软件学报 ISSN 1000-9825, CODEN RUXUEW
Journal of Software,2016,27(11):2814−2827 [doi: 10.13328/j.cnki.jos.004928]

E-mail: jos@iscas.ac.cn
http://www.jos.org.cn
Tel: +86-10-62562563

通用对弈游戏:一个探索机器游戏智能的领域*

张海峰, 刘当一, 李文新

(北京大学 信息科学技术学院,北京 100871)

通讯作者: 张海峰, E-mail: pkuzhf@pku.edu.cn

摘 要: 通用对弈游戏(general game playing,简称 GGP)是致力于提高机器的通用游戏智能的研究领域.与专用游戏智能程序不同,GGP 玩家直到游戏开始时才获得游戏规则,从而避免依赖于人类关于特定游戏的经验.GGP 研究发展至今,已在游戏表示、搜索算法、状态估值等方面做了深入探索,并在知识迁移等方面做出了尝试.GGP 研究的进展在一定程度上代表了通用人工智能的发展,因而是值得关注的.

关键词: 通用;游戏;逻辑;人工智能;知识表示

中图法分类号: TP18

中文引用格式: 张海峰,刘当一,李文新.通用对弈游戏:一个探索机器游戏智能的领域.软件学报,2016,27(11):2814−2827. http://www.jos.org.cn/1000-9825/4928.htm

英文引用格式: Zhang HF, Liu DY, Li WX. General game playing: A research field for exploring machine intelligence in games. Ruan Jian Xue Bao/Journal of Software, 2016,27(11):2814−2827 (in Chinese). http://www.jos.org.cn/1000-9825/4928.htm

General Game Playing: A Research Field for Exploring Machine Intelligence in Games

ZHANG Hai-Feng, LIU Dang-Yi, LI Wen-Xin

(School of Electronics Engineering and Computer Science, Peking University, Beijing 100871, China)

Abstract: General game playing (GGP) is a research field for improving the general gaming intelligence of machines. It is different from specific game playing in that GGP players do not know game rules before the game begins, which makes them independent from human experience on specific games. Until now, GGP researchers have deeply explored many problems, such as game representation, search algorithm, and state evaluation. Also, some efforts have been made on knowledge transfer. The progress of GGP research represents the development of artificial general intelligence to some extent, which makes it remarkable.

Key words: general; game; logic; artificial intelligence; knowledge representation

通用人工智能是人工智能领域的一项重大挑战,其目标是让机器能够像人一样完成各种各样的任务.具体到游戏领域,通用对弈游戏(general game playing,简称 GGP)致力于开发一种能够以人类水准玩任意已知或未知游戏的人工智能系统.

众所周知,IBM 的超级计算机“深蓝”战胜国际象棋冠军卡斯帕罗夫,标志着人工智能到达了一个新的高度.然而,“深蓝”的智能至少在两个方面还存在局限性:其一是只能玩国际象棋这一种游戏,不具有通用性;其二是依赖于大量的人类游戏经验,不具有完全自主学习的能力.GGP 研究的目标就是突破这些局限,它设置的环境要求机器必须在没有人类游戏经验的指导下玩各种各样的游戏.因此,GGP 研究的进展反映了机器游戏智能在通用性和自主学习方面的发展,GGP 比赛则成为一种评价机器游戏智能的标准.

* 收稿时间: 2015-06-23; 修改时间: 2015-07-17; 采用时间: 2015-10-09; jos 在线出版时间: 2015-11-15
CNKI 网络优先出版: 2015-11-16 09:22:19, http://www.cnki.net/kcms/detail/11.2560.TP.20151116.0922.002.html

1 GGP 研究的基本框架

Genesereth 等人于 2005 年提出了 GGP[1],同时开始举办一年一度的国际 GGP 比赛.GGP 研究发展至今,已经形成了一个特定的游戏集合,并在不断地发展扩大.游戏集合中,所有游戏的规则都用一种统一的游戏描述语言描述,GGP 玩家在每场游戏对局的开始阶段获得游戏规则.在一场游戏对局中,参与游戏的 GGP 玩家与一台扮演裁判的游戏主机进行交互,共同完成游戏.组织游戏对局的通常是 GGP 平台或比赛,它们成为衡量 GGP 研究进展的标尺.

1.1 游戏集合

目前,公开的 GGP 游戏约有几百个,它们在玩家人数、复杂度、对称性等方面各不相同,具有很好的多样性.有些游戏之间还具有关联性,这些关联性为游戏知识的迁移提供了发挥作用的空间.

1.1.1 典型游戏

GGP 游戏中最常见的是棋盘游戏,此外还有地图游戏、博弈论游戏等游戏形式.图 1 展示了几个比较典型的 GGP 游戏.

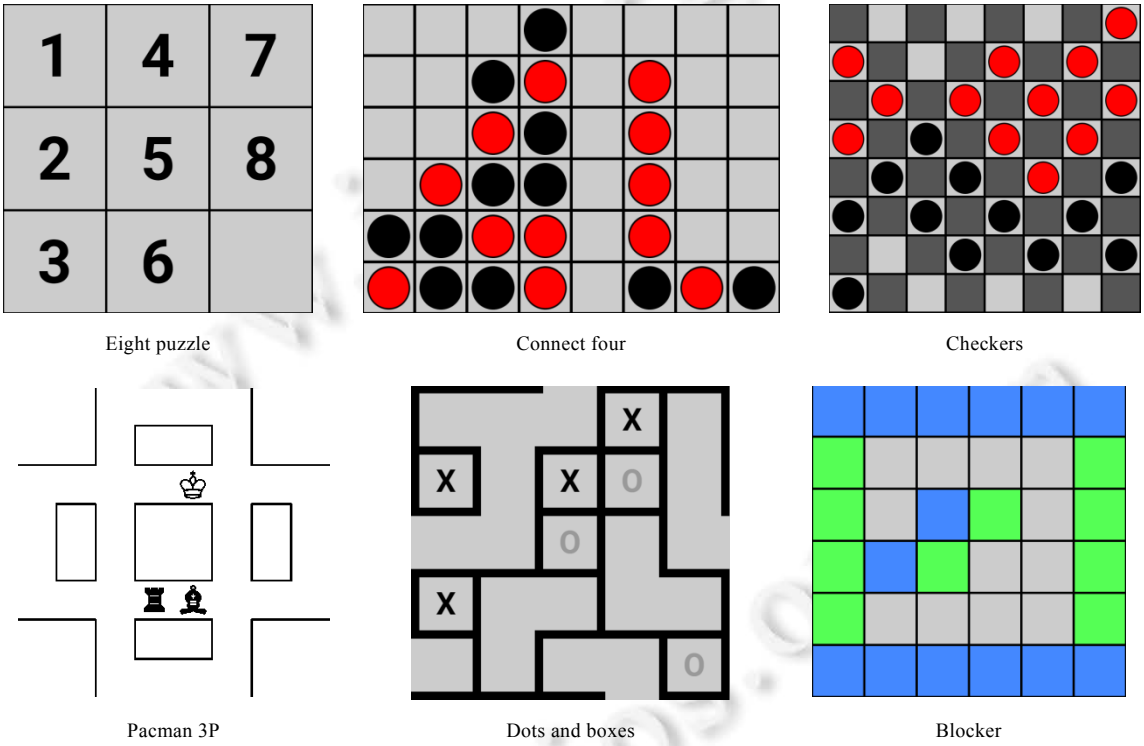

Fig.1 Examples of games

图 1 游戏举例

Eight Puzzle 是一个单人游戏,玩家每个回合可以移动一个数码,目标是使全部数码按照 1~8 的顺序排列.Connect Four 是一个双人棋盘游戏,双方按一定规则轮流落子,目标是率先形成“四子连珠”.Checkers(西洋跳棋)是一个经典的双人棋盘游戏,它的棋子只能减少不能增加,这与 Connect Four 正好相反.Pacman 3P 是一个不对称的三人地图游戏,其中,一个玩家的目标是在地图上吃尽量多的豆子,另两个玩家的目标是围捕第 1 个玩家.Dots and Boxes 是一个双人游戏,玩家轮流在地图上筑墙,筑四面墙能够圈一块地,圈地多者获胜.Blocker 是一个不对称的双人游戏,蓝、绿双方每个回合同时占领一个格子,绿色玩家的目标是连接左右,蓝色玩家的目标是阻

止绿色玩家.

可见,GGP 游戏丰富多彩,既有单人的也有多人的;既有同步的也有异步的,既有竞争的也有配合的,既有对称的也有非对称的.

1.1.2 生成的游戏

为了鼓励GGP玩家实现游戏知识的迁移,GGP游戏集合中存在着一些从已有游戏生成的游戏,包括组合游戏和相似游戏.

组合游戏是由几个相同或不同的游戏组合而成的游戏.比如,可以将两个 Connect Four 棋盘组合成一个新游戏,玩家每个回合只能选择其中一个棋盘行动,率先在任何一个棋盘获胜的玩家获胜.游戏还有其他组合方式,比如两个棋盘都获胜的玩家才获胜,或者指定其中一个棋盘获胜的玩家才获胜.

相似游戏是由另一个游戏修改某些条件变化而来的游戏.其中最为典型的是自杀游戏,即,游戏结果与原游戏完全相反的游戏.比如,Connect Four Suicide 规定率先形成“四子连珠”的玩家失败,这与 Connect Four 正好相反.此外,修改游戏玩家数、修改棋盘大小也是常见的构造相似游戏的方法.

1.2 游戏描述方式

游戏的规则可以用不同的方式加以描述,人类常用的是自然语言,数学定义上则常用状态机模型.从抽象程度上来说,这是两种极端的表述方式.高度抽象的自然语言虽然简洁,但对于计算机程序来说存在太多歧义,如何正确地解读还是一个难题.而高度具体的状态机模型需要描述太多细节,对于很多游戏来说仅存在理论上的可用性,比如国际象棋具有约 10^{50} 个状态[2],远远超过当今计算机存储空间的数量级.因此,要让计算机理解游戏规则,需要采用抽象程度适中的、无歧义的、相对简洁的游戏描述方式.

Love 等人提出的游戏描述语言(game description language,简称 GDL)[3]是专门用来描述游戏规则的一种语言,已经被 GGP 比赛广泛使用.GDL 是一种声明式的逻辑编程语言,它的抽象程度介于自然语言和状态机模型之间,是无歧义且相对简洁的,适于计算机理解.GDL 可以被认为是状态机模型的一种改写,它将状态机的状态用命题集合的方式编码,再将状态间的转移关系改写成命题间的逻辑关系,这种改写极大地简化了游戏的描述.

1.2.1 游戏的状态机模型

状态机模型是用集合论描述游戏规则的方式.状态机的每一个状态对应游戏中的一个局面,每一个状态下,每个玩家具有若干个合法的动作,玩家各自选择一个合法动作,共同决定游戏的下一个状态.游戏具有一个初始状态和若干个结束状态,结束状态决定各个玩家的得分.其形式化定义[4]如下:

定义 1(游戏的状态机模型). 设Σ是一个可数的字符集,$S\subseteq 2^{\Sigma}$是一个状态集,$A\subseteq\Sigma$是一个动作集,则一个离散的、同步的、确定性的、完全信息的游戏可以表达为一个六元组(R,s_0,T,L,u,G),满足:

- $R\subseteq\Sigma$,表示玩家;
- $s_0\in S$,表示游戏初始状态;
- $T\subseteq S$,表示游戏结束状态;
- $L\subseteq R\times A\times S$,表示各玩家在各状态下的合法动作;
- $u:(R\rightarrow A)\times S$,表示在各玩家动作共同作用下状态的转移;
- $G\subseteq R\times\mathbb{N}\times S$,表示各玩家在各状态下的得分.

1.2.2 GDL 的语法

GDL 用命题的集合编码游戏状态.对于一个游戏,GDL 定义一个状态命题集合 P,并在游戏状态集合 S 与 P 的幂集 2^P 之间建立一个单射 $f:S\rightarrow 2^P$,使得任意一个游戏状态对应 P 的一个子集.对于任意命题 p 和任意状态 s,若 $p\in f(s)$,则称 p 在 s 下为真,否则为假.GDL 还定义玩家的动作命题集合 A,使得玩家在任何游戏状态下的合法动作命题集合 $L_{s,r}$ 都是 A 的一个子集.此外,GDL 还规定了一些描述游戏元素的命题,它们的含义见表 1.

在此基础上,GDL 定义这些命题之间一系列的逻辑关系,包括:

- 当前状态为否是结束状态的逻辑,即由 true 命题推出 terminal 命题的逻辑;
- 当前状态下各玩家的合法动作的逻辑,即由 true 命题推出 legal 命题的逻辑;

- 当前状态如何转移到下一状态的逻辑,即由 true 命题和 does 推出 next 命题的逻辑;
- 当前状态下各玩家得分的逻辑,即由 true 命题推出 goal 命题的逻辑.

有了这些定义,GDL 的描述能力就与状态机模型等价了.此外,实际的 GDL 规则中通常还含有辅助性命题,辅助性命题的引入本质上不改变上述特殊命题之间的逻辑关系,只是为了描述的方便性与可读性.

Table 1 Propositions describing elements of games in GDL

表 1 GDL 中描述游戏元素的命题

命题	命题为真时的含义
(role *r*)	*r* 是一个玩家
(init *p*)	状态命题 *p* 在初始状态下为真
(true *p*)	状态命题 *p* 在当前状态下为真
(legal *r a*)	玩家 *r* 在当前状态下具有合法动作 *a*
(does *r a*)	玩家 *r* 采取了动作 *a*
(next *p*)	状态命题 *p* 在下一状态下为真
terminal	当前状态是结束状态
(goal *r n*)	当前状态下玩家 *r* 的得分是 *n*

1.2.3 GDL 的例子——井字棋

井字棋(Tic-Tac-Toe)游戏是一个简单易玩的游戏,它规定游戏双方在 3×3 的棋盘上轮流落子,率先将己方落子连成三子一线(横、竖或对角线)的玩家获胜.井字棋的 GDL 描述如下所示.

```
1.  (role xplayer) (role oplayer)
2.  (init (cell 1 1 b)) (init (cell 1 2 b))...(init (cell 3 3 b))
3.  (init (control xplayer))
4.  (⇐(legal ?w (mark ?x ?y)) (true (cell ?x ?y b)) (true (control ?w)))
5.  (⇐(legal xplayer noop) (true (control oplayer)))
6.  (⇐(legal oplayer noop) (true (control xplayer)))
7.  (⇐(next (cell ?m ?n x)) (does xplayer (mark ?m ?n)) (true (cell ?m ?n b)))
8.  (⇐(next (cell ?m ?n o)) (does oplayer (mark ?m ?n)) (true (cell ?m ?n b)))
9.  (⇐(next (cell ?m ?n ?w)) (true (cell ?m ?n ?w)) (distinct ?w b))
10. (⇐(next (cell ?m ?n b)) (does ?w (mark ?j ?k)) (true (cell ?m ?n b)) (or (distinct ?m ?j) (distinct ?n ?k)))
11. (⇐(next (control xplayer)) (true (control oplayer)))
12. (⇐(next (control oplayer)) (true (control xplayer)))
13. (⇐(row ?m ?x) (true (cell ?m 1 ?x)) (true (cell ?m 2 ?x)) (true (cell ?m 3 ?x)))
14. (⇐(column ?n ?x) (true (cell 1 ?n ?x)) (true (cell 2 ?n ?x)) (true (cell 3 ?n ?x)))
15. (⇐(diagonal ?x) (true (cell 1 1 ?x)) (true (cell 2 2 ?x)) (true (cell 3 3 ?x)))
16. (⇐(diagonal ?x) (true (cell 1 3 ?x)) (true (cell 2 2 ?x)) (true (cell 3 1 ?x)))
17. (⇐(line ?x) (or (row ?m ?x) (column ?m ?x) (diagonal ?x))
18. (⇐open (true (cell ?m ?n b)))
19. (⇐(goal xplayer 100) (line x))
20. (⇐(goal xplayer 50) (not (line x)) (not (line o)) (not open))
21. (⇐(goal xplayer 0) (line o))
22. (⇐(goal oplayer 100) (line o))
23. (⇐(goal oplayer 50) (not (line x)) (not (line o)) (not open))
24. (⇐(goal oplayer 0) (line x))
25. (⇐terminal (or (line x) (line o) (not open)))
```

GDL 采用的是前缀逻辑,其中,⇐是推理符号,其后的第 1 项表示推理结果,其余若干项表示推理条件.以问

号开头的字符串表示变量.第 1 行描述了游戏有两个玩家.第 2 行、第 1 行描述了游戏的初始状态,即棋盘 9 个格子都为空,初始控制权属于 xplayer.第 4 行~第 6 行描述了当前状态下各玩家的合法动作.第 7 行~第 12 行描述了当前状态如何转移到下一状态.第 13 行~第 18 行描述了一些辅助性命题,如“三子一线”等概念.第 19 行~第 24 行描述了当前状态下各玩家的得分.第 25 行描述了当前状态是否为结束状态.

1.2.4 GDL 的描述能力

和状态机模型一样,GDL 可以描述离散的、同步的、确定性的、完全信息的游戏.

GDL 只能描述离散的游戏,因为它所定义的游戏状态是离散的.对于具有连续状态的游戏,可以通过大量离散状态逼近的方式使其离散化,再用 GDL 加以描述.事实上,由于计算机的工作方式本身是离散的,任何依赖于连续概念的游戏都只能用离散的方式被计算机处理.

GDL 描述的游戏是同步的,即所有玩家在某一状态下同时采取动作,共同影响状态的转移.事实上,异步的游戏可以通过增加“无效的”动作转化成同步的游戏.“无效的”动作是指对状态转移没有任何实质影响的动作,如上述井字棋游戏,虽然是双方轮流落子的异步游戏,但通过赋予非落子方一个“无效的”noop 操作转化成了同步的游戏(见第 5 行、第 6 行).

GDL 描述的游戏都是确定性的和完全信息的.非确定性的游戏包括一些需要掷骰子的游戏,在这类游戏中,状态的转移不仅依赖于玩家的动作,还依赖于某些带有随机性的事件.非完全信息的游戏包括一些牌类游戏,在这类游戏中,玩家看不到其他玩家的牌,因此不能完全知晓当前所处的游戏状态.为了扩大 GDL 的描述能力,Thielscher 等人提出的 GDL-II(game description language with incomplete information)[5]引入了用于描述随机性的 random 命题和用于描述非完全信息的 sees 命题,从而具备了描述非确定性、非完全信息游戏的能力.虽然如此,目前 GGP 的研究主要还是建立在 GDL 上,GDL-II 只作为一种可能的扩展方式而存在,是否是唯一或最佳的扩展方式尚未有定论.

1.3 游戏对局交互方式

GGP 对局采用一种独特的交互机制,它规定玩家的程序各自在本地运行,并提供一个网络接口供游戏主机远程访问.游戏主机是对局的裁判,它负责与各个玩家通信以进行游戏,玩家之间不可以通信.

交互机制如图 2 所示.

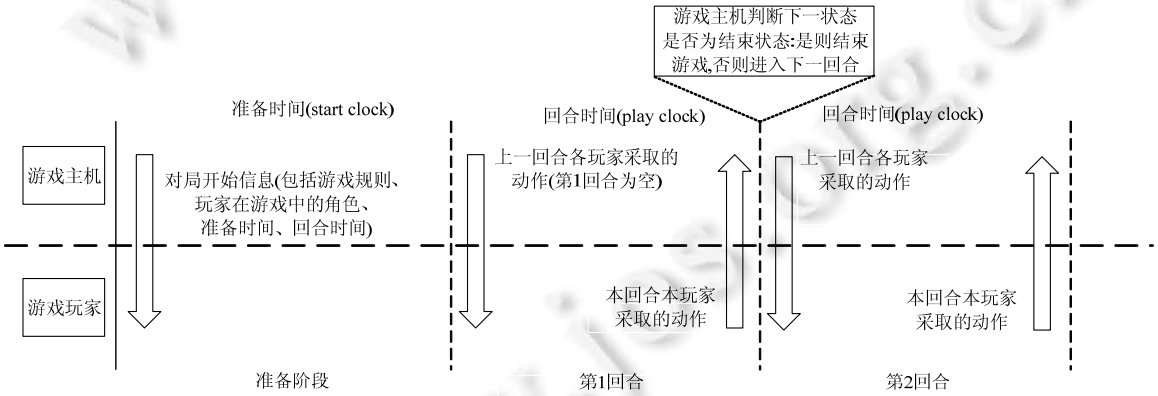

Fig.2 Game communication mechanism in GGP

图 2 GGP 对局交互机制

游戏主机首先将游戏开始的信息同时发送给各个玩家,这个信息中包含游戏规则、玩家在游戏中的角色以及比赛的准备时间和回合时间.收到游戏开始的信息后,各个玩家有一段时间做准备,这段时间称为准备时间(start clock),通常约 30s~120s.准备时间结束后,游戏正式进入第 1 回合.

每个回合各个玩家具有一段思考时间,这段时间称为回合时间(play clock),通常约 15s~60s.在每回合开始

时,游戏主机会将上一回合各个玩家采取的动作发送给所有玩家,玩家可以据此更新游戏状态,并计算当前回合的合法动作.在回合时间结束前,玩家需要选择一个合法动作发送给游戏主机.游戏主机收到各玩家的动作后,计算出下一回合的游戏状态,并判断该状态是否为游戏结束状态:若是,则结束游戏并计算得分;若不是,则开始下一回合.

1.4 游戏平台与比赛

机器玩通用游戏的能力不容易设定一个绝对的标准,因而 GGP 平台与比赛是衡量 GGP 研究进展的主要媒介.GGP 平台收集了大量的游戏,提供常年的在线评测服务,主要用于日常的 GGP 研究.GGP 比赛则定期在线上或线下举行,为 GGP 研究者提供交流的机会和场所,比赛的结果具有权威性.

1.4.1 GGP 平台

GGP 平台是常年运行 GGP 游戏主机的在线平台,一般提供数百个游戏的游戏规则,以及这些游戏的历史对局信息.游戏玩家可以通过 GGP 平台与别的游戏玩家进行游戏,游戏的结果将影响玩家在平台上的排名.GGP 平台采用上文所述的统一的游戏交互机制,为世界各地 GGP 研究者的交流提供了方便.

1.4.1.1 Tiltyard

Tiltyard 是由斯坦福大学运行的 GGP 平台,建立之初主要为斯坦福大学的 GGP 课程提供服务,现已成为 GGP 研究者重要的交流平台.GGP 玩家与 Tiltyard 建立连接后,Tiltyard 从游戏库中随机抽取游戏并组织 GGP 玩家进行对局.Tiltyard 统计对局的历史信息对玩家进行排名,并根据一定规则计算游戏的难度和平衡性.

1.4.1.2 Dresden GGP Server

Dresden GGP Server 是由德国德累斯顿工业大学运行的 GGP 平台,主要为德累斯顿工业大学的 GGP 课程提供服务.它的最大特点是加入了对非完全信息游戏描述语言(GDL-II)的支持.此外,它还具有人为选择游戏和对手的功能.

1.4.1.3 GGPZone

GGPZone 是由北京大学运行的 GGP 平台,也是目前国内唯一的 GGP 平台.它致力于为 GGP 研究提供更加丰富的游戏、研究工具、比赛模式.

1.4.2 GGP 比赛

GGP 比赛在 GGP 研究中扮演关键角色,它是评价 GGP 研究进展的重要标尺.国际 GGP 比赛已成功举办多届,吸引了来自世界各地的爱好者参加.GGP 比赛与其他游戏比赛一道探索机器游戏智能的极限.

1) 国际 GGP 比赛

国际 GGP 比赛(Int’l General Game Playing Competition)每年举办一次,主办方为斯坦福大学,2014 年之前在 IJCAI 或 AAAI 会议现场比赛,自 2014 年起改为在线比赛.国际 GGP 比赛的参赛者来自世界各地,其优胜者代表 GGP 玩家的最高水平.

由于 GGP 研究正处于不断发展的阶段,国际 GGP 比赛赛制也在逐年改进.2014 年的赛制分为资格赛和决赛.资格赛共有 6 轮,任何人可以在线报名任意多轮资格赛,只要在其中一轮中获得出线权即可参加决赛.每轮资格赛 GGP 玩家需要完成若干单人和多人游戏,主办方将根据 GGP 玩家的表现来决定是否给予决赛资格.决赛为“双败淘汰制”,每对相遇的玩家进行 3 局或 5 局对局,最终从进入决赛的 16 名玩家中选出最终的冠军.

国际 GGP 比赛历年的冠军见表 2.

国际 GGP 比赛冠军在各自夺冠的时期做出了突破性工作,Cluneplayer 和 Fluxplayer 采用的是基于评估的搜索算法[6,7],Cadiaplayer 首次采用基于模拟的搜索算法[8],Ary 使用集群分布式地进行蒙特卡洛树搜索[9],Sancho 对命题网络的推理速度等做了大量优化.

2) GGP 比赛与其他游戏比赛的比较

从比赛内容的角度看,GGP 比赛是针对通用游戏的,这是它与围棋、象棋等特定游戏比赛相比最大的不同之处.虽然 GGP 游戏集合是有限的,但在正式比赛时通常采用原创的新游戏,以免玩家预先针对特定游戏做准备.GGP 比赛不是唯一的通用游戏比赛,GVGP(general vedio game playing)[10,11]是针对通用视频游戏的比赛.此

外,Google 公司对 Atari 游戏的研究[12]也与 GVGP 的设定是类似的.

从比赛规则的角度看,GGP 比赛不限制参赛玩家使用的硬件资源,这是它与统一硬件资源的比赛的不同之处.因此,GGP 比赛不仅考验参赛者编写软件算法的能力,还鼓励参赛者探索更新的硬件或组织更大规模的硬件来解决问题.随着并行计算技术的进步,GGP 比赛的竞技水平必将进一步发展.

Table 2 Annual champions of international GGP competitions (from Wikipedia)

表 2 国际 GGP 比赛历年冠军(摘自维基百科)

年份	GGP 玩家	作者	学校
2005	Cluneplayer	Jim Clune	UCLA
2006	Fluxplayer	Stephan Schiffel, Michael Thielscher	Dresden University of Technology
2007	Cadiaplayer	Yngvi Björnsson, Hilmar Finnsson	Reykjavik University
2008	Cadiaplayer	Yngvi Björnsson, Hilmar Finnsson, Gylfi Þór Guðmundsson	Reykjavik University
2009	Ary	Jean Méhat	Paris 8 University
2010	Ary	Jean Méhat	Paris 8 University
2011	TurboTurtle	Sam Schreiber	–
2012	Cadiaplayer	Hilmar Finnsson, Yngvi Björnsson	Reykjavik University
2013	TurboTurtle	Sam Schreiber	–
2014	Sancho	Steve Draper, Andrew Rose	–

2 GGP 研究的主要问题

GGP 研究的目标是让计算机以人类水平玩所有的游戏,在现阶段是离散的、同步的、确定性的、完全信息的游戏.设计一个不出错的按 GDL 规则玩游戏的程序是容易的,但使其掌握尽可能好的游戏策略却相当困难.理论上,任何一个离散的、同步的、确定性、完全信息的游戏都是有确定解的,即所有游戏状态的价值都是可计算的,这就决定了玩家天然具有某种最优策略.但由于大部分游戏状态数过大,所有游戏状态的价值无法在有限的计算资源下被计算出来,才使得寻找尽可能好的游戏策略成为一个值得研究的问题.如果把对游戏的理解看成知识,GGP 研究的问题可以归纳为游戏知识的表示、获取和迁移.

2.1 游戏知识表示

为了设计好的游戏策略,必然要先深入理解游戏,理解游戏的第 1 步就是表示游戏.博弈树和命题网络是两种常用的表示游戏的工具,事实上,它们与上文提到的用于描述游戏规则的状态机模型和 GDL 是等价的.博弈树和命题网络从两种不同的角度看待游戏,在理解游戏的过程中起到相辅相成的作用.

2.1.1 博弈树

博弈树是一棵表示游戏的树,它的节点表示游戏状态,边表示玩家的动作.博弈树的根节点表示游戏的初始状态,叶子节点表示游戏的结束状态.对于异步的游戏,博弈树的各节点处只有一个玩家可以行动,各节点作为父节点发出的各条边对应可行动玩家的各个合法动作,边所连接的子节点对应动作作用于父节点的结果.对于同步的游戏,博弈树的每个节点处所有玩家均可以行动,因此,各节点作为父节点发出的边对应的是所有玩家的一个合法动作组合,边所连接的儿子节点是这个动作组合作用于父节点的结果.

博弈树与状态机的区别在于,同一个游戏状态在状态机中只对应一个节点,而在博弈树中可能对应多个节点.这是因为从初始状态到某个游戏状态可能存在多条路径,在状态机中,这些路径都汇聚到一个节点上;而在博弈树中,这些路径导向完全不同的节点.

借助博弈树的结构,可以自底向上地按照极大极小搜索[13]的思想计算游戏状态的价值.具体来说,叶子节点是游戏结束状态,其对于各个玩家的价值(以下称价值组合)直接从游戏规则得出.假设某个节点的所有儿子节点的价值组合都已计算得出,则这个节点的价值组合取决于该节点处玩家如何行动.若该节点处只有一个玩家可以行动,则该玩家一定选择对于自身价值最高的儿子节点.因此,待求节点的价值组合等于对于行动玩家价值最高的儿子节点的价值组合.由此,可以自底向上地计算出整棵博弈树的节点价值,同时确定各玩家的最优策略.对于同步的游戏也是类似的,不同之处在于每个回合可行动的玩家有多个,因而每个回合都是一次博弈,需要求解纳什均衡才能得到最优策略.

博弈树从最为细致的角度刻画了游戏,为完全理解游戏提供了理论上的可能性.但现实中,大型博弈树的结构受制于计算资源而无法被完全知晓,只能了解其局部的信息.局部信息包括节点的估值、树的结构等,它们是非常重要的游戏知识.

2.1.2 命题网络

命题网络(propositional net,简称 Propnet)是一种借助逻辑电路表示游戏的方式.命题网络首先用状态命题集合编码游戏状态,使任意一个游戏状态对应一个状态命题集合的子集;其次,命题网络定义了一系列命题表示游戏元素,如合法动作、游戏结束、游戏得分、状态转移等;最后,命题网络在这些命题之间建立逻辑连接,使得它能够完成表 3 中的推理.

Table 3 Reasoning achieved by propositional net

表 3 命题网络能够完成的推理

输入	输出
当前状态命题集合,各玩家的动作组合	下一状态命题集合
当前状态命题集合	各玩家合法动作集合
当前状态命题集合	结束命题
当前状态命题集合	得分命题

命题网络的定义与上文所述 GDL 的定义是一脉相承的.事实上,命题网络就是 GDL 的展开形式.GDL 通过一系列等价变换(消除变量、辅助命题,只保留描述基本游戏元素的命题及它们之间的逻辑关系),就得到了命题网络.通常可以用一个逻辑电路图来表示命题网络,如图 3 所示.

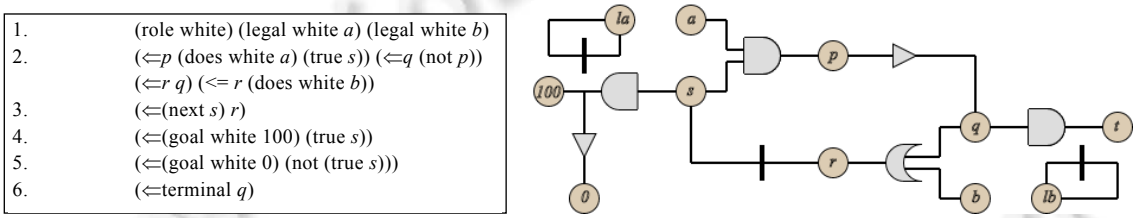

Fig.3 Illustration of propositional net (from Stanford CS227b)

图 3 命题网络示意图(摘自斯坦福课程 CS227b 课件)

命题网络中的与或非门与逻辑电路图类似.图中的竖实线表示存储器,它存储当前回合上级电路传来的值,并在下一回合将这个值传给下级电路.

博弈树的每一条边描述前后两个状态间的关系,而命题网络将前后状态分别作为输入和输出,相当于将博弈树所有的边汇聚到一起,用一个逻辑电路统一描述了所有前后状态间的关系.可以想象,如果将一棵随机生成的博弈树转化成命题网络,由于编码的无规律性,并不能降低描述的复杂度.但现实中,大多数游戏的状态具有良好的局部性,命题网络可以利用这种局部性降低描述的复杂度.例如,棋盘游戏的状态是由一个个落子点组合而成的,玩家的某个动作通常只对某个或某几个落子点产生影响,这就是一种局部性.命题网络用命题推理的形式统一描述这种局部的关系,合并了大量博弈树中描述这种关系的边,从而降低了描述的复杂度.

命题网络从一个相对精简的角度刻画了游戏,其描述是建立在游戏状态的局部关系之上的.它的结构也包含了许多游戏知识,例如,Cox 等人利用命题网络检测游戏的可拆分性[14].

2.2 游戏知识获取

游戏知识可能有无穷多种,这里讨论比较具有代表性的两种:游戏状态估值和游戏结构.游戏状态估值一般通过搜索博弈树获取,而游戏结构则主要通过分析命题网络得到.

2.2.1 游戏状态估值

如第 2.1.1 节所述,游戏状态的价值是游戏的关键知识.由于资源有限,计算游戏状态的精确值是不现实的,

估值便成为常见的替代方案.最朴素的估值方法是构造一个静态估值函数,它以游戏状态为输入,估值为输出.这种估值方法强烈依赖于函数构造者的游戏经验.更进一步的估值方法是局部展开法,它在待估值节点处局部展开博弈树,对展开部分的节点做静态估值,再通过某种方式计算出待估值节点的估值.这种估值方法将博弈树的局部结构信息与静态估值相结合,通常更为准确.局部展开博弈树的过程也是搜索博弈树的过程,在 GGP 实践中,主要有基于评估的搜索和基于模拟的搜索.

2.2.1.1 静态估值函数

静态估值函数是对一个游戏状态直接评估价值的函数.对于特定的游戏,人们往往能从游戏状态中归纳出一些有用的局部结构,这些局部结构被称为特征.常见的特征包括五子棋中的“活三”、象棋中的“车”、围棋中的“眼”等.一个状态可以具有多个特征,一个特征也可以被许多状态共同拥有.人们认识到,同一个特征对不同状态的估值的贡献往往是相似的.因此,可以从少量状态归纳出贡献明显的特征,再将这些特征应用于其他状态的估值,以此来构造静态估值函数.

在 GGP 框架下,机器玩家从接触游戏规则到开始玩游戏只有很短的时间,这使得发现并利用特征构造静态估值函数成为挑战.GGP 游戏的特征可以大致分为如下 3 类:

- 适用于特定游戏的专用特征.由于游戏规则不可预知,专用特征无法被事先写入程序,必须由程序自行从游戏规则中生成.由于特征的数量太大,目前尚未出现有效的算法自动生成有效的专用特征.Kuhlmann[15]提出了特征树(feature tree)结构表示游戏的特征,并提出了一种生成和筛选特征的方法,但该方法在实验中效果并不理想.Haufe 等人[16]提出的 State Sequence Invariants 可以用于专用特征的表示,但是其工作未涉及专用特征的生成和筛选.
- 适用于所有游戏的通用特征.通用特征是与特定游戏无关的,人们可以事先总结和定义.Clune[6]应用了行动力(mobility)特征,即玩家在某一游戏状态下的合法动作数量,这种特征在一定程度上反映了玩家在游戏中的生存能力(一般认为合法动作数量越多,生存能力越强).
- 介于通用和专用之间的特征.这类特征具有一定的适用范围,适用于某一类别的游戏.Kuhlmann 等人[17]用语法结构模式匹配的方法检测游戏是否具有棋子、棋盘等特征,这类特征仅适用于棋盘游戏.

2.2.1.2 基于评估的搜索

在 GGP 的语境中,基于评估的搜索(evaluation-based search approach)是指传统的极大极小搜索及其变种.这种搜索方式有两个特点:一是以平衡的方式扩展博弈树,二是需要一个静态估值函数.

以平衡的方式扩展博弈树是违反人类经验的.人在下分支因子很大的棋时,通常花很短的时间识别并排除前途较差的走法,而把大部分时间花在深入探索前途较好的走法上.α-β剪枝和静止期搜索(quiescence search)等技术应用了人的这种思想,对重要的分支做更深入的搜索,在一定程度上打破了平衡.但真正将非平衡扩展的思想贯彻到底的,是后文将要讨论的基于模拟的搜索.

如第 2.2.1.1 节所述,在 GGP 框架下构造静态估值函数是困难的.即使构造出了有效的静态估值函数,其所消耗的计算资源也将限制搜索的节点数.

在早期的国际 GGP 比赛中,基于评估的搜索占据着主导地位,后来逐渐被基于模拟的搜索取代.虽然如此,静态估值函数的构造还是值得继续探索的.

2.2.1.3 基于模拟的搜索

在 GGP 的语境中,基于模拟的搜索(simulation-based search approach)是指蒙特卡洛树搜索(Monte Carlo tree search,简称 MCTS)算法[18]及其变种.与基于评估的搜索方法不同,基于模拟的搜索方法以非平衡的方式扩展博弈树,且不需要静态评估函数.

基于模拟的搜索算法由选择(selection)、扩展(expansion)、模拟(simulation)、反向传播(backpropagation)这 4 个步骤重复执行而实现,如图 4 所示[19].第 1 步,从博弈树的根节点开始,递归地从子节点中选取最佳的节点,直到选中一个叶节点为止;第 2 步,扩展选中的叶节点,即,将叶节点的一个子节点加入博弈树;第 3 步,从新加入的子节点开始做蒙特卡洛模拟,即随机地进行游戏直到结束;第 4 步,将蒙特卡洛模拟的结果反馈到叶节点及其

各个祖先节点处.其中,第 1 步中最佳节点的定义是 Upper Confidence Bounds(UCB)值最大的节点,待选择节点 i 的 UCB 值的计算公式[20]如下:

$$v_i + C \times \sqrt{\frac{\ln N}{n_i}},$$

其中,C 为常数;v_i 代表节点 i 的估值;n_i 代表节点 i 被访问的次数;N 代表所有待选择节点被访问的总次数,即 $\sum n_i$.在 MCTS 中,v_i 的值通过第 4 步反向传播得到,即 v_i 等于从节点 i 及其子孙节点开始的蒙特卡洛模拟结果的平均值.

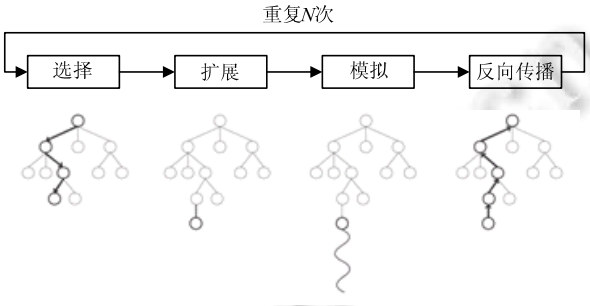

Fig.4 Illustration of MCTS algorithm [19]

图 4 蒙特卡洛树搜索算法示意图[19]

自从 Björnsson 等人将基于模拟的搜索引入国际 GGP 比赛[21]以来,该算法表现突出,目前已成为主流的算法.它用随机模拟的方法评估游戏状态,回避了通用游戏难以构造静态评估函数的问题.此外,大量的随机模拟适合并行计算也使得基于模拟的搜索具有优势.

尽管基于模拟的搜索已经在实践中取得成功,但其依然具有很大的改进空间.例如,如何在搜索时获取和应用游戏知识以提高搜索效率,就是一个值得关注的问题.Finnsson 等人尝试使用游戏知识指导“模拟”过程[22],在一定程度上提升了搜索的效率.

2.2.2 游戏结构

游戏结构也是一种游戏知识.博弈树、状态机、命题网络、GDL 等不同的游戏表示方式从各自的角度描述游戏,其反映的游戏结构也是不同的.博弈树和状态机规模太大,利用它们的结构特征还比较困难,因此,这里主要讨论命题网络和 GDL 的结构.

2.2.2.1 游戏对称性和等价性

游戏对称性包括玩家的对称性、动作的对称性、状态的对称性等,现实中的游戏大多具有对称性.发现游戏的对称性,有助于提高搜索的效率,因为对称的游戏元素可以共享估值.GDL 和命题网络用命题的形式表达玩家、动作、状态等游戏元素,因此,这些游戏元素的对称就表现为命题的对称.发掘命题的对称,可以通过分析 GDL 或命题网络的结构实现[23].

游戏等价性是指两个不同游戏的玩家、动作、状态之间存在某个一一映射的关系,在这个映射关系下,游戏的状态更新、合法动作、终局得分等都是等价的,比如 Tic-tac-toe 和 Number Scrabble[24].如果能够发现一个新游戏和已知的游戏等价,就可以利用已知游戏的知识来玩新游戏.虽然等价游戏存在的可能性很低,但它是讨论游戏相似性的基础,而知识在相似游戏之间的迁移是很有思考价值的问题.

证明两个游戏等价的过程与发掘一个游戏的对称性的过程具有很强的关联性.事实上,判断游戏是否等价的方法是考察两个游戏的 GDL 或命题网络是否同构,而发掘游戏对称性的方法是考察 GDL 或命题网络的所有自同构.因此,这两个问题在解决方法上是相通的.

2.2.2.2 游戏拆分

在近年的国际 GGP 比赛中,组合游戏被作为一个考察的重点.如第 1.1.2 节所述,组合游戏是指将若干个相

同或不同的游戏通过某种方式组合在一起,形成一个新的游戏.与之相对应的技术就叫做游戏拆分(game factoring).

游戏有不同的组合方式,一种简单的组合方式是两个玩家同时玩两个相同的游戏,每个回合只能选择其中一个游戏做动作.比如,两个玩家同时下两盘国际象棋,但每个回合只能选择其中一个棋盘下,最后,任意一个棋盘结束则整个游戏结束.在这样的规则下,玩家识别出两个相互独立的棋盘是十分重要的,它可以使整个游戏过程的分支因子下降一半.游戏拆分可以通过分析命题网络实现[14].通过分析命题之间的依赖关系,有可能把命题网络拆分成几个部分,形成独立的几个子游戏.

2.3 游戏知识迁移

GGP 的框架为讨论知识在不同游戏之间迁移提供了平台,这是 GGP 研究不同于特定游戏研究之处,也是极具挑战之处.哪些知识可以在哪些游戏之间迁移?知识迁移可以在多大程度上提高对新游戏的认识?这些问题都有待深入探索.

2.3.1 知识在相同游戏之间迁移

如第 2.2.2.1 节所述,表现形式不同的游戏可能在本质上可能是相同的.发现这一点,就可以将旧游戏的知识迁移到新游戏中.除完全相同外,部分相同和动态相同也是值得关注的.

考虑到一个游戏可能可以拆分成若干个子游戏,若两个游戏存在相同的子游戏,则可以在子游戏之间传递知识,这种情况就是游戏的部分相同.

动态相同是指两个游戏在初始状态下不同,但在游戏过程中可能达到相同的状态,如正常中国象棋与“让子”的中国象棋可以在游戏过程中达到相同的状态.两个动态相同的游戏可以共享状态估值等知识.

2.3.2 知识在相似游戏之间迁移

游戏的相似性包括多种多样的情况,一些常见的情况包括:

- 仅棋盘大小或形状不同,如棋盘大小为 15×15 的五子棋和棋盘大小为 16×16 的五子棋;
- 仅棋子不同,如围棋和“让 2 子”的围棋;
- 仅步数限制不同,如不限回合数的国际象棋和限制 200 回合结束的国际象棋.

在这些情况下,关于游戏特征、游戏结构的知识大多是可迁移的.例如,人们下五子棋时,一般不会因为棋盘大小相差一行一列而改变游戏策略;下“让 2 子”围棋时,对于各种棋型的估值与正常围棋是相同的;下限制 200 回合的国际象棋时,开始阶段几乎不会考虑回合数的限制.

然而,量变会引发质变.棋盘为 5×5 和 15×15 的五子棋、正常围棋和“让 50 子”的围棋、不限回合数和限制 30 回合结束的国际象棋肯定有很大的不同,知识在它们之间的传递要打很大的折扣.如何用一般的方法量化游戏的相似程度,从而确定知识在多大程度上可以传递,是一个有待讨论的问题.

此外还需要注意到,有时游戏规则微小的改动会导致游戏变得很不一样.比如,考虑正常的五子棋游戏和“先连成五子者失败”的自杀五子棋游戏,二者具有相同的游戏状态空间、合法动作、结束条件,而只有结束状态下的游戏得分不同,其游戏策略就大相径庭,很难说清有哪些知识是可以在二者之间迁移的.可见,量化两个游戏的相似程度是一个相当困难的问题.

对于这个问题,Kuhlmann 做了较为深入的探索[15].

(1) 定义了几种游戏相似的情况(如棋盘大小、步数限制等);

(2) 通过将游戏规则转化成 Rule Graph 并分析两个游戏 Rule Graph 的相似性,得到游戏的相似性;

(3) 对于相似的游戏,定义了对称性、估值函数、蒙特卡洛搜索树这 3 种可以迁移的知识,并分别提出了迁移的方法,最终在对称性迁移和估值函数迁移的实验中取得了理想的结果.

Kuhlmann 的工作证明了知识在相似游戏之间迁移的可行性,但也存在着诸多局限,包括游戏相似情况的局限性、估值函数迁移仅适用于用强化学习(reinforcement learning)训练估值函数、蒙特卡洛搜索树迁移取得了负面的实验结果等.

2.3.3 知识在任意游戏之间迁移

能够在任意游戏之间迁移的知识,可被称为通用知识.通用知识一定是关于最基本的游戏元素的知识,这些游戏元素在任何游戏中都存在,比如博弈树的局部结构、命题网络的局部结构等.通用知识一旦被掌握,即可应用于所有的游戏,因而是非常重要的.

通用知识包括上文提到的通用游戏特征的估值.博弈树的局部结构就是一种通用游戏特征.在一个游戏中对不同的博弈树局部结构进行估值,再将估值结果应用于另一个游戏是有可能的.Banerjee 等人[25]提出了针对特定游戏生成常见的博弈树局部结构的方法,并将这些局部结构作为游戏的特征.实验结果表明,在已知游戏和未知游戏使用相同的游戏特征进行强化学习训练的前提下,将已知游戏的训练结果作为未知游戏的初始训练参数,将有助于提高未知游戏训练的质量或缩短训练时间.

此外,上文提到的获取游戏知识的方法,比如蒙特卡洛树搜索算法,在某种意义上也是一种通用知识.在这种意义上,任何有助于通用游戏程序进步的方法都可以被视为游戏的通用知识.

3 GGP 研究的若干展望

可以预见,GGP 研究在游戏知识表示方面将稳步扩展,在游戏知识获取方面将更加深入,而在游戏知识迁移方面将迎来曙光.下文分别从 GGP 规则、GGP 玩家和 GGP 平台的角度展望 GGP 研究的未来发展方向.

3.1 GGP规则展望

GGP 规则包括 GGP 比赛采用的语言、赛制等,它决定了 GGP 研究涵盖的范围.GDL 作为目前的 GGP 游戏描述语言,其描述能力具有相当的局限性,如第 1.2.4 节所述.随着研究的深入,扩展游戏语言的描述能力是必然的趋势.

- 描述非确定性的、信息完全的游戏,GDL-II 已经做出了很好的探索;
- 描述非回合制的游戏,如星际争霸、帝国时代等即时战略游戏.

此外,放弃描述游戏规则也是可能的发展方向.如 Google 公司对 Atari 游戏的研究[12],机器玩家不知道确切的游戏规则,只通过模拟玩游戏,得到游戏状态、动作、得分的历史经验来学习游戏策略.

3.2 GGP玩家展望

现阶段,GGP 玩家在知识获取方面已经形成了比较成熟的方法,而在知识迁移方面还可以做更多的探索.以 2014 年国际 GGP 比赛冠军 Sancho 为例,它采用并行的蒙特卡洛树搜索算法,且在实现层面对命题网络进行了大量优化.它能在游戏开始阶段对游戏进行分析,从而检查游戏的可拆分性以及调整各种启发函数的参数.它获取游戏知识的能力超过了竞争对手,这是它成功的主要原因.此外,它在游戏知识迁移方面也做了一些探索,比如判断新游戏是否与已知的游戏相同,从而利用已有的游戏经验.

3.2.1 注重游戏知识对蒙特卡洛树搜索的帮助

虽然基本的蒙特卡洛树搜索算法不需要游戏知识,但结合游戏知识可以使它变得更好.游戏知识可以在多方面提升蒙特卡洛树搜索算法,包括:

- 在 UCB 公式中加入游戏知识相关的项,影响“选择”阶段;
- 在“模拟”阶段对合法动作或后继状态做快速的评估,从完全随机的模拟改为带概率分布的模拟;
- 利用游戏结构的局部性,动态地拆分游戏,运用局部搜索提升搜索效率.

3.2.2 应用其他人工智能算法

目前,许多人工智能算法被成功地应用到了 GGP 的研究中,最典型的就是蒙特卡洛树搜索算法.将来更多的人工智能算法或许可以被应用到 GGP 研究中,包括:

- 用遗传算法生成和筛选游戏特征,可以从最基本的游戏特征通过逻辑组合逐渐演变出复杂的特征;
- 用深度神经网络构建游戏特征,深度神经网络的层次对应游戏特征的复杂程度;
- 将更多的专用游戏算法引入 GGP.

3.2.3 使用更加强大的硬件

GGP 比赛没有限制硬件的使用,因而使用更强大的硬件是提升 GGP 玩家水平的重要方法.目前,硬件方面的提升主要体现在使用更多计算核心,实现蒙特卡洛树搜索的并行计算.将来,更多的硬件可能被利用:

- 使用 FPGA 实现命题网络,极大地加快游戏状态的更新过程;
- 使用 GPU 做并行的蒙特卡洛模拟;
- 使用超级计算机做大规模的特征生成和筛选、蒙特卡洛模拟等.

3.3 GGP平台展望

GGP 平台具有引领 GGP 研究的作用,应该在许多方面开拓进取.

3.3.1 添加有针对性的游戏

目前,GGP 游戏的数量级还较小,而引入更多的游戏无疑将增强 GGP 研究的通用性.GGP 平台除了将各种已经存在的游戏翻译成 GDL 之外,还应该人为地创造一些具有特点的游戏.这些游戏可以反映人类对某些特定概念的认识,比如第 1.1.2 节介绍的组合游戏反映了人类对组合的认识.其他例子包括对单调性的认识等.这样的认识在人看来一目了然,但对于机器来说可能是很困难的.有针对性地创造这类游戏,将对机器认识到这些概念起到积极作用.

3.3.2 鼓励探索知识迁移

为了鼓励 GGP 研究者探索知识迁移,GGP 平台可以提供更有针对性的比赛机制.由于直接研究通用知识是比较困难的,可以先研究知识在相似游戏中的迁移.例如,将 Connect Four 游戏的各种变体划入一个集合,再将这个集合拆分为训练集和测试集.GGP 玩家可以拥有一段时间在无人指导的环境下对训练集中的游戏进行学习,然后在正式比赛时使用测试集中的游戏.在这种赛制下,善于将已有的知识迁移到相似游戏的 GGP 玩家将获得优势.

4 结束语

本文介绍了 GGP 研究的基本框架和主要问题,并对 GGP 未来发展做了若干展望.GGP 研究的目标是提升机器的通用游戏智能,这与特定游戏的研究是不同的,代表着人工智能未来的一个发展方向.GGP 研究至今已取得了一定的进展,但在深度和广度两方面都还有很大的拓展余地,因而是值得广大研究者关注的.

致谢 在此,向对本文的工作给予支持和建议的北京大学许卓群教授表示感谢.

References:

[1] Genesereth M, Love N, Pell B. General game playing: Overview of the AAAI competition. AI magazine, 2005,26(2):62−72. [doi: 10.1609/aimag.v26i2.1813]

[2] Allis LV. Searching for Solutions in games and artificial intelligence [Ph.D. Thesis]. Maastricht: the Netherlands: University of Limburg, 1994.

[3] Love N, Hinrichs T, Haley D, Schkufza E, Genesereth M. General game playing: Game description language specification. Stanford: Stanford University, 2008.

[4] Schiffel S, Thielscher M. A multiagent semantics for the game description language. In: Duval B, van den Herik J, Loiseau S, Filipe J, eds. Proc. of the Agents and Artificial Intelligence. Berlin, Heidelberg: Springer-Verlag, 2010. 44−55. [doi: 10.1007/978-3-642-11819-7_4]

[5] Thielscher M. A general game description language for incomplete information games. In: Proc. of the 24th AAAI Conf. on Artificial Intelligence. Menlo Park: AAAI Press, 2010. 994−999.

[6] Clune J. Heuristic evaluation functions for general game playing. In: Proc. of the 22nd AAAI Conf. on Artificial Intelligence. Menlo Park: AAAI Press, 2007. 1134−1139.

[7] Schiffel S, Thielscher M. Fluxplayer: A successful general game player. In: Proc. of the 22nd AAAI Conf. on Artificial Intelligence. Menlo Park: AAAI Press, 2007. 1191−1196.

[8] Finnsson H, Björnsson Y. Simulation-Based approach to general game playing. In: Proc. of the 23rd AAAI Conf. on Artificial Intelligence. Menlo Park: AAAI Press, 2008. 259−264.

[9] Méhat J, Cazenave T. A parallel general game player. Künstliche Intelligenz, 2011,25(1):43−47. [doi: 10.1007/s13218-010-0083-6]

[10] Levine J, Congdon CB, Ebner M, Kendall G, Lucas SM, Miikkulainen R, Schaul T, Thompson T. General video game playing. Artificial and Computational Intelligence in Games, 2013,6:77−83. [doi: 10.4230/DFU.Vol6.12191.77]

[11] Bellemare MG, Naddaf Y, Veness J, Bowling M. The arcade learning environment: An evaluation platform for general agents. Journal of Artificial Intelligence Research, 2013,47:253−279.

[12] Mnih V, Kavukcuoglu K, Silver D, Rusu AA, Veness J, Bellemare MG, Graves A, Riedmiller M, Fidjeland AK, Ostrovski G, Petersen S, Beattie C, Sadik A, Antonoglou I, King H, Kumaran D, Wierstra D, Legg S, Hassabis D. Human-Level control through deep reinforcement learning. Nature, 2015,518(7540):529−533. [doi: 10.1038/nature14236]

[13] Neumann JV. Zur theorie der gesellschaftsspiele. Mathematische Annalen, 1928,100(1):295−320. [doi: 10.1007/BF01448847]

[14] Cox E, Schkufza E, Madsen R, Genesereth M. Factoring general games using propositional automata. In: Proc. of the IJCAI Workshop on General Intelligence in Game-Playing Agents (GIGA). Menlo Park: AAAI Press, 2009. 13−20.

[15] Kuhlmann GJ. Automated domain analysis and transfer learning in general game playing [Ph.D. Thesis]. Austin: The University of Texas at Austin, 2010.

[16] Haufe S, Schiffel S, Thielscher M. Automated verification of state sequence invariants in general game playing. Artificial Intelligence, 2012,187:1−30. [doi:10.1016/j.artint.2012.04.003]

[17] Kuhlmann G, Dresner K, Stone P. Automatic heuristic construction in a complete general game player. In: Proc. of the 21st AAAI Conf. on Artificial Intelligence. Menlo Park: AAAI Press, 2006. 1457−1462.

[18] Browne CB, Powley E, Whitehouse D, Lucas SM, Cowling P, Rohlfshagen P, Tavener S, Perez D, Samothrakis S, Colton S. A survey of Monte Carlo tree search methods. IEEE Trans. on Computational Intelligence and AI in Games, 2012,4(1):1−43. [doi: 10.1109/TCIAIG.2012.2186810]

[19] Chaslot G, Bakkes S, Szita I, Spronck P. Monte-Carlo tree search: A new framework for game AI., 2008. In: Proc. of the 4th Artificial Intelligence and Interactive Digital Entertainment Conf. AAAI Press, 2008. 216−217.

[20] Auer P. Using confidence bounds for exploitation-exploration trade-offs. Journal of Machine Learning Research, 2002,3:397−422.

[21] Björnsson Y, Finnsson H. Cadiaplayer: A simulation-based general game player. IEEE Trans. on Computational Intelligence and AI in Games, 2009,1(1):4−15. [doi: 10.1109/TCIAIG.2009.2018702]

[22] Finnsson H, Björnsson Y. Learning simulation control in general game-playing agents. In: Proc. of the 24th AAAI Conf. on Artificial Intelligence. Menlo Park: AAAI Press, 2010. 954−959.

[23] Schiffel S. Symmetry detection in general game playing. In: Proc. of the 24th AAAI Conf. on Artificial Intelligence. Menlo Park: AAAI Press, 2010. 980−985.

[24] Pell B. Strategy generation and evaluation for meta-game playing [Ph.D. Thesis]. Cambridge: University of Cambridge, 1993.

[25] Banerjee B, Stone P. General game learning using knowledge transfer. In: Proc. of the 20th Int'l Joint Conf. on Artificial Intelligence. Menlo Park: AAAI Press, 2007. 672−677.

张海峰(1989－),男,福建浦城人,博士生,主要研究领域为人工智能.

李文新(1968－),女,博士,教授,博士生导师,CCF 杰出会员,主要研究领域为人工智能.

刘当一(1994－),男,学士,主要研究领域为人工智能.