文章编号 1004-924X(2002)02-0148-06

面向多机器人路径规划的一种基于模糊模型的再励函数结构

张 芳, 颜国正, 林良明
(上海交通大学 电子信息学院820所, 上海 200030)

摘要: 再励学习, 作为一种新兴的智能学习模式, 由于学习机制简单, 不需要任何先验知识, 也不需要样本数据, 被越来越多地用于未知环境模型系统的学习。而目前再励学习存在的问题之一是学习速度不高, 难以保证系统的实时性。在已有的再励学习系统中, 再励函数多采用无模型表示结构, 这种结构过于简单粗糙, 也是再励学习学习效率低下的主要原因之一。因此, 本文结合多机器人协调避障路径规划问题, 提出一种新的基于模糊模型的再励函数结构, 这种结构将反映机器人基本行为如躲避障碍物、其它机器人和趋向目标等的再励函数子函数进行分层建模, 并取模糊加权和来表示总的再励函数。仿真试验表明, 使用基于模糊模型的再励函数结构使再励学习的收敛速度要高于无模型结构。

关 键 词: 机器人; 再励学习; 再励函数; 模糊模型; 避障路径规划
中图分类号: TP242 文献标识码: A

1 引 言

再励学习是一种通过与环境的直接交互而获取知识的机器学习方法, 由于具有很好的实时在线学习能力, 对环境信息、系统模型和领域知识要求极低, 对未知环境下的控制问题是一个较好的解决途径, 如本文涉及的在未知环境下的多机器人协调避障路径规划问题。然而, 再励学习由于学习机制简单, 结构粗糙, 具有学习效率低下的缺点, 因此, 如何提高再励学习的学习效率是再励学习领域要解决的重要问题。有文献提出一些改进的学习算法或采用分层学习结构来加速学习, 较少涉及到对再励函数的研究, 而本文旨在通过改进再励函数表示结构来提高学习效率。

再励函数 $R(s, a)$ 反映在状态 s 下执行动作 a 所获得的即时报酬. 再励函数设计得合理与否, 直接影响到再励学习的收敛速度, 但其设计方法还很少有较为系统的研究。通常, 采用极其简单的目标奖励或动作惩罚表示的无模型结构。也有一些改进的无模型结构如 Mataric[1] 和 Balch[2,6] 在多机器人搜集(foraging)任务研究中将再励函数分解成反映执行任务事件发生、避障和趋向目标情况的子函数模块, 用这些相互独立的子模块和来表示总的再励函数, 但这仍然是一种比较粗糙的无模型非均匀表示结构。文献[3]在单机器人目标导航研究中提出基于先验知识的再励函数表示结构, 结合领域知识建立了一个精确的数学模型来表达再励函数。这种方法大大提高了学习效率, 但是当先验知识不足或系统复杂时, 将不能保证数学模型的正确性, 存在模型灾问题, 将严重影响系统的稳定性。

本文结合多机器人协调避障路径规划的应用研究提出一种新的再励函数表示结构– 基于模糊模型的再励函数表示结构。并将它与无模型非均匀表示结构进行了性能比较。这种再励函数结构既解决了精确数学模型的模型灾问题, 又克服了无模型结构的过于简单粗糙的缺陷。本文第二节介绍应用实例– 多移动机器人协调避障路径规划问题以及面向路径规划的无模型非均匀再励函数结构; 第三节对基于模糊模型的再励函数结构展开详细的讨论; 仿真实验和结论分别在第四、五节给出。

收稿日期: 2001-09-18; 修订日期: 2002-01-27

2　多移动机器人协调避障路径规划问题

多机器人协调避障路径规划是多移动机器人实现协调协作任务的基础, 反映了机器人在运动过程中对周围环境和其它机器人的交互能力。所谓多机器人协调避障路径规划, 是指在环境信息已知、部分已知或完全未知的情况下, 为机器人寻找一条从起点到终点的无碰路径, 同时又要保证某性能指标(距离或运动时间) 最优或次优。多移动机器人系统不仅要考虑躲避固定障碍物和移动障碍物, 还要考虑避免与其他机器人碰撞, 并且要保证各机器人最终趋向目标。本文假设在障碍物信息未知的环境下, 若干机器人以路径最短的原则找到一条从已知的起点到各自的已知目标点的无碰撞路径。

2.1　机器人模型

假设机器人模型如图 1 所示, 机器人前方安装超声波传感器, 用于检测[- 90°, 90°] 之间以相隔22. 5° 的均分的 9 个方向上的局部障碍信息 d_i, $i = 1 \sim 9$, 检测范围为 0. 3~ 3m。假设机器人装有全局定位系统, 能检测自身的全局位置和前进方向, 也就是机器人距离目标的距离 d_g 和前进方向与目标的夹角 θ_g。假设机器人装有彩色 CCD 而各个机器人具有不同于环境的颜色, 因此机器人根据简单的灰度处理就可以识别检测到的障碍物为固定障碍物或其他机器人. 机器人的控制动作有前进 0. 3m, 左转 22. 5°, 右转 22. 5° 三个, 没有后退动作, 因此机器人后面也不设传感器。

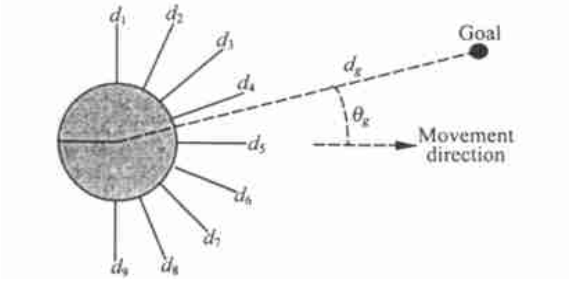


图 1　机器人模型

Fig. 1　Robot model.

2.2　无模型再励函数结构

基于再励学习的多移动机器人避障路径规划方法是在未知环境模型下路径规划问题的一个很好解决途径, 已引起较多学者对其进行研究。对于其中的再励函数表示结构, 多采用动作惩罚均匀表示结构, 以及一些改进的无模型结构, 我们也在文献[4] 中提出了面向多移动机器人协调避障路径规划问题的无模型非均匀再励函数结构。对应多移动机器人协调避障路径规划的三种基本行为, 将总的再励函数 R 分解成趋向目标再励函数 R_g、躲避固定障碍物再励函数 R_0 和躲避其它机器人再励函数 R_r 三个子函数模块。考虑多移动机器人协调避障路径规划的趋向目标行为和避障行为存在较大的相关性, 对子函数取加权和来表示总的再励函数:

$$R = w_1 R_g + w_2 R_0 + w_3 R_r$$

w_i, $i = 1 \sim 3$ 为相应的权值。

这种无模型非均匀再励函数结构尽管比极其简单的动作惩罚均匀表示结构有所改进, 但也是根据极少的领域知识人为设计的, 比较简单粗糙, 影响再励学习的学习速度。为此, 本文考虑采用一些知识性的描述方法来对再励函数建模, 提出基于模糊模型的再励函数表示结构。

3　基于模糊模型的再励函数表示结构

基于模糊模型的再励函数表示结构如图 2 所示。经过对传感器数据的处理后获得机器人与障碍物、其它机器人和目标的最短距离为 d_0、d_r 和 d_g, 机器人前进方向与障碍物、其它机器人和目标的最短距离方向之间的夹角分别为 θ_0、θ_r 和 θ_g。这样, 模糊模型的输入量为:

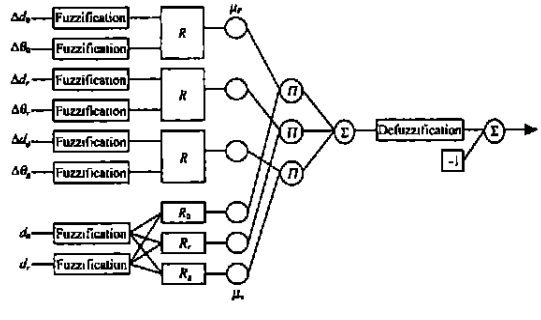


图 2　模糊模型结构

Fig. 2　Fuzzy model structure.

$$\Delta d_i = d_i^t - d_i^{t-1},$$

$$\Delta\theta_i = |\theta_i^t| - |\theta_i^{t-1}|,\ i = o, r$$

$\Delta d_g = d_g^{t-1} - d_g^t$，　$\Delta\theta_g = |\theta_g^{t-1}| - |\theta_g^t|$

其中，d_0^{t-1}、d_o^t 分别表示上一时间步和当前时间步机器人检测到与固定障碍物的最近距离。θ_o^{t-1}、θ_o^t 分别表示上一时间步和当前时间步机器人前进方向与最近距离所处的方向的夹角。d_r^{t-1}、d_r^t 分别表示上一时间步和当前时间步机器人检测到与其它机器人间的距离。θ_r^{t-1}、θ_r^t 分别表示上一时间步和当前时间步机器人前进方向与其它机器人的夹角。d_g^{t-1}、d_g^t 分别表示上一时间步和当前时间步机器人与目标间的距离。θ_g^{t-1}、θ_g^t 分别表示上一时间步和当前时间步机器人与目标间的夹角。

针对再励函数的状态空间较大，直接对此多输入多规则建模将使模型和规则变得极其复杂，为此，将整个模糊模型根据机器人行为分层，先分别对各子行为再励函数 $r_j, j = o, r, g$ 进行模糊推理，然后在对各子行为的模糊权 $w_j, j = o, r, g$ 建模，模糊推理规则设计为如图3所示。然后对各子行为再励函数 μ_{rj} 进行模糊加权乘积和，即为总的再励函数，用面积中心法去模糊化后再偏置－1即为再励函数的精确输出值 r。

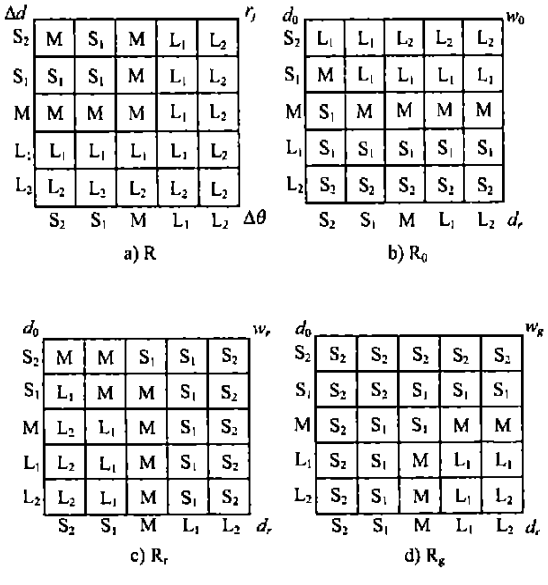


图3　模糊规则

Fig. 3　Fuzzy rule.

这里，隶属函数采用钟形函数 $r_j, w_j, j = o, r, g$ 和 r 的隶属函数如图4所示，而 $\Delta d_j \in [-0.3, 0.3]$，$\Delta\theta_j \in [-\pi, \pi]$，$j = o, r, g$ 和 $d_i \in [0, 3]$，$i = o, r$ 的隶属函数根据各变量的取值范围对其作相应的扩展。

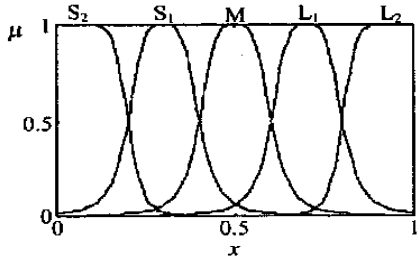


图4　隶属函数

Fig. 4　Member function.

4　仿真实验

假设在一个20m×20m的环境中散落着几个形状和位置未知的障碍物，要求几个机器人从已知起点向已知终点以路径最短原则移动。然后，增加一个机器人，并改变部分起点、终点和环境中障碍物的位置和大小，再进行试验，看机器人是否仍能够成功地到达目标点。表1为仿真实验时机器人的起点和终点位置。

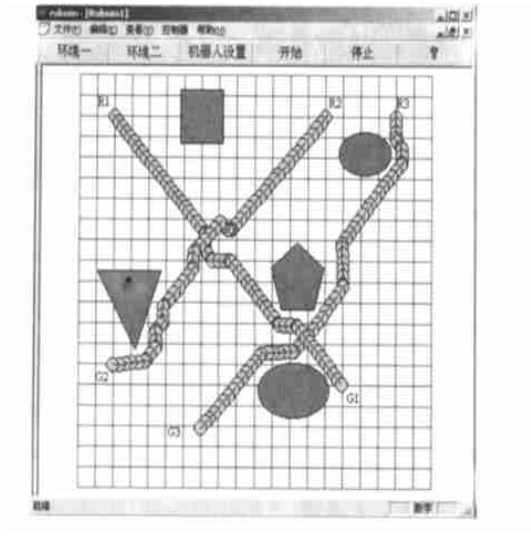
表1　机器人在仿真环境的起点和终点位置

Table 1　The start and end positions of robots in simulation environment

Robot	Initial environment		Changed environment	
	Starting point	End point	Starting point	End point
R_1	2, 18	15, 5	2, 18	16, 4
R_2	14, 18	2, 6	2, 4	13, 18
R_3	18, 18	7, 3	18, 19	11, 3
R_4			19, 12	1, 12

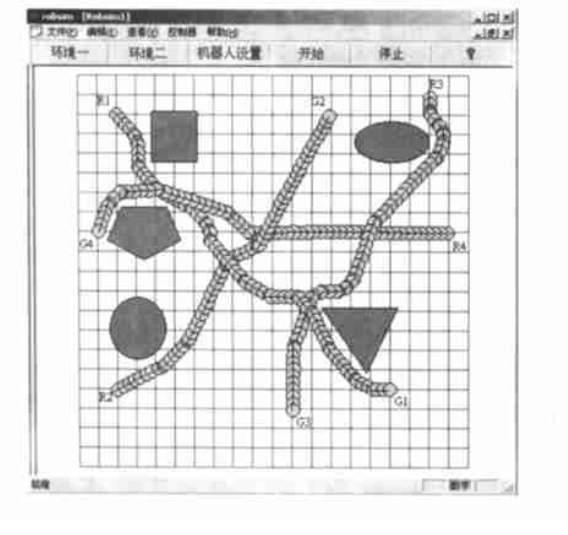
再励学习算法采用文献[5]提出的基于非均匀模糊分割的FCMAC函数逼近器的滞后更新多步Q学习算法。试验表明，算法中无论是采用无模型非均匀再励函数结构还是基于模糊模型的再励函数结构，算法均有效且有较好的鲁棒性。图5反映了机器人在原始环境和改变了的环境下经过学习后走出的路径轨迹。

比较两种再励函数的实验结果, 可见采用基于模糊模型的再励函数结构使算法具有更快的收敛速度, 大约试验 2900 次就收敛于最优策略, 而无模型非均匀结构的收敛速度大约是 3500 次。如图 6 所示。图 7 为机器人在不同试验次数下到达目标点所需的移动步数。



(a) 原始环境

(a) Initial environment.

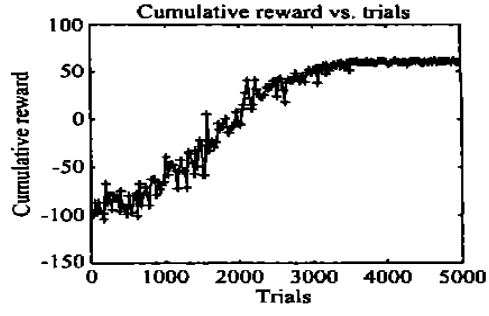


(b) 变化环境

(b) Changed environment.

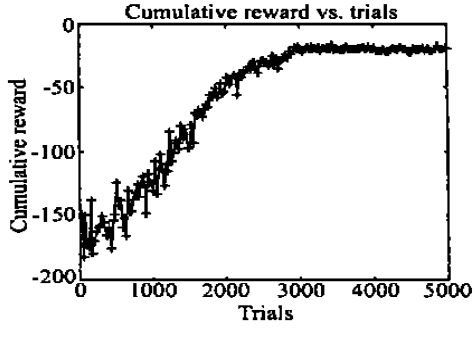
图 5　多机器人的移动轨迹

Fig. 5　The moving trajectory of multiple mobile robots.



(a) 无模型

(a) Model_free

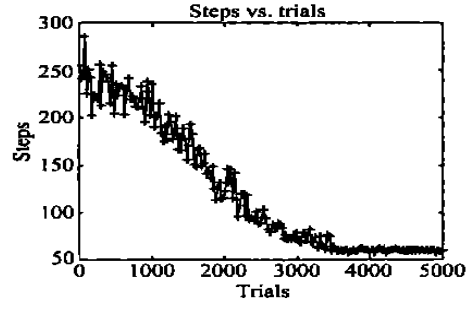


(b) 模糊模型

(b) Fuzzy model

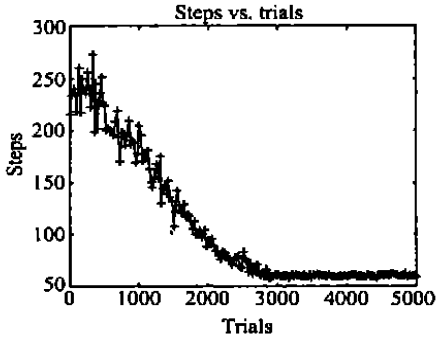
图 6　累积报酬和试验次数的关系

Fig. 6　Relationship between cumulative reward and trials.



(a) 无模型

(a) Model_free.



(b)模糊模型

(b) Fuzzy model.

图7　机器人移动步数和试验次数的关系

Fig. 7　Relationship between robots′ moving steps and trials.

5　小　　结

针对再励学习学习效率低下的缺陷,本文结合多移动机器人协调避障路径规划问题提出基于模糊模型的再励函数结构,这种结构用语言变量来描述复杂的非线性系统,用高级的仿人模糊推理机制来推理各子行为再励函数综合作用下的总的再励函数,使得这种函数结构表示的再励函数更确切,从而克服了无模型结构的过于简单粗糙的缺陷。实验表明,利用基于模糊模型的再励函数表示结构比无模型非均匀再励函数结构大大提高了再励学习的收敛速度。

尽管基于模糊模型的再励函数表示结构是结合多机器人协调避障路径规划的应用研究提出的,但它不失其一般性,可以扩展应用于其它领域。

在本文中,隶属函数结构和模糊规则都是根据先验知识人为设计的,缺乏学习功能辨识和自校正隶属函数和推理规则,因此,下一步的工作将考虑利用神经网络的学习能力和分布式连接结构,在模糊模型中引入神经网络来对模糊规则和隶属函数进行学习和调整。

参考文献:

[1] Mataric M J. Reinforcement learning in the multi_robot domain[J]. *Autonomous Robots*, 1997,4(1): 73- 83.

[2] Balch. Reward and diversity in multirobot foraging[A]. *IJCAI- 99 Workshop on Agents Learning About, From and with Other Agents*[C]. 1999.

[3] 李强. 复杂连续系统的再励学习系统- 算法设计及应用[D]. 上海: 上海交通大学,2000.

[4] 高志军, 颜国正, 　清. .多机器人协调与合作系统的研究现状和发展[J]. 光学 精密 工程,2001,9(4): 99- 103.

[5] 张芳, 颜国正, 林良 明. 一种基于非均匀模糊分割的模糊 CMAC 函数逼近器的再励学习方法[J]. 上海交通大学学报, 2002,(10): 15- 20.

[6] 陈忠泽, 颜国正, 林良 明, 等. 一种新的机械手最优轨迹的规划算法[J]. 光学 精密 工程,2001,9(3): 242- 246.

Multi_robot path planning_oriented and fuzzy model_based reinforcement function structures

ZHANG Fang, YAN Guo_zheng, LIN Liang_ming

(No. 820 Lab, College of Electronics and Information Technology, Shanghai Jiaotong University, Shanghai 200030, China)

Abstract: As a newly rising intelligent learning mode, reinforcement learning is being applied more and more in a learning system with unknown environment model because of its simple learning mechanism and no need of knowledge of the system or sample data in advance. However, one of the problems of the reinforcement learning method is that its learning speed is too low to ensure the real_time system. Researchers

have studied to speed up learning by improving learning algorithm and adopting intelligent exploration policy or applying the hierarchical reinforcement learning method, etc. However, how to describe the reinforcement function and how the reinforcement function affects the learning speed are seldom studied. In the existing reinforcement learning system, the model_free reinforcement function artificially defined is usually used. Its simple and rough expression is one of the causes of the low efficiency of learning. In this article, a new fuzzy model_based reinforcement function structure is presented. It is described according to the actual application in the conflict- free path planning problem of a cooperative multiple mobile robot system. In this system, the robot behaviors are divided into three basic kinds moving to the goal, avoiding obstacles and other robots. Then, the subfunctions reflecting these basic behaviors of robots are hierarchically and fuzzily modeled, and the final reinforcement function is expressed by the sum of fuzzy weighted sub- functions. The fuzzy model based reinforcement function has more accurate expression of the influence of each robot's action on the environment. The simulation shows that using the fuzzy model based reinforcement functions in reinforcement learning algorithm can further speed up the convergence than using model- free reinforcement functions.

Key words: robots reinforcement learning; reinforcement function; fuzzy model; conflict- free path planning

作者简介: 张　芳(1971-) , 女, 浙江省嵊州市人, 博士研究生, 研究领域: 多机器人协调、再励学习;
颜国正(1960-) , 男, 湖南省人, 博士生导师, 研究领域: 多机器人协调、微机械;
林良明(1939-) , 男, 福建省人, 博士生导师, 研究领域: 多机器人协调、微机械。