并行概率规划综述*

饶东宁[1], 李建华[1], 蒋志华[2†], 赵淦森[3]
(1. 广东工业大学 计算机学院, 广州 510006; 2. 暨南大学 信息科学技术学院 计算机科学系, 广州 510632; 3. 华南师范大学 计算机学院, 广州 510631)

摘 要: 自动规划针对特定领域的特定问题,生成一个由可应用动作构成的规划。经典规划中的动作效果是确定的,且在每个时间步内只能执行一个动作。但在实际问题中,动作的效果往往是不确定性的,且动作的执行具有并发性。因此,并行概率规划(parallel and probabilistic planning,PPP)被提出,并且它的应用前景正在引起规划研究学术圈的关注。有鉴于此,对其进行综述。具体内容包括定义 PPP 领域、问题和规划解,介绍其描述语言、基准领域及规划器,并对其中两个有代表性的规划器进行实际测试。实验表明在求解效率方面测试结果与比赛结果基本一致,但部分规划器的求解规模与竞赛不完全一致。这可能是比赛中的某些未开源代码或手工干预得到的。

关键词: 自动规划; 并行概率规划; 国际规划比赛; 规划领域; 规划器
中图分类号: TP182 **文献标志码:** A **文章编号:** 1001-3695(2016)06-1607-05
doi:10.3969/j.issn.1001-3695.2016.06.002

Survey on the parallel and probabilistic planning

Rao Dongning[1], Li Jianhua[1], Jiang Zhihua[2†], Zhao Gansen[3]
(1. *School of Computer, Guangdong University of Technology, Guangzhou* 510006, *China*; 2. *Dept. of Computer Science, School of Information Science & Technology, Jinan University, Guangzhou* 510632, *China*; 3. *School of Computer, South China Normal University, Guangzhou* 510631, *China*)

Abstract: Given a specific domain and problem, automated planning will generate plan solutions composed of applicable actions. In classic planning, actions have deterministic effects and are carried out sequentially. However, in real-world problems, the effects of an action might be non-deterministic and there might be concurrent actions. Therefore, the parallel and probabilistic planning(PPP) was proposed. The PPP has a strong application perspective and is drawing a great deal of attention in the planning community. Therefore, this paper presented a survey on the PPP. It first gave formal definitions of domains, problems and plan solutions of the PPP. Then it introduced competition languages, benchmark domains and competitive planners in the PPP. Finally, it tested two representative planners. Experiment results show that, for the efficiency, the results were similar to those in the competitions. However, the scalability was different. It might be caused by some unpublished source codes or some manual intervention.

Key words: automated planning; parallel and probabilistic planning(PPP); international planning competitions(IPCs); planning domains; planners

0 引言

自动规划(automated planning,以下简称规划)[1~3]是人工智能的重要分支。规划问题由领域描述、初始状态和目标状态组成,它用推理或搜索等方法寻找一个从初始状态到目标状态的有效动作序列。自动规划的理论与技术从 20 世纪 60、70 年代开始研究,至今已有四五十年的历史,研究范围从基于简单模型的经典规划到越来越复杂的非经典规划。特别地,在最近十五年,规划问题从表示语言到求解技术已经有了极大的进步,其中向应用靠拢是最大的驱动力。不少研究者尝试在车辆调度、交通灯控制、航空任务等实际领域应用规划的现有技术[4~7],但是,基于八大假设[2,3]的经典规划的影响力根深蒂固,使得应用难以顺利开展。如规划应用于网络服务合成时[8],网络服务(动作)具有不确定性,但经典规划假设动作效果确定;再如,在实际问题中动作可并发执行,但在经典规划中动作必须是串行的。可见,经典规划的假设不满足解决实际问题的需求。复杂规划领域成为目前规划研究学术圈的一个主要的研究方向。

并行概率规划(PPP)是最新出现的复杂规划问题之一。它具有并行规划[9]和不确定规划[10]的特征,允许非互斥的动作并发执行,动作的前提和效果满足全局性约束,放松了经典规划的五大假设。PPP 的出现使得基于效果模型的传统规划描述语言不再适用,因此需要提出能够充分描述全局性约束的新语言。新语言应基于因果规则,能够用统一的框架描述在相

收稿日期: 2015-09-08 **修回日期:** 2015-11-03 **基金项目:** 中央高校基本科研业务费专项资金资助项目(21615438);广州市云计算安全与测评技术重点实验室开放基金资助项目(GZCSKL-1408)

作者简介: 饶东宁(1977-),男,广东兴宁人,副教授,博士,主要研究方向为智能规划;李建华(1989-),男,硕士研究生,主要研究方向为智能规划;蒋志华(1978-),女(通信作者),副教授,博士,主要研究方向为智能规划(tjiangzhh@jnu.edu.cn);赵淦森(1977-),男,教授,博士,主要研究方向为云计算.

同或不同状态下的流文字的各种依赖关系,而不是把全局约束分散到各个动作模型中。PPP 的出现也使得需要提出新的规划求解方法。由于并行性和不确定性的组合,状态搜索空间在初始的几步内就会变得非常庞大,所以传统的搜索策略和启发式估值效率低下。新的求解方法有待于基于概率演算或模拟。规划学术圈对 PPP 非常重视,在各个层面都投入了关注,包括建立新的比赛、提出新的语言以及开发新的规划器。本文对 PPP 的应用前景也非常看好。尽管它现有的求解技术还不够成熟,但是假以时日一定会有很大的改进,再加上时态规划[4]、最优规划[5]等其他特征的融入,PPP 必定会成为将规划方法推向实际应用的桥梁。

国际规划大赛(international planning competitions,IPCs)是规划发展的重要环节,而近几年出现了专门针对 PPP 的比赛。IPC 始自 1998 年,两到三年一届,其发布的基准领域是规划器比较的基准。近两次 IPCs 推出概率规划比赛(international probabilistic planning competitions,IPPCs)并提供了 PPP 基准领域。IPC 使用规划领域定义语言(planning domain definition language,PDDL),它从早期的动作描述语言发展到 PDDL 3.0[11]及其概率版本(probabilistic planning domain definition language,PPDDL)[12]。IPPC-2011 提出了 PPP 的描述语言(relational dynamic influence diagram language,RDDL)[13]。

规划器是规划比赛的参与者,优胜者所用方法可能成为规划研究里程碑。如 IPC-1998 冠军 Blackbox[14]引入规划图[15]成为规划领域特有数据结构;再如基于前向搜索的系统(FF[16]、LAMA[17]、SymBA*-2[18]等)囊括了 IPC-2000、IPC-2002、IPC-2008、IPC-2011 和 IPC-2014 等确定性规划比赛冠军。在概率规划方面,FPG[19]在 IPC-2006 中获奖,它基于策略空间中的随机局部搜索;在 IPPC-2011 和 IPPC-2014 上 SPUDD[20]和 PROST[21]分获冠军,它们分别基于动态规划和蒙特卡洛采样。

1 并行概率规划定义

1.1 不确定规划

不确定规划研究始于 20 世纪 70 年代[10]。其中不确定性包括初始状态的不确定性和动作效果的不确定性。一个规划领域可以对应多个具体的规划问题。每个规划问题 Π 由规划领域描述 Σ、初始状态 I 和目标条件 G 构成,即 $\Pi=\langle\Sigma, I, G\rangle$。不确定动作是具有多组效果的动作,且这多组效果具有共同的前提。在本文中,为了区分方便,一个具有 n 组效果的不确定动作记为 $\tilde{a}=\langle \mathrm{pre}(\tilde{a}), \mathrm{eff}_i(\tilde{a})\rangle (1\leqslant i\leqslant n)$,其动作效果为 $\mathrm{eff}_i(\tilde{a})=\langle \mathrm{add}_i(\tilde{a}), \mathrm{del}_i(\tilde{a})\rangle$,序对中的元素分别为增加效果和删除效果。$\tilde{a}$ 可视为含 n 个确定动作的集合 $A_a=\{a_1,\cdots,a_n\}$,$a_i=\langle \mathrm{pre}(a), \mathrm{add}_i(\tilde{a}), \mathrm{del}_i(\tilde{a})\rangle (1\leqslant i\leqslant n)$。由于不确定哪一组动作效果会发生,在状态 s 中应用 $\tilde{a}$ 将导致多个可能的后继状态(即信念状态)。不确定规划的求解就在信念状态空间里进行。在信念状态 S_a 中应用下一个不确定动作 $\tilde{b}$,可得下个信念状态 $S_b=\bigcup_{s\in S_a} T(s,\tilde{b})$,如果 $\tilde{b}$ 在 s 中是可应用的,T 是转换函数。

不确定规划中,动作的不确定性带来了系统转换的不确定性。不确定规划中概率规划占主要地位,其中的转换是个转移概率分布。总之,它有三方面特征:a)动作的效果是一个集合;b)执行动作后系统的演化是一个转移分布;c)在不可观测的情况下,系统处在由多个具体状态组成的信念状态中。

1.2 并行规划

并行规划在一个时间步可执行多个动作[9],其优点包括:相互独立的动作可并行执行,不用考虑动作间顺序;规划的总时间步大大减少。并行规划要考虑动作之间的互斥关系。不管是由于冲突还是需求竞争,互斥的动作不能同时执行,因为这会使得世界模型处于不稳定的状态。如果一个动作集合中的所有动作都是非互斥的,则称这个动作集合为非互斥的动作集合。动作集合 A' 在状态 s 下是可用的,当且仅当对任意动作 $a\in A'$,其前提 $\mathrm{pre}(a)\subseteq s$ 成立。在并行规划中,在某状态下应用一个非互斥的动作集合,其效果等同于各个非互斥动作效果的累积。由于各个动作既不冲突又无需求竞争,所以无论应用的顺序如何,其最后结果都是一样的。

动作的并行性带来了同时执行动作的前提和效果判断的困难。在动作效果确定的假设下,目前已有的工作将注意力集中在同时执行的动作互斥的判断上。在前提均被满足且不互斥的情况下,动作可同时执行。并行性带来如下三方面的特征:a)同一规划步中的动作不能互斥;b)执行动作后系统的演化是确定性的;c)规划解的最优性可以从多个方面定义,如总的规划长度最短、总的动作数最少或者使得某种代价最低。

1.3 并行概率规划

PPP 是概率规划和并行规划的结合,它具有动作可并行、动作效果由全局约束决定、效果具有不确定性等性质。在给出定义前本文将动作互斥定义扩展到不确定动作上。

定义 1 不确定性动作的互斥。对任意两个不确定性动作 $\tilde{a}$(即 $A_a=\{a_1,\cdots,a_n\}$)和 $\tilde{b}$(即 $A_b=\{b_1,\cdots,b_m\}$),若存在 i $(1\leqslant i\leqslant n)$ 和 $j(1\leqslant j\leqslant m)$ 使 $a_i\in A_a$ 和 $b_j\in A_b$ 互斥的,则 $\tilde{a}$ 和 $\tilde{b}$ 互斥。

定义 1 表明,在两个不确定动作中,若有任意一组分动作互斥,则这两个不确定动作也互斥。在无观测情况下任何一组分动作的组合都可能发生,因而分动作的互斥可能导致不确定动作互斥。进一步地,在一个不确定动作集合中,若任意两个不确定动作是非互斥的,则该不确定动作集合才是非互斥的。

定义 2 并行概率规划领域。一个并行概率规划领域是一个四元组 $\Sigma_{\mathrm{PNDP}}=\langle P,S,\tilde{A},\tilde{T}\rangle$,其中:

a)P 是一个非空有限的命题集合;

b)$S\subseteq 2^P$ 是状态集合;

c)$\tilde{A}=\{\tilde{a}\mid \tilde{a}=\langle \mathrm{pre}(\tilde{a}), \mathrm{eff}_i(\tilde{a})\rangle\}$ 是不确定动作集合;

d)$\tilde{T}: S\times 2^A\to 2^S$ 是状态转换函数。设 $\tilde{A}_1\in 2^A$ 是 n 个不确定动作的集合,$\tilde{A}_1=\{\tilde{a}_1,\cdots,\tilde{a}_n\}$,$\forall \tilde{a}_i,\tilde{a}_j\in\tilde{A}_1 (1\leqslant i,j\leqslant n)$ 都非互斥,动作 $\tilde{a}_i (1\leqslant i\leqslant n)$ 含 m_i 组效果,$\tilde{a}_i=\langle \mathrm{pre}(\tilde{a}_i), \mathrm{eff}_k(\tilde{a}_i)\rangle$,$1\leqslant k\leqslant m_i$,其中 $m_1,\cdots,m_n$ 均为正整数。若 $\tilde{A}_1$ 在 $s\in S$ 中可用,则 $\tilde{T}(s,\tilde{A}_1)$ 表示后继状态集合 $S_{A_1}=\{s'\mid s'=s+(\mathrm{add}_{k_1}(\tilde{a}_1)+\cdots+\mathrm{add}_{k_n}(\tilde{a}_n))-(\mathrm{del}_{k_1}(\tilde{a}_1)+\cdots+\mathrm{del}_{k_n}(\tilde{a}_n)), \forall k_1,\cdots,k_n, 1\leqslant k_1\leqslant m_1,\cdots,1\leqslant k_n\leqslant m_n\}$。

定义 2 中的状态转换函数表明,若不确定性和并行性并存,则任一组不确定效果组合都将产生一个可能的后继状态。整个后继状态集合的势等于不确定效果数的笛卡尔积。在 PPP 中允许并发动作,每个动作步可应用一个动作集合,因此后继状态是应用所有非互斥动作的一个合并状态。从以上定义可见,PPP 的状态空间非常庞大,通过穷举状态或者可达路

径来进行估值是极其困难的;再加上概率性的计算,状态空间呈现一定的概率分布。这些因素大大增加了 PPP 求解方法的复杂性。因此,如何进行概率演算以及如何设计高效的启发式函数,成为 PPP 求解方法中的关键问题。

定义 3　并行概率规划解。并行概率规划问题 $\Pi_{\text{PPP}} = \langle \Sigma_{\text{PPP}}, I, G \rangle$ 的规划解是一个策略 π_{PPP},该策略是由状态—动作集序对构成的集合:$\pi_{\text{PPP}} = \{\langle s, \widetilde{A}_1 \rangle \mid s \in S,\ \widetilde{A}_1 \in 2^A$,且 $\widetilde{A}_1$ 在 s 中是可应用的$\}$。

应用规划解策略可使规划问题从初始状态转换为目标状态。

2　并行概率规划比赛

对问题实例来说,RDDL 是可分解的马尔可夫过程(Markov decision process, MDP)。如果该问题是部分可观测的,那么就是一个部分可观测的马尔可夫过程(partial observable Mar-kov decision process, POMDP)。尽管 IPPC 分为 MDP 领域和 POMDP 领域,但实际上解决 POMDP 的方法更简单。这是因为部分观测带来的抽象意味着状态空间的缩小,所以本文主要讨论 MDP 领域。下面分别介绍 IPPC-2011 和 IPPC-2014。

IPPC-2011 有八个领域,比赛平台为 Amazon Elastic Compute Cloud (EC2)。比赛允许在手动和规划器交互,如终止运行、调试错误、改变参数、重新启动等。比赛对于所有实例的每次运行记录一个分数,再对每个实例计算归一化分数,80 个实例的归一化分数的平均值为最终得分。这届比赛参加布尔型 MDP 比赛的规划器有五个,其排名为:SPUDD[20]、Glutton[22,23]、PROST[21]、MIT-ACL 和 Beaver[24]。

IPPC-2014 竞赛领域和实例设置与 IPPC-2011 基本相同,区别主要是在比赛规则上。IPPC-2014 上,每个实例的运行有时间限制,且比赛开始后不允许手动调整参数。这一届参加布尔型 MDP 比赛的有 PROST、G-Pack、PPUDD 以及 LRTDP。其中 G-Pack 是 Glutton 的后继版本,PPUDD 是 SPUDD 的概率版本,LRTDP 类似非迭代的 Glutton。最后 PROST 和 G-Pack 获胜。

3　并行概率规划描述语言与基本领域

3.1　并行概率规划的描述语言

PPP 领域用一种新的基于规则的语言来建模,即 RDDL[13]。在 RDDL 中,动作效果的全局性约束使得基于效果的动作模型不再适用,即 PDDL 家族[11],包括 PPDDL[12] 都不可直接使用。全局约束要用规则来定义,于是新规划语言就基于规则。从语义上粗略地讲,RDDL 就是一个动态贝叶斯网络(dynamic Bayesian network,DBN),它用一个简单的影响图扩展节点代表立即回报。目标函数用来指定这些立即奖励将如何在最短时间内被优化。它有如下特征[13]:

a)一切都是具有参数的变元,包括状态、动作和观测。

b)变元可以是一个流(fluent)文字也可以不是。流是一种关系,其真值随着状态变化;而非流文字则在任意状态中保持真值不变。

c)转换是状态和动作的函数。其输入是当前状态和一些可应用的动作,而输出是下一个状态。

d)报偿函数是一个伯努利分布或者一个逻辑表达式。

e)转换和报偿函数都可含有表达式。

f)约束可以加在状态或者动作上。一个约束是含有流文字或者非流文字的逻辑或者关系表达式,它可以含有算术的或者关系的操作,以及蕴涵、量词甚至函数。

尽管 RDDL 是一种新的语言,但是它根植于 PDDL 家族中。比如,RDDL 的转换关系使用规则来表示,而这样的规则在 PDDL 中被称为派生谓词规则或公理;再比如,在 RDDL 中非常灵活且作用巨大的约束在 PDDL 中也出现过。

RDDL 为 PPP 领域描述提供了语法规范,但是目前这种新语言的语义解释并不完备。如 DBN 是在命题空间的,而 RDDL 的推理基于一阶规则;一般语言的语义应在一阶层面上定义。

3.2　并行概率规划的基本领域

在国际规划大赛 IPPC-2011(http://users.cecs.anu.edu.au/~ssanner/IPPC_2011/)和 IPPC-2014(https://cs.uwaterloo.ca/~mgrzes/IPPC_2014/index.html)上主要有八个规划问题领域:

a)Crossing_traffic 领域描述的是路径规划问题。在一个房间里面,机器人需要自动规划路线,避开障碍物从而到达目的地。

b)Elevator 领域描述的是电梯运作问题。有多个电梯在同时运作,乘客到达具有某种概率分布。电梯调度正确则给予一定的奖赏,反之则获得惩罚。

c)Game_of_life 领域描述的是细胞存活实验。动作可以使得细胞存活或者灭亡,但是细胞存活的概率会随着周围活细胞数目的增多而增大。求解目标是使得尽量多的细胞生存下来。

d)Navigation 领域描述的也是机器人路径规划的问题。只不过,与 crossing_traffic 领域不同的是,每个位置都有时间窗口让机器人不能停留。因此除了路径规划,还有时间上的调度。

e)Recon 领域描述的是太空任务。具体任务包括水检测、生命检测和拍照。Agent 对任务的执行必须满足一定的因果关系,例如必须确定有生命痕迹才能进行拍照。

f)Skill_teaching 领域描述的是技能学习问题。通过对学生传授一系列技能,使得学生最终考试成绩提高的概率增大。

g)Sysadmin 领域描述的是系统配置问题。管理员需制定最优配置从而使得电脑系统的整体性能提高,如重启次数最少。

h)Traffic 领域描述的是交通控制问题。动作可以控制十字路口的交通灯信号,从而尽量避免车流拥塞。

4　并行概率规划器

PPP 问题的求解可基于模拟或迭代深化等技术。IPPC-2011 有五个规划器参加 MDP 比赛,而 IPPC-2014 上参赛的都是它们的后继版本。因此,本章先介绍 IPPC-2011 上的五个规划器及其后继,再对 IPPC-2014 上获得冠亚军的 PROST 和 Glutton 验证其开放源码在无干预时的实际性能,最后分析结果。

4.1　并行概率规划器介绍

IPPC-2011 上规划器共有五个,它们是 SPUDD、Glutton、PROST、MIT-ACL 和 Beaver。分别介绍如下:

a)SPUDD[20] 用动态规划技术但不枚举完整状态。它用动态抽象方法来求解 MDP,该方法用代数决策图 ADDs[19] 来表示值函数和策略。SPUDD 假设已经存在一个 ADDs 输入,然后通过操作 ADDs 来实现。ADDs 往往是非常紧凑的,因此期望值的计算量大大降低。SPUDD 用值迭代的方法求贝尔曼方程直到超时。在 IPPC-2014 上出现的是 SPUDD 的概率版本:

PPUDD 和 ATPPUDD。它们把 MDP 近似地看做为一个概率模型,这样可以降低计算复杂度,而不会降低太多的解质量。但是观测会导致概率节点的数量增多,所以会出现状态数目爆炸现象。PPUDD 把过渡和奖励决策图转换成概率 ADDs,然后解贝尔曼方程的概率版本。与 SPUDD 不同,PPUDD 采用逐步增强的规划,它保持从初始状态得到的所有可能状态,限制当前的可达状态。PPUDD 采用离线算法,ATPPUDD 是其在线版本。后者在规划时对计算工作进行调度,但是在复杂的奖励制度下,目前的自动解释可能导致一个很差的策略。此外,输入的 SPUDD 域的 ADDs 实例在优化之前容易超出内存的大小。

b) Glutton[22] 使用的主要算法是 LR^{2}TDP,它基于 LRTDP[26] 的反向迭代加深算法,用一系列优化算法来提高在复杂问题上用大分支因子来解决问题的能力。其中最重要的是二次抽样转换函数,它可分离出高速缓存转移函数的样本。在 IPPC-2011 中,动作实例的数量往往与问题的状态空间的大小等数量级。而 RTDP 在贝尔曼备份过程中要枚举动作,所以往往失败。Glutton 用状态动作采样来解决这个问题。然而采样过程仍然过于耗时,所以 Glutton 缓存采样的动作成果。为了提供一个良好的实时规划,Glutton 采用迭代加深的方式来运行 RTDP。对 H 步有限域问题,它先解决一个 1 步版本,然后产生一个 2 步版本,如此循环。一旦时间耗尽,Glutton 为每个动作提供策略和候选策略。此外,Glutton 采取的优化策略是动态地分配更多的时间给难的问题,并先尝试最简单的实例。在 IPPC-2014 上,Glutton 的后继版本 G-pack 参加了比赛,但算法没有本质改变。

c) PROST[21] 基于 UCT[27] 算法,该算法是一个蒙特卡罗树抽样过程。在 UCT 中,在每个样品中的动作选择(被称为 rollout)依赖于先前所有的 rollouts。尽管可证明,UCT 在极限情况下是最优的,但是它需要相当长的一段时间才能达到收敛,因为初始的 rollouts 是随机的。为了改善这种情况,PROST 改进了 UCT,它用一个简单的结果确定化方法来计算两步的前向启发式。PROST 保存所有计算过的状态—动作对,它还包括一个能够识别多余行动从而大大减少分支因子的方法。它在搜索树的未访问节点时用一个一般的初始值来代替 UCT 的初始随机步,从而更快收敛。在 IPPC-2014 上,PROST 提供了一个新的版本:PROST 2014。PROST 2014 在有限域 MDPs 内,基于启发式树搜索,并使用不合理动作修剪、报偿锁检测和基于迭代加深搜索的启发式等方法。若在未超时之前能求解根节点,那么 PROST 2014 将内部模拟几次运行并确定最佳的 30 个连续模拟。它用总运行次数来确定一个好的停止时机。

d) MIT-ACL 采用基于价值的增强学习方法来解决 MDPs。其值函数用线性函数逼近,其中的特征是存在于顶层的二进制状态维度的逻辑功能。更具体而言,状态维度构成了最初的特征。最后,还使用了贪心策略作为缺省方法。

e) Beaver[24] 使用双向的在线概率规划方法,结合了决策理论规划(decision theory planning, DTP)和前向搜索。DTP 是一个求解概率规划的方法,它把问题看做 MDP 并产生一个策略。DTP 经典解决方案要枚举整个状态空间,通常不可行。在 DTP 中,可通过分解为 MDP 提供解法。这些解法是抽象的,仅仅需要枚举相关条件。这些条件将状态划分为等价类,而不列举整个状态空间。这类基于分解的 MDP 求解器在解决规划问题方面一直非常成功,并在过去的十年内产生了许多变种算法。

4.2 实验对比与分析

从两届 IPPCs 的结果来看,最有代表性的概率规划器就是 PROST 和 Glutton,分别代表着基于模拟和基于迭代深化的方法。考虑到 IPPC-2011 允许手动调节参数和人工干涉,而 IPPC-2014 规划器无法获得源码,本文在自有实验平台上验证它们的实际性能,并加以分析。

本文的测试环境如下。操作系统:Linux Ubuntu 12.04;CPU:2.27 GHz(i3 350);内存:2 GB。测试对全部基准领域各 10 个实例每个运行 3 次,取平均的运行时间和奖赏值。实验结果表明两者的求解质量类似且均非最优,Glutton 在求解能力方面不如 PROST,但是在求解效率方面却优于 PROST。两者都能全部求解的基准领域是 nevigation 领域(如图 1 所示)和 skill_teaching 领域。在有些领域,比如 recon 和 traffic 领域,Glutton 对大部分问题都无法求解,而在另一些领域,如 game_of_life 领域,PROST 对大部分实例都运行超过 30 min(被终止)。

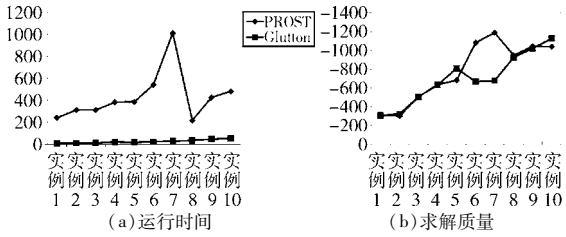

图 1 PROST 和 Glutton 在 navigation 领域的运行时间和求解质量实验对比(时间单位为 s,规划解质量为奖赏值)

在 IPPC-2011 中,所有规划器都对大部分问题能求解,这说明有人工调节参数甚至干预来促成其求解。在规划器的算法方面,上述实验结果表明基于模拟的方法性能稍差,而基于迭代方法时间估计困难导致求解规模较小。此外,大部分现有 PPP 规划器都要求转换输入文件的格式,即不完全支持 RDDL。但 RDDL 基于规则,PPDDL 基于动作效果,二者的表示能力不等价,如对中间变量[13] 上述规划器及其输入文件都不能支持。

5 结束语

PPP 是最新的复杂规划领域,AAAI2015 上有近 10 篇文章讨论与其相关的问题[28~30]。PPP 的特殊性在于要求描述语言基于规则,而非效果模型,其优势在于非常适合描述允许并发性和不确定性同时存在的规划问题。现有 PPP 规划器的求解技术多基于 UCT 模拟或迭代深化搜索,它们各有所长。本文的综述为对 PPP 感兴趣的学者,特别是国内的学者,提供了重要的参考资料。尽管 PPP 的研究推动了复杂规划领域研究的热潮,然而要做到完美地求解还有很多方面的工作要进一步开展。例如在理论方面,需要完备 RDDL 的语义解释。目前 RDDL 用于描述 PPP 领域和问题,但其语义模型仅粗略地归结于 DBN,其中的转换网络可通过领域描述中的转换规则来构建。但是 RDDL 描述是一阶的,其语义模型也应上升到一阶网络结构,即 RDBN(relational dynamic bayesian network)[31]。RDBN 是目前最复杂的 DBN,是一阶逻辑和动态层次网络结合的产物。因此,用 RDBN 来解释 RDDL 的语义是可行的,但需要严格和完备的论证。再如在算法方面,两类基本求解算法的优缺点明显,要互相借鉴、取长补短。具体地说,基于模拟的方

法应采用更高效的启发式以加快收敛速度,从而减少求解时间;而基于迭代搜索的方法应改善动作采样技术以应对大规模的问题实例,从而增加求解能力。在应用方面,一方面应尽量拓展 PPP 的基准领域,另一方面应针对电梯调度、交通灯控制等最有代表性的并发不确定领域,实现针对特定领域的规划求解器,以真正解决实际问题。此外,规划和学习的结合一直是两个学科融合的热点。因此,如何使用机器学习方法以自动获取 RDDL 领域描述模型[32]也是一个值得期盼的研究方向。

参考文献:

[1] Geffner H, Bonet B. A concise introduction to models and methods for automated planning [M]//Synthesis Lectures on Artificial Intelligence and Machine Learning. [S. l.]: Morgan & Claypool Publishers, 2013: 1-141.

[2] Ghallab M, Nau D, Traverso P. Automated planning: theory and practice [M]. San Francisco: Morgan Kauffmann Publishers, 2004: 1-635.

[3] Ghallab M, Nau D, Traverso P. 自动规划——理论和实践[M]. 姜云飞,杨强,凌应标,译. 北京:清华大学出版社, 2008.

[4] Piacentini C, Alimisis V, Fox M, et al. An extension of metric temporal planning with application to AC voltage control [J]. Artificial Intelligence, 2015, 229(12):210-245.

[5] Núñez S, Borrajo D, López C L. Automatic construction of optimal static sequential portfolios for AI planning and beyond[J]. Artificial Intelligence, 2015, 226(9):75-101.

[6] Brafman R, Domshlak C. On the complexity of planning for agent teams and its implications for single agent [J]. Artificial Intelligence, 2013, 198(5): 52-71.

[7] Hanheide M, Göbelbecker M, Horn G, *et al*. Robot task planning and explanation in open and uncertain worlds[J]. Artificial Intelligence, 2015(In Press).

[8] 饶东宁,蒋志华,姜云飞. 从 WSBPEL 程序中学习 Web 服务的不确定动作模型[J]. 计算机研究与发展, 2010, 47(3): 445-454.

[9] Milani A, Terragnolo M. Representing conflicts in parallel planning [C]//Proc of the 1st International Conference on Artificial Intelligence Planning Systems. San Francisco: Morgan Kauffmann Publishers, 1992: 291-292.

[10] Cimatti A, Roveri M, Traverso P. Strong planning in non-deterministic domain via model checking[C]//Proc of the 4th International Conference on AI Planning System. Palo Alto: AAAI Press, 1998: 36-43.

[11] Gerevini A, Long D. Plan constraints and preferences in PDDL3: the language of the Fifth International Planning Competition, RT 2005-08-47[R]. Brescia, Italy: University of Brescia, 2005.

[12] Younes H L S, Littman M L. PPDDL1.0: an extension to PDDL for expressing planning domains with probabilistic effects, CMU-CS-04-167[R]. Pittsburgh: Carnegie Mellon University, 2004.

[13] Sanner S. Relational dynamic influence diagram language (RDDL): language description [EB/OL]. (2011-01). http://users.cecs.anu.edu.au/~ssanner/IPPC2011/RDDL.pdf.

[14] Kautz H, Selman B. Blackbox: a new approach to the application of theorem proving to problem solving [C]// Proc of the 4th International Conference on AI Planning System Workshop on Planning as Combinatorial Search. Palo Alto: AAAI Press, 1998:58-60.

[15] Blum A, Furst M. Fast planning through planning graph analysis [C]//Proc of the 14th International Joint Conference of Artificial Intelligence. San Francisco: Morgan Kaufmann Publishers, 1995: 1636-1642.

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

[17] Richter S, Westphal M. The LAMA planner: guiding cost-based anytime planning with landmarks [J]. Journal of Artificial Intelligence Research, 2010,39:127-177.

[18] Torralba A, Alcazar V, Borrajo D. SymBA*: a symbolic bidirectional A* planner[C]//Proc of the 24th International Conference on Automated Planning and Scheduling. 2014:105-109.

[19] Buffet O, Aberdeen D. The factored policy gradient planner (IPC-06 version)[C]//Proc of the 16th International Conference on Automated Planning and Scheduling. 2006.

[20] Hoey J, Staubin R, Hu A, et al. SPUDD: stochastic planning using decision diagrams[C]//Proc of the 15th Conference on Uncertainty in Artificial Intelligence. [S. l.]: North-Holland Publishing Company, 1999: 279-288.

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

[22] Kolobov A, Dai Peng, Mausam D S, *et al*. Reverse iterative deepening for finite-horizon MDPs with large branching factors[C]//Proc of the 22nd International Conference on Automated Planning and Scheduling. Palo Alto: AAAI Press, 2012:146-154.

[23] Kolobov A, Mausam D S, Weld D. LRTDP vs. UCT for online probabilistic planning[C]//Proc of the 26th Conference on Artificial Intelligence. Palo Alto: AAAI Press, 2012:1786-1792.

[24] Raghavan A N, Joshi S, Fern A, et al. Bidirectional online probabilistic planning[C]//Proc of the 22nd International Conference on Automated Planning and Scheduling. Palo Alto:AAAI Press,2012.

[25] Bahar R I, Frohm E A, Gaona C M, *et al*. Algebraic decision diagrams and their applications[C]//Proc of IEEE/ACM International Conference on Computer-Aided Design. Piscataway: IEEE Press, 1993: 188-191.

[26] Bonet B, Geffer H. Labeled RTDP: improving the convergence of real-time dynamic programming[C]//Proc of the 13th International Conference on Automated Planning and Scheduling. Palo Alto: AAAI Press, 2003:12-21.

[27] Kocsis L, Szepesvari C. bandit based Monte-Carlo planning[C]// Proc of the 17th European Conference on Machine Learning. Berlin: Springer, 2006:282-293.

[28] Ghooshchi N G, Namazi M, Newton M A H, *et al*. Transition constraints for parallel planning[C]//Proc of the 29th AAAI Conference on Artificial Intelligence. Palo Alto: AAAI Press, 2015:3268-3274.

[29] Vianna L, Barros L, Sanner S. Real-time symbolic dynamic programming[C]//Proc of the 29th AAAI Conference on Artificial Intelligence. Palo Alto: AAAI Press, 2015:3402-3408.

[30] Srinivasan S, Talvitie E, Bowling M. Improving exploration in UCT using local manifolds[C]//Proc of the 29th AAAI Conference on Artificial Intelligence. Palo Alto: AAAI Press, 2015: 3386-3392.

[31] Manfredotti C, Messina E. Relational dynamic Bayesian networks to improve multi-target tracking [C]//Proc of the 11th International Conference on Advanced Concepts for Intelligent Vision Systems. 2009: 528-539.

[32] Rao Dongning, Jiang Zhihua. Learning planning domain descriptions in RDDL [J]. International Journal on Artificial Intelligence Tools, 2015, 24(3): 1550002.