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.

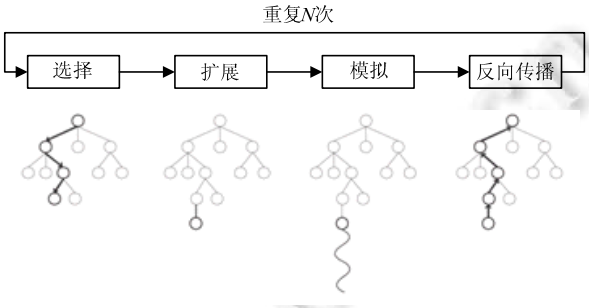

Fig.4 Illustration of MCTS algorithm [19]

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.