

基于分层 POMDP 的智能轮椅行为控制方法^①

陶 永^② 王田苗 魏洪兴 陈殿生

(北京航空航天大学机械工程及自动化学院 北京 100191)

摘 要 针对部分可观察马尔可夫决策过程(POMDP)模型在智能轮椅控制上的应用因模型参数难以确定而受到限制的问题,将 POMDP 模型决策过程中的动作空间分成一系列较小动作子集,进而提出了一种基于分层 POMDP 模型的智能轮椅行为控制方法。该方法综合了动作的不确定性和状态的部分可观测性,通过对环境的观测和信息的采集,得到不确定环境下的最优策略选择,进而选择相应的最优动作,从而提高了动作的执行效率。在室内家居环境下进行的交互任务与导航控制试验以及对实验结果进行的分析验证了这一方法的实时性、有效性和可靠性。

关键词 智能轮椅, 行为控制方法, 分层的部分可观测马尔可夫决策过程(POMDP)模型

0 引 言

具有自主避障、导航及人机交互等功能的智能轮椅是一种助老助残服务机器人,能够大大提高老年人和残疾人的生活质量,因而成为国内外研究的热点,目前国外研究出的智能轮椅有 Rolland III^[1], Sharioto^[2], RobChair^[3], Senarios^[4], AHM^[5], Mister Ed^[6], Wheellesley^[7]和 NavChair^[8]等。

老年人/残疾人的特点及智能轮椅使用的环境决定了安全导航、人机交互等是智能轮椅的重要组成部分。由于智能轮椅所处环境的不确定性,且用户对智能轮椅的操作行为通常是随机和不确定的,因此,如何处理智能轮椅在导航、人机交互过程中的不确定性与部分可观测性,如何规划系统的最优选择与动作,是当前智能轮椅控制方法研究的难点。但这个问题的研究并没有取得实质性进展,文献[8]只对智能轮椅控制系统的自主性和导航安全性进行了研究,而没有考虑用户对智能轮椅的控制与交互;文献[9]只对智能轮椅控制方法进行了深入的综述;文献[10]提出了完全通过用户的控制实现智能轮椅的导航和交互的设想,但认为要求用户具有连续操作及较强的规划能力,不适合老年人使用。部分可观察马尔可夫决策过程(partially observable Markov decision process, POMDP)模型是最广泛用于解决不

确定环境下规划问题的模型之一^[11],该模型在连续和部分可观测的环境中,通过环境的观测做出最优的控制进而规划相应的动作,但在智能轮椅控制的具体应用过程中较难确定模型的参数,从而限制了其应用。本文将 POMDP 模型决策过程中的动作空间分为一系列较小动作子集,提出了一种基于分层 POMDP 模型的行为控制方法,该方法通过对环境的观测和信息的采集,得到不确定环境下的最优策略选择,进而选择相应的最优动作,提高了智能轮椅在室内部分可观测环境下的决策效率,并通过实验验证了所提方法的实时性和有效性,为智能轮椅的应用提供了一种低成本、高可靠性的控制方法解决方案。

1 POMDP 模型

POMDP 模型结构由一个 6 元组 $[S, A, T, R, Z, O]$ 组成^[12]。这里,类似地定义 POMDP 模型的 6 元组如下:

- (1) 内部状态 S : 表示智能轮椅所处室内家居下的环境状态有限集合。
- (2) A : 表示智能轮椅所执行动作的有限集合。
- (3) 观察 Z : 表示观测值的有限集合。
- (4) $T: S \times A \rightarrow \Pi(s)$: 状态转移函数,表示智能轮椅在某一状态 s 执行动作 a 后并得到观测值 z , 下

① 863 计划(2007AA041701, 2007AA041702)资助项目。

② 男,1979 年生,博士,研究方向:助老助残服务机器人;联系人, E-mail: taoyong@me. huaa. edu. cn, taoyong1979@126. com (收稿日期:2009-04-28)

一个状态为 s' 的转移函数,用 $T(s, a, z, s')$ 表示。

(5) $R: Z \rightarrow R$: 奖励函数,表示得到观测值 z 所获得的奖赏,用 $R(z)$ 表示。

(6) $O: S \times A \rightarrow \Pi(Z)$: 表示智能轮椅执行动作 a 后,处在状态 s' 时得到状态值 z 的概率,等于 $p(z | s', a)$ 。

由于 POMDP 模型中真实状态是不确定的,所以,本文引入了信念状态 $B(S)$,表示和真实状态相同的概率。图1为 POMDP 模型与环境的交互过程,其主要包含两部分:

(1)状态估计器:接收环境的观测量,最新采取的动作及前面一个信念状态,进而产生更新后的信念状态;

(2)最优选择:输入为当前的信念状态,输出为将要采取的动作;在下一步的动作中采取最优步骤,使得奖励 R 最大化^[13]。

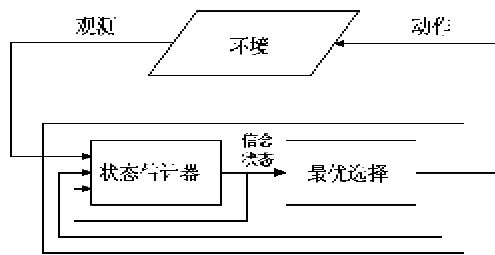


图1 POMDP 模型

2 基于分层 POMDP 模型的行为控制方法

2.1 分层 POMDP 模型

分层 POMDP 模型的提出是通过将动作空间分为更小的动作子集实现的。由于状态部分可观测,基于分层动作集的分层 POMDP 模型的计算量降低,在智能轮椅实际控制应用中能够较快地确定模型参数。图2所示为分层动作集的基本组成及智能轮椅控制方法的流程。

一个分层的动作空间呈树状,树叶代表整个 POMDP 模型中具体的动作,即每个动作 $a \in A_0$ 须由至少一个“树叶”表示,其中 A_0 为简单动作集;在内部的每个节点,都存在一个抽象动作^[14]。

通过设置分层的动作集,分层 POMDP 模型实现了用一系列小的 POMDP 模型替代原有的一个整体的较大的 POMDP 模型,从而有效降低了计算的复杂性和计算量。与一系列小的 POMDP 模型相对应,设置了一系列对应的局部最优策略。

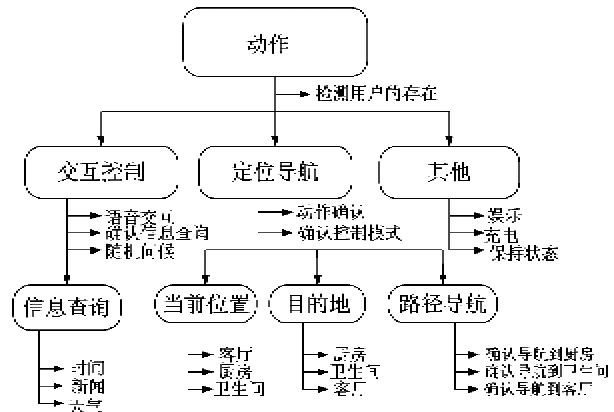


图2 分层的动作空间

在分层动作集中的每个内部动作节点 $\bar{a}_i$ 都对应一个子任务 P_i , 该子任务定义如下:

S_i : 状态矢量空间,与原有的整个动作空间 S 相对应。

O_i : 观测值空间,与原来整个观测值空间 O 相对应。

A_i : 动作子集,包含树状图中的子节点 $\bar{a}_i$ 。

π_{P_i} : 局部最优策略,与动作子集 A_i 相对应。

分层动作集包含了一系列单独的 POMDP 子任务,从而对不同子任务的最优策略分别进行优化决策判断,分别得到相对应的局部策略,提高了控制效率。

2.2 状态和观测值特征提取

分层 POMDP 模型通过简化的动作空间,完成对应的子任务从而降低了进行最优决策判断所需计算的复杂性。在此基础上,提出了状态和观测值的提取,进一步降低计算量;通过较小集合的表示,所定义的状态集和观测值集得以缩减。

状态集和观测值集的提取与分层 POMDP 模型的参数紧密相连。子任务 P_i , 动作集 A_i , 状态集 S_i 和观测值集 O_i , 对应状态集的状态特征为

$$X_i = \{X_1, X_2, \dots, X_m\} \quad (1)$$

对应观测值集的特征为

$$Z_i = \{Z_1, Z_2, \dots, Z_n\} \quad (2)$$

假设状态特征集合被分为两个不同的子集 X_+ 和 X_- , 观测值特征集合也对应地分为两个子集 Z_+ 和 Z_- 。此时,如果 $\forall a_j \in A_i$, 状态特征子集 X_+ 和观测值子集 Z_+ 与子任务 P_i 不相关:

$$R(X_+, X_-, a_j) = R(X_-, a_j) \quad (3)$$

图3所示为一个动态信念网络中的各状态子集和观测值子集的约束关系,其中状态特征子集 X_+

对奖励函数 R 没有影响,进而对子集 X_- 所对应的状态和观测特征子集也没有影响,可见,状态特征 X_+ 和观测值特征 Z_+ 对函数的值没有影响,因而可以忽略。其中,给定子任务的所有动作都须满足上述的特征对应关系。

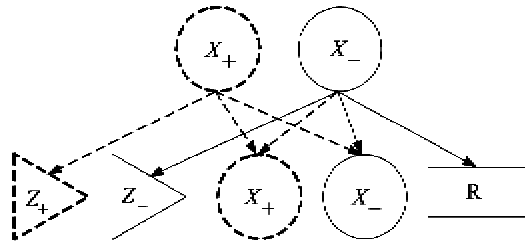


图3 动态的信念网络

在分层动作空间模型中,对每个子任务分别进行特征提取、缩减状态值集合和观测值集合,不影响对动作进行分解后的决策优化^[15]。状态值特征和观测值特征的提取是在进行动作分解后自然进行的结果,且动作空间分解和状态与观测值特征提取是可选的。对同一个结构在进行较好的动作空间分解后,可进行充分的状态和观测值特征提取,从而进一步降低计算量。

2.3 智能轮椅行为控制方法

本节提出了一种室内家居环境下的基于分层 POMDP 模型的智能轮椅行为控制方法,所研究的智能轮椅面向行动不便的老年人并提供信息交互及辅助导航的任务。

(1) 智能轮椅任务描述

本文主要研究智能轮椅的辅助导航、信息交互及提醒等功能。智能轮椅接收到手柄输入或语音控制命令后,根据行为控制系统做出路径规划决策及交互反应。在老人的日常生活、社会交往和锻炼过程中,特定的导航任务需要智能轮椅从室内的一个位置移动到另一个位置,智能轮椅还为用户提供信息交互服务,如重大社会事件的新闻、电视节目预告等,主要通过语音和屏幕输出进行交互。

(2) 基于分层 POMDP 的控制方法

针对上述任务,智能轮椅控制方法中包含一系列的数值状态特征。如表 1 和表 2 所示,智能轮椅行为控制方法包含 10 个方面的状态与观测值特征。

智能轮椅通过多种模式获得观测值,列出的观测值通常包含复杂的传感器信息。基于超声传感器和 RFID 标签,对传感器数据进行处理后,与地图对照以确定智能轮椅所处的已知地标位置。语音交互和触摸屏输入作为上述传感器所感知信息的冗余通

表 1 导航和交互任务的状态

状态特征	数值
定位	客厅,厨房,卫生间
用户是否存在	是,否
用户导航目标	厨房,卫生间,客厅,其他,无
用户操作模式	手柄输入,半自主,全自主
查询信息内容	时间,新闻,天气,无

表 2 导航和交互任务的观测值

观测值特征	数值
手柄输入	前进,后退,左转,右转,左前,左后,右前,右后,无
语音命令	是,不,前进,后退,左转,右转,停止,现在几点了,今天的天气怎样,今天的新闻
触摸屏输入	时间,天气,新闻,确定,取消,空,前进,后退,左,右,停止
RFID & 超声传感	在厨房,在卫生间,在客厅,在走廊
信息输出	时间,新闻,天气,无

RFID: radio frequency identification,无线射频识别

道,保证智能轮椅操作的可靠性和安全性。

室内家居环境不同的观测值对应了智能轮椅不同的动作,主要分为三类:

- 交互任务 = {检测用户的存在,播报当前时间,确认控制模式,播报天气,确认导航到厨房,确认导航到卫生间,确认导航到客厅,确认信息查询,确认时间查询,确认天气查询,确认执行}
- 导航任务 = {前往厨房,前往卫生间,前往客厅,前进,后退,左转,右转,停止}
- 其他 = {娱乐,充电,无}

上述单独的动作包含在智能轮椅预先定义的一系列操作中,例如,前往客厅:定位→地图中的目的地→路径规划→避障→移动到客厅,该导航任务中的动作序列通过运动规划转化为电机转动指令,在此过程中通过动态编程实现从智能轮椅当前位置到目的地的路径规划及基于超声传感器和 RFID 标签的避障。由于传感器信息的局限性,用户语音和触摸屏输入信息作为智能轮椅导航和交互任务的补充,在信息交互的决策策略中包含了同步的语音输出和液晶显示屏显示。

分层 POMDP 模型的奖励参数根据智能轮椅当前动作的资源消耗及对状态的影响程度进行设定。例如:

- 达到一个用户期望的目标时,进行正向奖

励:

$$R(a_i = \text{ConfirmGetWeather}) = +50$$

$$\text{if } s_i(\text{User-Info-Goal}) = \{\text{Weather}\}$$

- 当执行错误的动作时,进行负的奖励(惩罚):

$$R(a_i = \text{GotoKitchen}) = -100$$

$$\text{if } s_i(\text{User-Nav-Goal}) = \{\text{T, V Room}\}$$

该奖励函数满足基于动作分层的每个子任务的局部反馈要求。

基于分层 POMDP 的模型,提高了智能轮椅行为控制方法的上层决策效率,且最优策略可离线计算;智能轮椅的主控制器主要对状态和对应动作进行控制。智能轮椅通过信息感知动作对用户的意图进行预测和确认。

3 实验验证

本文搭建了如图 4 所示的智能轮椅试验平台,轮椅的尺寸为 $1.2\text{m} \times 0.7\text{m}$,最大运行速度 15km/h 。控制系统由 1.5GHz VIA 嵌入式处理器、 1G 内存、 2G CF 存储卡组成,运行 Windows XP 精简版操作系统,并使用 VC 6.0 进行控制编程。智能轮椅的环境感知与避障部分包含分别安装在前面和后面的 3 个超声传感器和 2 个红外传感器,室内地板上铺设的 RFID 标签的间隔为 1m ,通过智能轮椅底部的 RFID 读卡器读取不同位置的标签信息,用于室内定位及导航。用户可通过操纵手柄、语音命令及触摸屏发出控制命令,通过串口和电机驱动器进行通讯,控制电机执行动作。



图 4 智能轮椅试验平台

为了验证本文提出的基于分层 POMDP 的模型行为控制方法的有效性,对智能轮椅在如图 5 所示的家庭室内环境中进行了信息交互与导航控制的试验,即智能轮椅根据用户需求提供信息及以特定的速率到达用户期望的室内目标位置。房间内可能的

目标地址使用灰色方框表示,形成了目的地序列 $\{s1, s10, s14, s16, s19, s20, s24\}$ 。

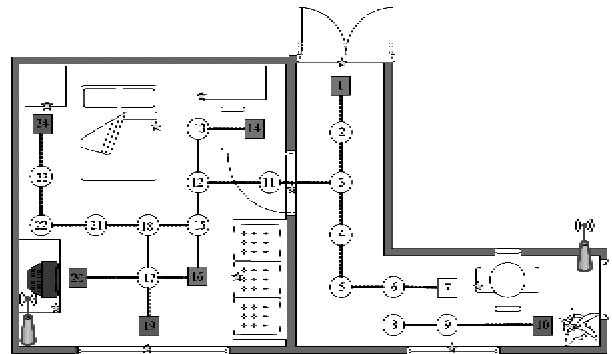


图 5 家庭室内环境下的试验

实验过程中,进行了通过语音交互和触摸屏输入输出的信息交互实验,共计进行了 69 次关于天气、新闻和时间的信息查询实验,其中 65 次正确试行,3 次由于背景噪声的干扰造成交互失败,1 次由于触摸屏的校准造成实验失败,成功率为 94.2% 。进行了室内导航试验,6 位不同的测试者对智能轮椅进行了 58 次试验。图 6 所示为其中一次导航实验:移动通过走廊,导航通过房门,避开障碍物(椅子),成功到达目的地后通知用户。在试验中,所研究的状态变量为一个二维向量 $s = (x, y)$,即用户的意图为智能轮椅以 y 的速率到达家庭房间的 x 位置。用户命令集由 10 个命令组成,包含前进、后退、左转、右转、左前、左后、右前、右后、加速、减速、停止。通过用户操作手柄、语音命令和触摸屏的输入及超声、红外和 RFID 传感器得到每个状态的观测值;观测值根据各个子任务的进行依次获取,内部信念状态随之进行更新,进而根据局部最优策略的选择,执行对应的子动作集合,达到新的状态;该处理过程反复进行,直到到达导航任务的最终目标点。



图 6 室内导航试验

在导航实验中,55次成功地完成了导航任务,2次由于未到达指定目的地而导航失败,1次由于与障碍物发生了碰撞而导航失败,成功率为94.8%。通过实验,验证了本文提出的基于分层 POMDP 的智能轮椅行为控制方法在室内环境下的有效性与可靠性。

4 结 论

本文提出了一种基于分层 POMDP 模型的智能轮椅行为控制方法,将 POMDP 模型的动作空间分为一系列小的动作子集从而提高最优决策的效率,并对智能轮椅的任务、状态值与观测值的特征提取、最优决策的优化判断进行了详细的阐述,最后在室内环境下的实验,验证了本文所提出的分层 POMDP 控制方法的有效性与可靠性。

参考文献

- [1] Mandel C, Huebner K, Vierhuff T. Towards an autonomous wheelchair: Cognitive aspects in service robotics. In: Proceedings of the Towards Autonomous Robotic Systems, London, United Kingdom, 2005. 165-172
- [2] Vanhooydonck D, Demeester E, Nuttin M, et al. Shared control for intelligent wheelchairs: an implicit estimation of the user intention. In: Proceedings of the International Workshop on Advances in Service Robotics, Bardolino, Italy, 2003. 176-182
- [3] Pires G, Araujo R, Nunes U, et al. Robchair—a powered wheelchair using a behaviour-based navigation. In: Proceedings of the International Workshop on Advanced Motion Control, Coimbra, Portugal, 1998. 536-541
- [4] Beattie P, Bishop J M. Localization of the senario wheelchair. In: Proceedings of the MobiNet Symposium Mobile Robotics Technology for Health Care Services, Athens, Greece, 1997. 287-293
- [5] Bourhis G, Moumen K, Pino P, et al. Assisted navigation for a powered wheelchair. In: Proceedings of the IEEE International Conference on Systems, Man Cybernetics, Le Touquet, France, 1993. 553-558
- [6] Connell J, Viola P. Cooperative control of a semi-autonomous mobile robot. In: Proceedings of the IEEE International Conference on Robotics Automation (ICRA), Cincinnati, USA, 1990. 1118-1121
- [7] Yanco H A, Mittal V, Yanco H, et al. Wheelchair, a robotic wheelchair system: Indoor navigation and user interface. In: Proceedings of the Lecture Notes Artificial Intelligence: Assistive Technology Artificial Intelligence, New York, NY, USA, 1998. 256-268
- [8] Levine S, Bell D, Jaros L, et al. The NavChair assistive wheelchair navigation system. *IEEE Trans Rehabi Eng*, 1999, 7(11): 443-451
- [9] Simpson R. Smart wheelchairs: A literature review. *Journal of Rehabilitation Research and Development*, 2005, 42: 423-436
- [10] Simpson R, LoPresti E, Hayashi S, et al. The smart wheelchair component system. *Journal of Rehabilitation Research and Development*, 2004, 41(5): 429-442
- [11] Amalia F, Panos T. Real-time hierarchical POMDPs for autonomous robot. *Navigation, Robotics and Autonomous Systems*, 2007, 55(7): 561-571
- [12] Kaelbling L P, Littman M L, Cassandra A R. Planning and acting in partially observable stochastic domains. *Artificial Intelligence*, 1998, 101:99-134
- [13] 周继恩, 刘贵全, 张春阳等. 基于内部信念状态 POMDP 模型在用户兴趣获取中的应用. 小型微型计算机系统, 2004, 25(11): 1979-1783
- [14] Martha E P, Sandra E, Sebastian T. Pearl: A Mobile Robotic Assistant for the Elderly. In: Proceedings of the American Association Artificial Intelligence Technical Report WS-02-02, Menlo Park, USA, 2002
- [15] Pineau J, Thrun S. High-level robot behavior control using pomdps. In: Proceedings of the American Association Artificial Intelligence Technical Report WS-02-02, Menlo Park, USA, 2002

The intelligent wheelchair behavior control method based on hierarchical POMDP model

Tao Yong, Wang Tianmiao, Wei Hongxing, Chen Diansheng

(School of Mechanical Engineering and Automation, Beijing University of Aeronautics and Astronautics, Beijing 100191)

Abstract

In view of the problem that the application of the partially observable Markov decision process (POMDP) model in intelligent wheel chair behavior control is limited due to the difficulty in determining the model parameters, the paper proposes a method for control of intelligent wheelchair behaviors based on a hierarchical POMDP model. The method considers the nondeterministic actions and uncertainty of the state, and decomposes the actions into a set of small ones that can be done more efficiently. The model can account for both state transition and observation uncertainty. This behavior control method based on hierarchical POMDP was implemented onboard an intelligent wheelchair in the context of interactive service and navigation tasks. Finally the course of experiments conducted in an assisted living facility demonstrated that it could autonomously provide service and information for users, and the analyses of the experimental results verified the effectiveness and real-time property of the method.

Key words: intelligent wheelchair, behavior control method, hierarchical partially observable Markov decision process (POMDP) model