_{leader}, \theta_{leader}$  代入上式, 并将求得的  $t_{leader}$  代入下式, 由此可得各成员机器人在机器人管理者约束下的成功等级概率。

$$p_i = p(\varepsilon_{i,x}^2 + \varepsilon_{i,y}^2 \leq \Sigma_0^2 | t_i \leq t_{leader}) \quad (8)$$

### 3.3. 解决多机器人冲突的集成适应度函数

多机器人路径规划是要求各机器人能避开动态运行的障碍物, 且机器人之间不能相碰。我们把机器人是否在安全范围内, 结合机器人本身预测回报值和各机器人等级概率三方面考虑来设计解决机器人冲突的集成适应度函数。

#### 1) 是否在安全之内

我们把所有的障碍物和机器人都视为质点, 每个障碍物都有一个安全半径, 若机器人与障碍物的距离大于安全半径, 则认为是安全的; 若小于安全半径, 则认为是不安全。障碍物与机器人每一个路点  $(x_i, y_i)$  之间的距离  $d$  与安全半径  $R$  之间的关系如下:

$$fit1 = \begin{cases} 0 & d > R \\ -1 & d \leq R \end{cases} \quad (9)$$

其中  $d = \sqrt{(x_o - x_r)^2 + (y_o - y_r)^2}$ , 障碍物坐标  $(x_o, y_o)$ , 机器人坐标为  $(x_r, y_r)$ 。  $fit1$  的值表明只要机器人质点与障碍物的距离在安全半径之外, 则其适应度值为 0, 若有一个路径点与障碍物的距离在安全半径之内, 则适应度为-1。

#### 2) 机器人的预测回报适应度函数

$$fit2 = F_j \quad (10)$$

#### 3) 机器人的成功等级概率适应度函数

$$fit3 = p_i \quad (11)$$

综合上述因素, 机器人的集成适应度函数为:

$$f = (1 + fit1) * fit2 * fit3 \quad (12)$$

### 3.4. 解决多机器人路径规划冲突的等级策略[10]

本文采用离散启发式机器学习算法,结合 Q 学习和 XCS 算法共同解决多机器人路径规划中的冲突问题。LS-SVM 获得的最优学习策略作为 XCS 的初始规则集。XCS 通过与环境的交互,可以发现一组用于指导机器人避碰的学习规则,为机器人选择动作提供实时、动态的反馈,使机器人自主地学习到最优避碰规划策略。

Step 1: 最小二乘支持向量机产生的最优学习策略作为 XCS 的初始规则集,并随机执行其中的学习策略。

Step 2: 将消息列表中的消息与当前避碰规则的条件部分进行匹配,匹配的避碰规则送入匹配规则集。

Step 3: 从匹配规则集中选择合适的避碰编码规则送往效应器,引导机器人产生相应的避碰动作,并依据反馈值来判断避碰效果。

Step 4: 重复 Step 2~Step 3 过程,不断从匹配规则集中选择合适的避碰动作编码规则送往效应器,引导机器人产生相应的动作,直到匹配规则集为空或定时启动避碰规则发现机制。

Step 5: 规则发现系统采用遗传算法和规则覆盖算法构造新的避碰规则。

Step 6: 对产生的新的避碰规则进行规则合并,使其能概括以前的两个样本,滤除相似的个体,减少基因的单一性,加快了学习收敛速度。产生的新学习个体规则送入规则集,经过 XCS 发送到其公共规则集中,供其它机器人 XCS 产生新避碰规则时共享。

Step 7: 如果机器人未达到避碰效果,则返回 Step 2。

## 4. 仿真实验和分析

为了验证算法在多机器人的路径规划冲突中的有效性,以及提高多机器人系统解决路径规划冲突的能力的可行性。下面实验对所提算法进行全方位的验证比较,实验在不同的场景中进行。S 和 G 分别为机器人初始和目标位置,O1~O3 为动态障碍物,其它为一般静态障碍物。

实验场景一、U 型障碍物的动态狭隘环境

图 1 为 U 型障碍物的形状,所有图中 G 表示机器人目标位置,S 表示机器人初始位置。空心圆表示动态障碍物。

图 2 为多机器人系统在连续五个 U 型狭隘环境下的运动轨迹,动态障碍物(编号 1~3)分别在所属的 U 型狭隘环境中作上下运动。尽管路途中有连续多个 U 型陷阱,机器人经过前期个体和团队训练阶段的强化学习,机器人均能摆脱 U 型陷阱的局部最优,四个机器人最终均顺利到达目标位置 G 点。

实验场景二、狭窄通道的多障碍物动态狭隘环境

图 3 为机器人在狭窄通道环境下的运动轨迹。由于图 3 左图中危险品为静态障碍物,机器人在此环境下能够顺利绕过危险品障碍物的运动轨迹。图 3 右图内有两个作水平运动的动态障碍物(编号 1~2),机器人在绕过危险品障碍物后,能够精确预测到第一个动态障碍物 2 的位置,且直线快速通过动态障碍物 2 的移动区域,然后绕过第二个动态障碍物 1,从而完成机器人在狭窄通道环境下的整体运动规划,都能产生平滑的轨迹。

图 4 为多机器人在狭窄通道环境下的运动轨迹,动态障碍物(编号 1~3)均在所属环境下作上下直线运动,所有参与规划的机器人均取得较满意的模拟曲线。在动态狭隘环境下,由于 XCS 具有较强预测回报能力和环境学习能力,所有机器人都能取得满意的运动轨迹和稳定的收敛效果。

实验场景三、凸凹多障碍物的动态狭隘环境

图 5 为凸凹多障碍物形状。图 6 为多机器人系统在有多个凸凹多障碍物狭隘环境下的运动轨迹。尽

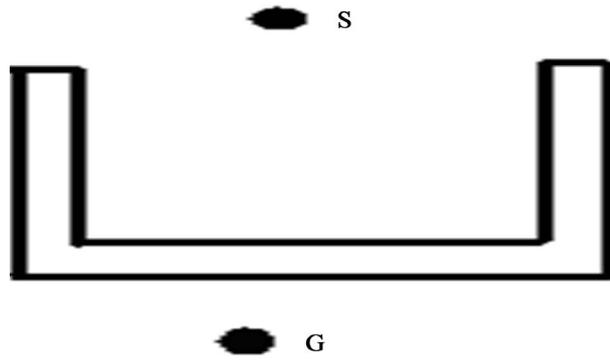


Figure 1. The shape of U-shaped obstacle

图 1. U 型障碍物的形状

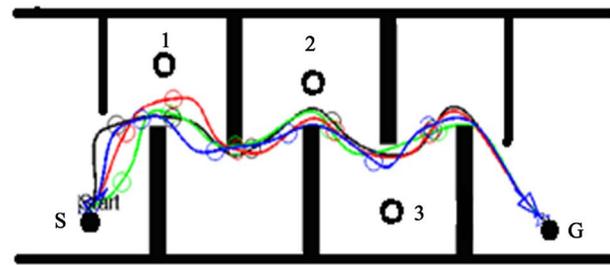


Figure 2. Multi-robot motion trajectories under U environment

图 2. U 型环境下的多机器人运动轨迹

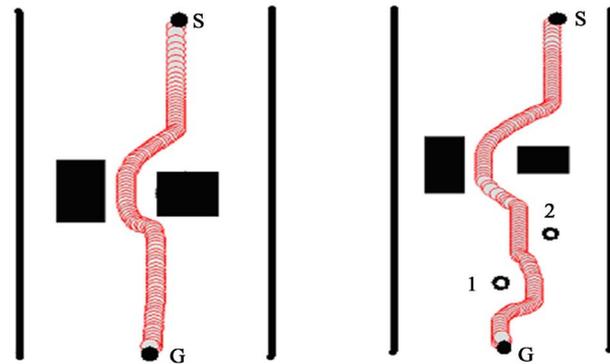


Figure 3. Robot trajectory in narrow channel environment

图 3. 狭窄通道环境下的机器人运动轨迹

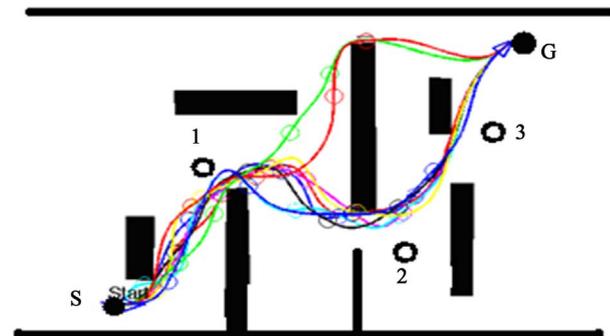


Figure 4. Multi-robot trajectories in narrow channel environment

图 4. 狭窄通道环境下的多机器人运动轨迹

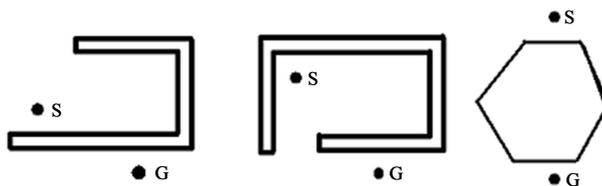


Figure 5. The shapes of many convex-concave obstacles  
图 5. 凸凹多障碍物形状

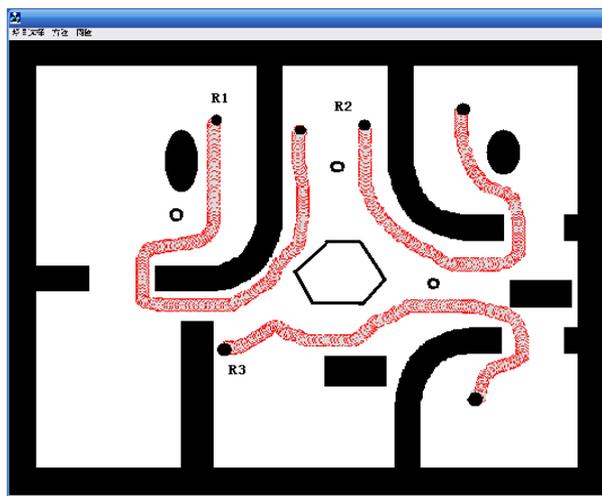


Figure 6. Multi-robot trajectories in convex-concave obstacles environment  
图 6. 凸凹多障碍物环境下多机器人运动轨迹

管有多个易于陷入局部最小的风险陷阱，但机器人(R1、R2、R3)在文献[10]所述的行为控制方法 follow\_wall、avoid\_obstacle、move\_to\_goal 的引导下，结合本文提出的 XCS 集成算法，机器人 R1、R2 和 R3 均顺利躲避附近移动的动态障碍物，所有机器人能够非常稳定的到达各自的目标。在安全性方面，图 6 最危险的地带在凸凹障碍物周围，在 30 次试验中机器人 R1、R2 和 R3 分别成功 29、29 和 30 次，整个多机器人系统路径规划成功率在 95% 以上，表现出良好的动态避障性能。

## 5. 结束语

本文提出采用基于 XCS 的离散启发式方法用于单个机器人的路径规划，机器人的优先权由其 XCS 的预测回报值来确定，而由集中式高水平的规划机器人动态分配各机器人的优先顺序，以解决多机器人间的冲突，共同完成多机器人系统的路径规划。很大程度上解决了遗传算法的早熟收敛、局部最优解、占据存储空间较大、收敛数度慢等问题，提高了多机器人发现安全路径的能力。

## 基金项目

2013 河南省高等学校“专业综合改革试点”项目；河南省教育厅重点基金资助项目(12B520047)。

## 参考文献 (References)

- [1] Masouri, M. and Aliyari, M. (2008) Teshnehlab. Integer GA for mobile robot path planning with using another GA as repairing function. *International Conference on Automation and Logistics*, Qingdao, 1-3 September 2008, 135-140.
- [2] 孟伟, 黄庆成, 韩学东, 等 (2005) 一种动态未知环境中自主机器人的导航方法. *计算机研究与发展*, 9, 1538-

1543.

- [3] 邵杰, 杨静宇, 万鸣华, 黄传波 (2010) 基于学习分类器的多机器人路径规划收敛性研究. *计算机研究与发展*, **5**, 948-955.
- [4] 孟偲, 王田苗 (2008) 一种移动机器人全局最优路径规划算法. *机器人*, **3**, 217-222.
- [5] Gao, Y. and Sun, S.-D. (2009) A collision based local path planning of mobile robot. 2009 *International Asia Conference on Informatics in Control, Automation and Robots*, Piscataway, 2009, 185-190.
- [6] 曲道奎, 杜振军, 徐殿国, 等 (2008) 移动机器人路径规划方法研究. *机器人*, **2**, 97-106.
- [7] Holland, J.H. (1986) A mathematical frame work for studying learning in classifier systems. In: Farmer, D., Lapedes, A., Packard, N. and Wendroff, B., Eds., *Evolution, Games and Learning*, North-Holland, Amsterdam, 307-317.
- [8] Musilek, P. (2005) Enhanced learning classifier system for robot navigatio. *Intelligent Robots and Systems 2005 International Conference*, Piscataway, 2005, 3390-3395.
- [9] Larry, B., Mattew, S., Anthony, B., et al. (2007) Learning classifier system ensembles with rule-sharing. *IEEE Transactions on Evolutionary Computation*, **11**, 496-502.
- [10] 邵杰, 杨静宇 (2011) 基于多 LCS 和人工势场法的机器人行为控制. *计算机科学*, **1**, 948-951.