文章编号 :1002 - 0411(2003)03 - 0229 - 05

基于替代传导径迹增强式学习的自主式微直升机控制

杨玉君　程君实　陈佳品　张　琛　肖永利

(上海交通大学信息存储研究中心　上海　200030)

摘　要 :随着微电子机械系统(MEMS) 的迅猛发展 ,自主式微直升机的研究也已成为这一领域内的研究热点之一 .由于微直升机尺寸的限制 ,不能安装功能很强的传感器和处理器 ,难以获得完全的环境信息 ,所以传统的基于模型的控制方法不适用于环境是动态的自主微直升机控制 .基于行为的控制方法采用累次逼近的方法 ,不需要环境的精确模型 ,因此系统的稳定性较好 .本文采用基于替代传导径迹的增强式学习 ,结合即时差分方法 ,提高其学习效率 ,仿真实验验证了该学习算法的有效性 .最后 ,本文介绍了微直升机控制中存在的一些问题和我们以后的改进方向 .*

关键词 :微电子机械系统 ;自主微直升机 ;增强式学习 ;替代传导径迹 ;即时差分

中图分类号 :TP24　　文献标识码 :B

AUTONOMOUS MICRO-HELICOPTER CONTROL BASED ON REINFORCEMENT LEARNING WITH REPLACING ELIGIBILITY TRACES

YANG Yu-jun　CHENG Jun-shi　CHEN Jia-pin　ZHANG Chen　XIAO Yong-li

(*Information Storage Research Center , Shanghai Jiaotong Univ . , Shanghai* 200030 , *China*)

Abstract : With the rapid development of MEMS , study of micro helicopter has been a hotpot in this field . Because of its overall size , the micro helicopter could not be equipped with strong sensors and MPU , which affect the helicopter to get the whole environment information , therefore , the traditional control method disagrees with the helicopter in the uncertain environment . However , the method based on behavior only uses trial and error without the exact model of the environment . We adopt reinforcement learning with replacing eligibility traces to be combined with the temporal difference learning , which improves the efficiency and speed convergence . The results of simulation prove the validity of the learning algorithms . At last , this paper introduces the existent problems with the helicopter control and gives the future study trend .

Keywords : MEMS , autonomous micro helicopter , reinforcement learning , replacing eligibility trace , temporal difference

1　引言 (Introduction)

近年来 , MEMS 技术在国防军工等方面的应用已受到越来越广泛的重视 ,各国都在不断加速武器装备的小型化和信息化进程 ,这使得以 MEMS 技术为基础 ,具有广泛军事用途的微型飞行器(MAV) 研制已成为世界上许多国家的研究热点之一 .上海交通大学信息存储研究中心充分发挥了研制微马达的成功经验和在微细加工及微精密装配技术上的优势 ,成功研制出以直径 2 mm 电磁型微马达作为驱动器能离地飞行的微直升机[1] .

由于微直升机的尺寸的原因 ,不能安装功能很强的传感器(例如 ,CCD 摄像头和声纳传感器) ,所以微直升机不能获得完全的状态信息 .传统的基于模型[2]的控制方法是建立在可以获得外界精确模型的假设之下的 ,这在实际系统中不容易实现 ;实际物理系统一般是多维的 ,获取状态空间的所有数据比较困难 ,即使在给定精确模型的条件下 ,在高维空间中建立依据维数成指数级增长的最优策略也非常复杂(维数灾) .相对而言 ,基于行为的控制方法[3]利用并行的策略来解决这些问题 .每种行为只需要完成当前任务的少量信息 ,避免了信息过载 .由于任务的划分 ,系统不需要构造和维护环境的整体模型 ,在一定程度上也减少了计算量 ;另一方面 ,基于行为的方法可以构造分层的体系结构(例如 , Brooks[3] 的 Sub-

* 收稿日期 :2002 - 07 - 16
基金项目 :国家自然科学基金资助项目(69889050) ;863 计划资助项目(863-512-04-01) ;总装备部资助项目

sumption Architecture) ,高层行为可以抑制和调整低层行为,系统可以方便地增加功能,而不丧失已经产生的低层行为能力,具有很强的扩展功能.

增强式学习是一种适合于基于行为方法的半监督学习,我们的微直升机采用替代传导径迹的增强式学习进行学习控制(learning control) ,这种学习比累积传导径迹的增强式学习有更好的收敛性,并且没有偏差[4],仿真实验验证了该方法的有效性.

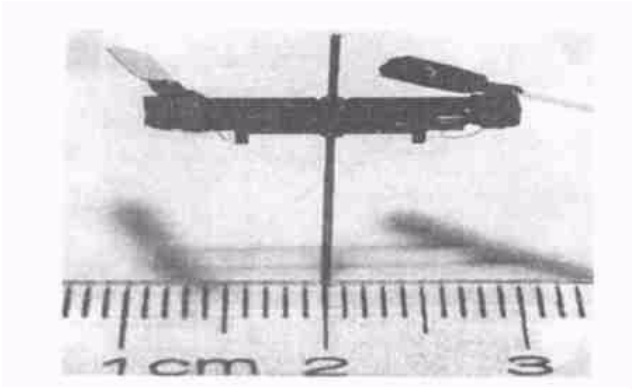

图 1 微型直升机照片

Fig .1 Photo of the micro helicopter

2 微直升机(Micro helicopter)

上海交通大学信息存储研究中心在总装备部的支持下,已经研究成功了以直径 2mm 电磁型微马达作为驱动器.离地飞行的微型直升机[1],如图 1 所示.直升机整体尺寸仅为 18.8 × 2.5 × 4.6mm^3,重量为 106.7mg, 其结构如图 2 所示.

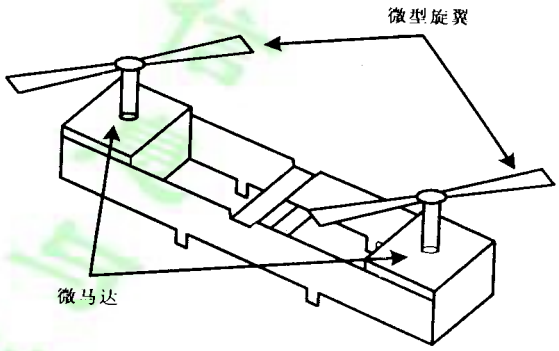

图 2 微型直升机结构

Fig .2 Micro helicopter configuration

在此基础之上,综合考虑各种因素,提出一种四旋翼结构形式的微直升机设计思想,下文将做介绍.

2 .1 四旋翼微直升机结构

四旋翼的微机器人直升机的整体尺寸为 28 × 28 × 4 .6mm^3,其结构形式如图 3 所示.该微型飞行器由四个直径仅 2mm 的电磁型微电机作为驱动部件,微电机转轴上装有可产生升力的微型旋翼.该微型飞行器的结构特点是外型简单紧凑,易于微细加工和微精密装配;圆型机身可为以后加载有效负载提供空间;陶瓷薄片与微电机相连有利于微电机散热;机身对旋翼气流流通影响很小;整体结构体积小、重量轻、强度高[5].

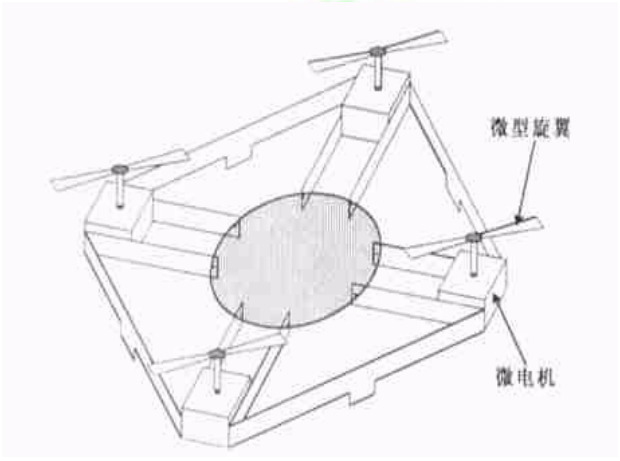

图 3 四旋翼结构形式的微直升机

Fig .3 Configuration of four-rotary-wing micro helicopter

2 .2 四旋翼微直升机的基本行为

旋翼飞机是通过改变旋翼拉力大小以及旋翼桨盘方向来控制飞行器运动方向的.微直升机采用直流电机作为驱动器,而电机转速的调整速度很快,因此旋翼拉力大小容易改变和控制.另外,微直升机设计结构也是通过改变旋翼拉力大小来调整旋翼桨盘方向,从而可控制飞行器飞行方向.

从飞行要求来看,微型飞行器具备空间六个运动自由度和四个可以控制的基本运动状态.其中四个基本运动状态分别是:(1) 上下飞行;(2) 前后飞行;(3) 侧向飞行;(4) 水平转动.我们定义微直升机的 5 种基本行为,就是在四种运动状态的基础之上再增加一种自身姿态控制行为.我们采用分层的行为体系结构,低层的行为负责基本运动和姿态控制,高层行为负责完成指定的任务,如图 4 所示.

由于微直升机能力有限,现阶段我们主要研究导航飞行,姿态平衡是一个很重要的基本行为,只有在直升机姿态平衡的基础上才可以做其他飞行动作.当然,这些行为最后都要转化为对四个 2mm 微马达的协调控制[5],如图 5 所示,本文只讨论基本行为的控制.

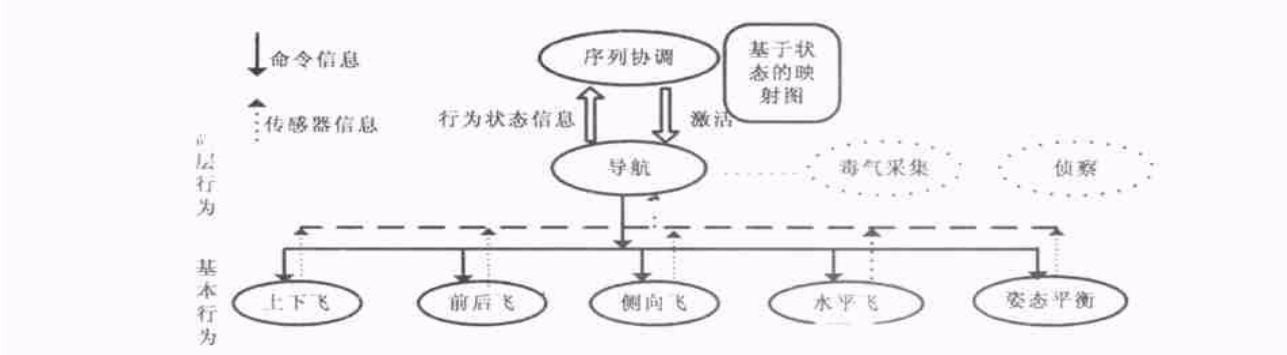

图 4　基于行为的控制结构

Fig .4　Control architecture based on behavior

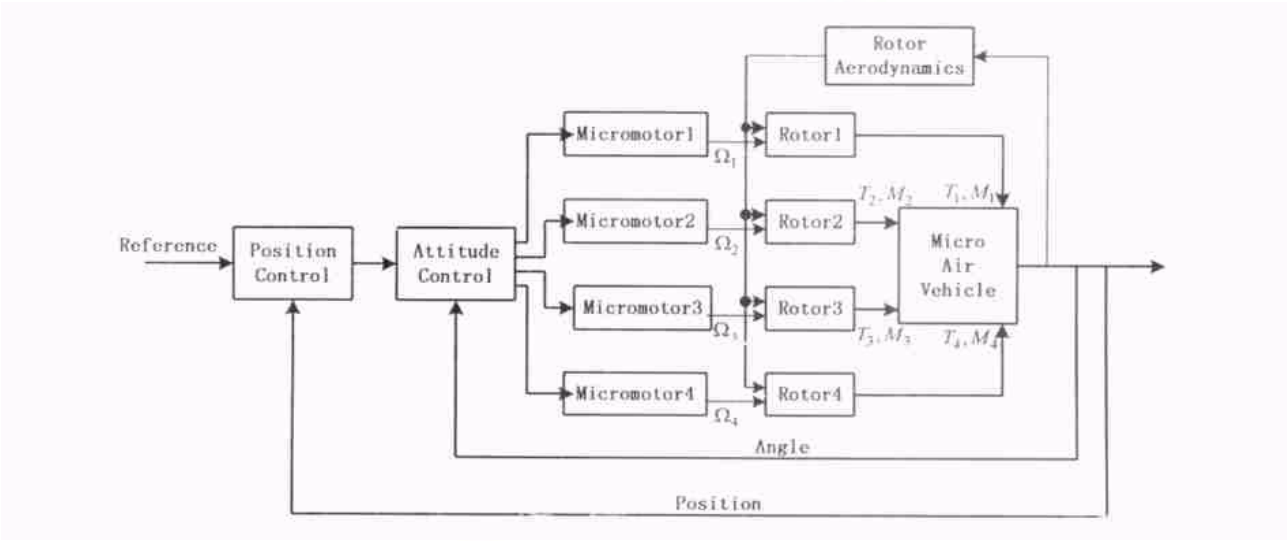

图 5　微马达控制系统框图

Fig .5　Diagram of micro motors control system

其中, Ω_i (i =1 ,2 ,3 ,4) 分别为微直升机四个马达的转速, T_i 、M_i (i =1 ,2 ,3 ,4) 分别为四个微马达对应旋翼的拉力和扭矩.

3　替代传导径迹的增强式学习 (Reinforcement learning with replacing eligibility traces)

增强式学习 (RL) 已经成功地用于智能体系统并取得很好的效果,而且是一种几乎不需要获得环境的动力学知识的学习方法.增强式学习基于逐渐逼近的机理,模仿人类的学习策略,智能体和环境交互之后,通过接受来自环境的增强式信号进行学习,这使得系统很鲁棒.而基于模型的智能体的学习通常需要环境的确切模型,所以并不适用于动态环境.

3.1　传导径迹

在增强式学习中,有两种基本机理来处理滞后报偿,一种是即时差分学习 (例如 $TD(\lambda)$ 算法和 Q 学习),另外一种就是传导径迹机理[4].径迹是指每一次状态访问时初始化的短期记忆过程,它会随着时间的推移而逐渐衰减,径迹将状态定义为学习的适应度 (eligibility).在状态适应度不为零的情况下,信念分配可以根据径迹的适应度来完成.传导径迹有两种,一种是累积传导径迹,一种是替代传导径迹.在累积传导径迹中,径迹在每次访问状态时建立,而在替代传导径迹中,径迹复位为 1,替代原来的径迹.在累积传导径迹中,最近和经常被访问的状态获得更多的信念,而在替代传导径迹中,消除了状态被访频率对信念分配的影响,对性能的改进有很大的影响.

累积传导径迹和替代传导径迹分别如式 (1) 和 (2) 所示:

$$e_{i+1} = \begin{cases} \gamma\lambda e_t(s) & \text{if } s \neq s_t \\ \gamma\lambda e_t(s) + 1 & \text{if } s = s_t \end{cases} \tag{1}$$

$$e_{i+1} = \begin{cases} \gamma\lambda e_t(s) & \text{if } s \neq s_t \\ 1 & \text{if } s = s_t \end{cases} \tag{2}$$

其中, $\lambda(0 \leqslant \lambda \leqslant 1)$ 表示衰减因子, $\gamma(0 \leqslant \gamma \leqslant 1)$

表示折扣因子,$e_t(s)$是状态 s 在 t 时的径迹,并且 s_t 为时间 t 时的实际状态.

3.2 即时差分算法

即时差分($TD(\lambda)$)算法结合了传导径迹的即时差分学习来预估值函数(value function),离散形式的 $TD(\lambda)$算法定义如下:

$$\Delta V_t(s) = \alpha_t(s)[r_{t+1} + \gamma V_t(s_{t+1}) - V_t(s_t)] \cdot e_{t+1}(s) \quad \forall s, \forall t \text{ s.t. } s_t \neq T \tag{3}$$

其中,$V_t(s)$是 $V(s)$在时间 t 的估计值,$\alpha_t(s)$是正的步长参数,$e_{t+1}(s)$是状态 s 的传导径迹,$\Delta V_t(s)$是估值 $V(s)$在 t 时的增量.在线 $TD(\lambda)$的估值在每一个时间步长都增加:$V_{t+1} = V_t(s) + \Delta V_t(s)$,而在离线 $TD(\lambda)$学习中,$\Delta V_t(s)$在终点状态才被考虑,它的报偿被延滞了.

4 仿真实验 (Simulation experiment)

4.1 仿真环境

仿真环境如图6,环境为10×10的网格空间,每个网格的大小为28 mm×28 mm.在环境中有一长方型的障碍物,微直升机通过增强式学习,避开障碍物,寻找最短的路径到达目的地.微直升机的位置由重心在网格环境中的坐标决定.我们设定微直升机的初始位置为(1,2),终点(terminal)位置为(7,5).在仿真实验中,我们只考虑微直升机在水平面内的二维运动.仿真平面是不可包绕的(unwrap),微直升机不能通过界面绕到对面.

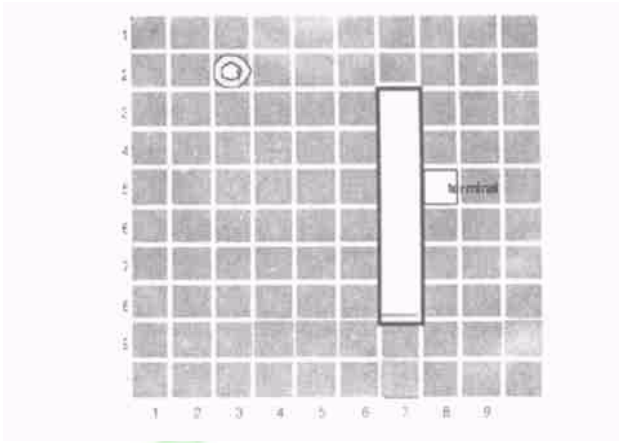

图6 仿真环境

Evironment of simulation

4.2 仿真结果

我们采用 $TD(\lambda)$学习中的在线 Sarsa 算法[6],$V_t(s) = \max_{a \in A} Q(s, a)$, S 为微直升机当前的状态,a 为当前状态可以选择的行为.我们对仿真中的参数进行设定:$\alpha_t(s) = 0.3$, $\gamma = 0.9$, $\lambda = 0.9$,仿真步长为100步,实验次数为50.替代传导径迹和累积传导径迹的仿真结果分别如图7,8所示.替代传导径迹增强式学习只需要21次实验就收敛于最优解,而累积传导径迹增强式学习需要34次实验.仿真结果表明:替代传导径迹的增强式学习适合于微直升机行为控制,并且收敛性很好.

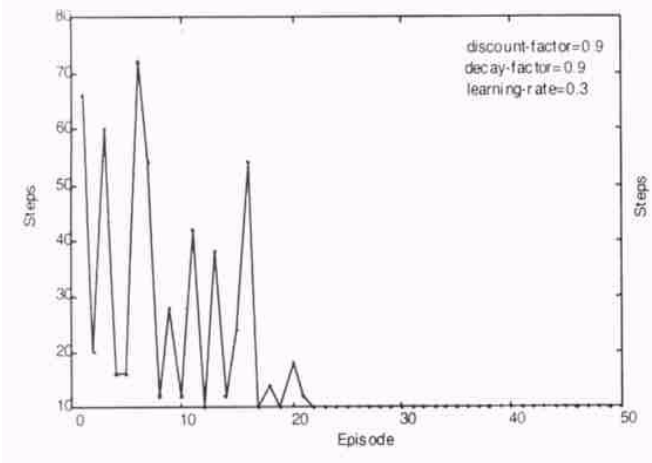

图7 替代传导径迹的增强式学习

Fig.7 Reinforcement learning with replacing eligibility traces

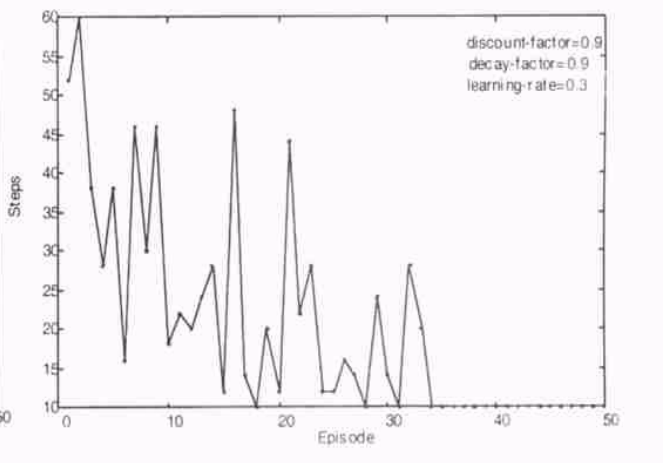

图8 累积传导径迹的增强式学习

Fig.8 Reinforcement learning with accumulating eligibility traces

4.3 存在的问题

仿真环境中,微直升机的状态是离散的,我们采用一种查表(look-up table)的机制,其占用的空间为$|S| \times |A|$,当状态空间很大的时候,查表机制不能满足需要.

由于状态空间是离散的,每个状态和行为都独立地存储,没有任何联系,为了提高效率,就必须泛化(generalization)状态空间,使得相似状态具有相似

的行为.CMAC 可以用来泛化状态空间,但本质上是一种智能查表,状态空间特别大时效率也低,神经网络 MLPs 可以很好地解决状态空间的泛化问题[6].

5　结论(Conclusion)

自主式微直升机的研制是一项包含了多种交叉学科的高、精、尖技术,其研究在一定程度上可以反映一个国家的科学技术水平.我们充分发挥微纳米加工国防重点实验室的技术和设备的先进性,利用目前处于国际领先水平的微马达这一科研成果,开展以微直升机实用化为目标的研究.为了实现微直升机的自主飞行,我们已经完成微直升机的悬停和飞行控制研究,并且采用替代传导径迹的增强式学习来实现微直升机的学习控制,仿真结果验证了方法的有效性.我们在今后将注重研究控制算法的效率和实时性,并且将采取 FPGA 来硬件实现控制算法,达到实时要求.为了提高通信能力,我们采用 Bluetooth 来实现微直升机的通信,芯片小且能耗低,适合微直升机对能源的要求.

下阶段我们将适当放大微飞机的尺寸以提高承载能力,可以装配微陀螺仪定位和微型传感器进行信息采集,在总装备部的支持和其他院校及研究所的协助下,研制出可以实地飞行的直升机,进一步缩小同国外的差距.

参　考　文　献(References)

1　Xiao Y L, Zhang C. Design and fabrication of 2mm-Diameter electromagnetic micromotor-based micro-helicopter[J]. High Technology Letters, 2000,12:77 ~ 79

2　Nilsson N. Logic and artificial intelligence[J]. Artificial Intelligence, 1991,47:71 ~ 87

3　Brooks R A. A robust layered control system for a mobile robot[J]. IEEE Journal of Robotics and Automation, RA-2(1), 1986:14 ~ 23

4　Singh S P. Reinforcement learning with replacing eligibility traces[J]. Machine Learning Journal, 1996,**22**:123 ~ 158

5　Xiao Y L. Study and Degign on a Centimenter Sized micro-rotorcraft[D]. China:Shanghai Jiaotong university, 2001

6　Rummery G A. Problem Solving with Reinforcement Learning[D]. British:Cambridge University Engineering Department, 1995

作者简介

杨玉君(1975 ~),男,博士生.研究领域为多智能体系统,机器学习,智能控制.

程君实(1939 ~),男,教授,博士生导师.研究领域为智能控制,机器人.

陈佳品(1960 ~),男,教授.研究领域为机器人控制,微小型移动机器人.