DOI: 10.13245/j.hust.15S1014

基于 LMDP 的机器人任务规划建模

王雯珊 曹其新

(上海交通大学机器人研究所, 上海 200240)

摘要 针对经典规划模型和马尔可夫决策过程(MDP)模型的不足,提出了一种轻量马尔可夫决策过程(LMDP)模型.此模型在 MDP 模型上作了简化,使其既能描述实际任务中不确定性的特点,又有效降低了状态转移的分支系数,从而适用于大规模的问题.另外,利用经典规划领域的启发函数对 LMDP 问题进行初始化,能够大大加快收敛速度.最后以机器人酒吧任务为例,将此模型与基于 MDP 模型的 Prost 规划器在不同问题规模下进行对比,实验结果表明此模型能有效加快求解速度,并能够更好地适应大规模实际环境.

关键词 机器人; 任务; 规划; 建模; 马尔可夫决策过程; 不确定性

中图分类号 TP242 **文献标志码** A **文章编号** 1671-4512(2015)S1-0058-04

Modeling for robot task planning based on light-weighted Markov decision process

Wang Wenshan Cao Qixin

(Research Institute of Robotics, Shanghai Jiao Tong University, Shanghai 200240, China)

Abstract To overcome the deficiencies of the classical planning model and the Markov Decision Process (MDP) model, a light-weighted Markov decision process (LMDP) model was proposed. This is a simplification of the MDP model, retains the uncertainty feature for realistic environment, and is able to handle large-scale problems by reducing the branching factor of state transitions. Besides, the convergence time is greatly reduced by initialization methods based on heuristic functions in classical planning domain. The planning method based on LMDP model was compared with the Prost planner, which is based on MDP model. The results show that using LMDP model, the planning efficiency is improved, and the planning results better adapt the physical world.

Key words robot; task; planning; modeling; Markov decision process; uncertainty

任务规划是智能机器人的关键组成部分[1],其主要目标是根据抽象的指令,以及当前环境中机器人的状态,规划出一系列机器人能够直接执行的子任务.无论是家庭服务[2]、工业生产[3]、安防救援[4]还是外空探索[5]的自主机器人都涉及任务规划问题.对于这种任务级别的规划,在结构化的环境中最常用的是基于有限状态机的方法[6].然而针对大规模、非结构化环境,设计者很难预先定义好所有可能的状态.因此此次研究目标是针对领域无关的任务规划问题,建立一个通用而又有效的任务规划模型,使其能够应用于大规模非结构化环境的多种领域.

任务规划方法一般采用状态转移系统模型,用状态的转移来表示执行的动作的效果.在此基础上,通常须作一些假设来使问题简化.经典任务规划模型有许多比较高效的求解算法,然而由于假设过于严格,很难应对实际环境中的动态性和不确定性.针对动态不确定环境,当前广泛采用马尔可夫决策过程(MDP)模型,并用迭代方法求解.然而,求解大规模的 MDP 模型会遇到维数爆炸问题,而一些常用的状态近似方法也无法用在面向通用领域的任务规划问题上,因此 MDP 模

收稿日期 2015-06-30.

作者简介 王雯珊(1986-),女,博士研究生,E-mail: amigo@sjtu.edu.cn.

型在实际的任务规划问题中也很难应用. 本研究提出轻量马尔可夫(LMDP)模型，在 MDP 模型上作了简化，使其既能应对实际任务规划中大部分的不确定性问题，又能降低求解难度，适用于大规模的实际问题.

1　经典规划模型和 MDP 模型分析

经典任务规划模型对状态转移系统作的假设包括完全可观性、确定性、静态性、离线规划等. 在上述假设下，经典任务规划模型是一个五元组 $\Sigma=(S,A,c,I,G)$，S,A,c,I 和 G 分别表示系统的状态集合、动作集合、动作耗费、初始状态和目标状态. 规划结果用动作序列 $P=\langle a_1,a_2,\cdots,a_n\rangle$ 表示. 规划的目标是求得动作序列 P，使得系统从初始状态 I 转移到目标状态 G，且总耗费最小化.

当前的经典规划算法主要包括图规划、命题可满足方法和启发式搜索方法[7]. 但是这些方法很难在实际应用中使用，主要的原因是经典规划模型对系统所作的大量简化，使得其难以完成实际任务. 例如对于确定性简化，经典模型假设动作完成以后，状态一定会发生相应的改变. 但是在实际情况中，动作执行经常会发生错误和失败，动作执行的效果也不是固定的，如移动机器人随着电力下降更容易发生错误等.

经典自动规划方法对模型的假设太过严格，MDP 模型对经典模型中的假设做出了扩展，主要解决了不确定性问题，并且允许外部随机事件，支持在线规划.

MDP 模型用一个四元组来表示随机系统 $\Sigma=(S,A,P,R)$，S,A,P 和 R 分别表示系统状态集合、动作集合、状态转移概率和回报函数. 规划问题的解是一个策略 $\pi:S\mapsto A$，是状态到动作的映射. 这个问题是要求解一个最优策略，使得系统能够获得最大的回报值.

求解这个最优化问题的算法有动态规划方法(DP)、蒙特卡罗方法(MC)和时间差分算法(TD)等[8]. 限制这种方法用于实际的主要问题是维数爆炸.

MDP 问题的分支系数非常大，求解的复杂度随着状态变量数指数增加. 在强化学习和其他规划领域，求解大规模 MDP 模型主要采用状态近似和分层规划的方法. 然而，由于任务规划的领域无关性质，缺少状态近似和分层规划所需要的领域知识，难以取得很好的效果.

2　轻量马尔可夫模型

为了兼顾模型的实用性和求解效率，提出 LMDP 模型，使其既能满足实际任务规划中动态和不确定性要求，又能降低求解难度. 观察实际机器人执行动作，发现动作执行的结果可以简化成成功和失败两种情况，因此在状态 s 执行动作 a 后，最多只会转移到两种可能状态. 进一步，假设一个动作执行失败后，不改变原来系统的状态. 因此状态转移后的两种状态为 s 和 s'，如图 1 所示. 相比于 MDP 模型，LMDP 大大降低了状态空间的分支系数，又能符合实际规划任务的需要.

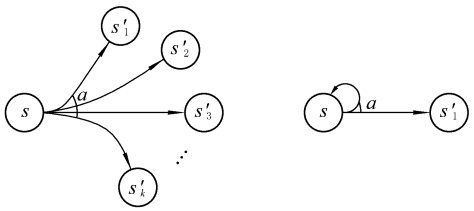


图 1　动作 a 的状态转移模型

定义　LMDP 规划问题是一个五元组 $\Sigma=(S,A,c,I,G)$，$S=\{s_1,s_2,\cdots\}$ 为有限个状态的集合；$A=\{a_1,a_2,\cdots\}$ 为有限个动作的集合，$a\in A$，$a=(n_a,p_a,e_a,s_a)$ 分别为动作的名称、前提条件、执行效果和成功概率，且 $s_a\in(0,1]$；$c:A\mapsto R_0^+$ 为动作的耗费；$I\subseteq S$ 为系统的初始状态；$G\subseteq S$ 为系统的目标状态.

LMDP 的规划解也是一个策略 $\pi:S\mapsto A$，是状态到动作的映射. 对于状态 s 定义值函数，表示从 s 出发执行策略 π 的耗费值用 $V^\pi(s)$ 表示，

$$V^\pi(s)=c(a_0)+c(a_1)+\cdots=\sum_{i=0}c(\pi(s_i)). \tag{1}$$

规划问题是要找到一个最优策略 π^*，使得耗费函数取最小值. 与 MDP 模型类似，最优的耗费函数 $V^*(s)$ 满足

$$V^*(s)=\min_{a\in A}\Big[c(a)+\sum_{i=0}P_a(s'\mid s)V^*(s')\Big]=\min_{a\in A}[c(a)+s_aV^*(s')+(1-s_a)V^*(s)], \tag{2}$$

式中 $P_a(s'|s)$ 表示在状态 s 执行动作 a 后转移到状态 s' 的概率. 最优策略 π^* 是使得耗费函数取最小值的策略，

$$\pi^*(s)=\arg\min_{a\in A}[c(a)+s_aV^*(s')+(1-s_a)V^*(s)]. \tag{3}$$

根据式(2)和(3)可得如下动态规划算法 1.

算法 1 Modified TD

Initialize cost array V arbitrarily

Repeat

$\Delta \leftarrow 0$

For each $s \in S$:

temp$\leftarrow V(s)$

$V(s) \leftarrow \min_a [c(a) + s_a V(s') + (1 - s_a) \cdot V(s)]$

$\Delta \leftarrow \max(\Delta, |\text{temp} - V(s)|)$

until $\Delta < \theta$(a small positive number)

$\pi(s) = \text{argmin}_{a \in A} [c(a) + s_a V(s') + (1 - s_a) V(s)]$

在动态规划算法中,按第 6 行迭代修改耗费值,最终耗费值会收敛到最优值 V^*. 相比于 MDP 模型,LMDP 模型的迭代过程更简单,收敛更快.

更重要的是,LMDP 模型很容易被松弛成一个确定性问题,利用经典规划模型的求解方法,对耗费值进行初始化,取代上述算法中第 1 行的随机初始化. 这个初始化信息能够大大加快收敛速度,如算法 2 所示.

算法 2 Value Initialization

def heuristicSearch(s, a)

$s' \leftarrow$apply(s,a)

if $s' \subseteq G$ then return $c(a)$

bestcost$\leftarrow \infty$

for $a' \in A$ with best heuristic value:

futurecost$\leftarrow$heuristicSearch(s',a')

if futureCost$<$bestCost

then bestcost$\leftarrow$futurecost

return $c(a)$+bestcost.

3 实验结果

通过机器人酒吧任务,验证 LMDP 的有效性. 在机器人酒吧任务中,机器人根据用户的点单,将酒水饮料送到用户的餐位. 实验系统由一系列机器人组件组成,包括负责人机交互的仿人机器人、负责运送饮料的移动机器人和负责识别和抓取机械臂. 图 2 显示了酒吧任务的规划结果以及执行过程,图 2(a)显示了仿人机器人通过语音交互获取订单;在图 2(b)中移动机器人运动到吧台负责运送饮料;图 2(c)和(d)分别显示了机械臂抓取饮料并放置在移动机器人上的过程.

如前所述,对 LMDP 问题的高效求解很大程度上得益于松弛化以后的初始化方法. 如图 3 所示,利用经典规划领域的 FF 启发函数[9],对

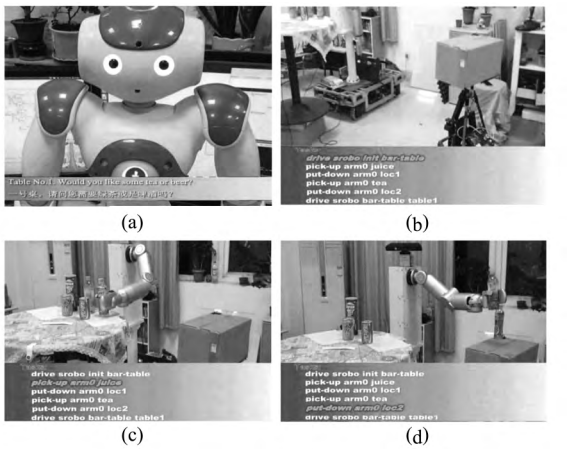
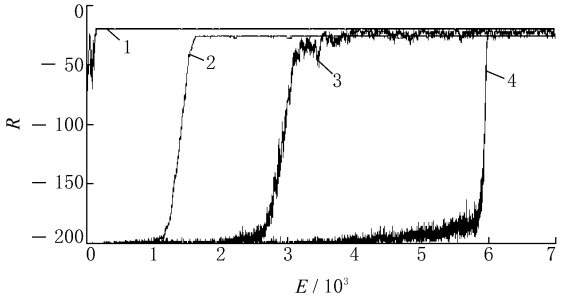


图 2 基于 LMDP 的机器人酒吧任务规划与执行

LMDP问题进行初始化. 相比于不使用初始化方法的 TD(λ)[8],回报值(R)收敛速度提高近 100 倍. 这是因为没有初始化信息的方法在迭代初期经历了类似于随机搜索的过程,而启发函数提供的初始化信息消除了这一过程.

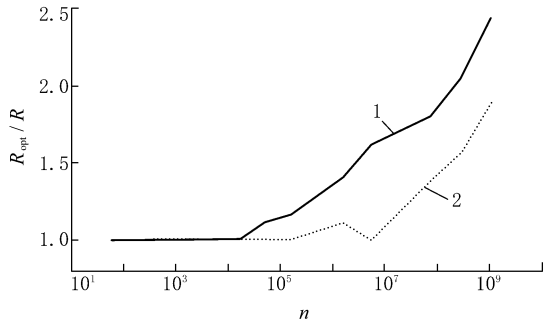


1—TD(0.0)FF 启发函数; 2—TF(0.9)未初始化;
3—TF(0.5)未初始化;4—TF(0.0)未初始化.

图 3 用确定化模型提供初始化信息的收敛速度对比

有些 MDP 规划器(如 Prost)也使用了松弛成确定性问题以后进行初始化的方法[10]. LMDP 模型低分支系数的特点使得其更容易松弛成确定性问题,并能够使用经典规划领域的启发函数提供更加有效的初始化信息. 对比了 MDP 模型规划方法和本 LMDP 模型规划方法在 10 个不同规模(n)问题上的结果,如图 4 所示. 两种方法对 10 个问题各迭代 1 000 次进行求解,曲线显示了最优回报值(R_{opt})与规划器求得的策略的回报值(R)的比值. 基于 LMDP 模型的方法在小规模问题上都能收敛到最优策略,并且对于大规模问题也能得到更好的规划结果.

面向与领域无关的机器人任务规划问题,建立了一个通用而又有效的任务规划模型. 此模型既能满足实际问题的不确定性的要求,又有效降低了状态转移的分支系数,使其能够应对大规模的问题. 实验表明 LMDP 模型能有效加快求解速



1—MDP 模型；2—LMDP 模型.

图 4　LMDP 模型规划方法与 MDP 模型规划方法的对比

度，并能够更好地适应大规模实际环境.

参　考　文　献

[1] Galindo C，Fernández-Madrigal J A，González J，et al. Robot task planning using semantic maps[J]. Robotics and Autonomous Systems，2008，56(11)：955-966.

[2] 宋沐民，路飞，陆娜，等. 智能空间下基于分层任务网络的服务机器人任务规划[J]. 控制理论与应用，2014，31(7)：901-907.

[3] Radaschin A，Voda A，Minca E，et al. Task planning algorithm in hybrid assembly/disassembly process[C]// Proc of 14th IFAC Symposium on Information Control Problems in Manufacturing . Bucharest：IFAC，2012：571-576.

[4] 梁志伟，沈杰，杨祥，等. RoboCup 机器人救援仿真中基于拍卖的任务分配算法[J]. 机器人，2013，35(4)：410-416.

[5] Sherwood R，Mishkin A，Chien S，et al. An integrated planning and scheduling prototype for automated Mars rover command generation[C]// Proc of Sixth European Conference on Planning. [s. l.]：AAAI，2014：292-302.

[6] Loetzsch M，Risler M，Jungel M. XABSL-a pragmatic approach to behavior engineering[C]// Proc of IEEE/RSJ International Conference on Intelligent Robots and Systems. Beijing：IEEE，2006：5124-5129.

[7] Ghallab M，Nau D，Traverso P. Automated planning：theory and practice[M]. Singapore：Elsevier，2004.

[8] Barto A G. Reinforcement learning：An introduction[M]. London：MIT press，1998.

[9] Hoffmann J，Nebel B. The FF planning system：fast plan generation through heuristic search[J]. Journal of Artificial Intelligence Research，2001，14：253-302.

[10] Keller T，Helmert M. Trial-based heuristic tree search for finite horizon MDPs[C]// Proc of 23rd International Conference on Automated Planning and Scheduling. Rome：AAAI，2013：135-143.