

doi: 10.3969/j.issn.1674-8425(z).2014.04.021

# 博弈机器人的行为规划

张小川<sup>a,b</sup>, 侯鑫磊<sup>a,b</sup>, 涂 飞<sup>a</sup>

(重庆理工大学 a. 计算机科学与工程学院; b. 人工智能系统研究所 重庆 400054)

**摘 要:** 博弈行为是博弈机器人智能的外在表现, 行为选择是提高机器人智能的重要手段之一, 而博弈行为体系是行为选择的重要基础。通过模拟人类下棋过程, 提出了博弈机器人的行为过程结构, 依据系统分类思想, 构建了博弈机器人的博弈行为体系, 提出了博弈行为规划方法, 借助六子棋博弈平台, 验证了研究成果的有效性。

**关 键 词:** 机器博弈; 博弈机器人; 行为规划; 六子棋

中图分类号: TP18

文献标识码: A

文章编号: 1674-8425(2014)04-0099-05

## Behavior Planning for the Game Robot

ZHANG Xiao-chuan<sup>a,b</sup>, HOU Xin-lei<sup>a,b</sup>, TU Fei<sup>a</sup>

(a. School of Computer Science and Engineering; b. Institute of Artificial Intelligent ,  
Chongqing University of Technology , Chongqing 400054 , China)

**Abstract:** The game behavior is an outward manifestation of the game robot's intelligence , and the behavior selection is one of the important methods to improve robot intelligence. And the game behavior system is also an important research foundation of behavior selection. This article simulated the process of human game playing , and put forward the behavior process structure for the computer game robot. According to the system classification thought , the article built the robot's computer game behavior system , put forward the behavior planning method , and verified the effectiveness of the research result with the Connect6 game platform.

**Key words:** computer game; game robot; behavior planning; Connect6

智能是生命体最精彩的功能。人类从未停止过对自身及其他自然生命体的思考: 智能的本质是什么? 长期以来, 这是困扰人类的重要科学问题。国内学者钟义信<sup>[1]</sup>认为: 关于智能的本质, 可从“信息-知识-策略-行为”的转换过程获得初

步的认识; 行为是生命体外在的表现形式, 是生命体内在智能的外部表象。Tyrell 认为<sup>[2]</sup>: “行为选择就是从一组可能的候选集中选择最适合的行为。”因此, 行为选择是生命体智能的高级形式, 对行为选择的研究将有助于揭示智能过程和智能的

收稿日期: 2014-01-16

基金项目: 国家自然科学基金资助项目(60443004); 重庆市教委科学技术研究项目(KJ120824)

作者简介: 张小川(1965—), 男, 四川邻水人, 硕士, 教授, 主要从事人工生命、计算机软件方面的研究。

引用格式: 张小川, 侯鑫磊, 涂飞. 博弈机器人的行为规划[J]. 重庆理工大学学报: 自然科学版, 2014(4): 99-103.

**Citation format:** ZHANG Xiao-chuan, HOU Xin-lei, TU Fei. Behavior Planning for the Game Robot [J]. Journal of Chongqing University of Technology: Natural Science, 2014(4): 99-103.

本质,具有重要的科学意义,而行为本身的研究又是行为选择研究的重要基础。

本文以机器博弈为研究原型,以博弈机器人的竞争性行为作为研究对象,首先建立了博弈机器人的行为过程结构,提出了机器博弈的行为体系,为研究机器博弈中的联想推理行为奠定基础。

## 1 博弈机器人的行为过程结构

在双方完备信息博弈活动中,双方都需要考虑对手各种可能的方案、着法,并试图发现、选择对己方有利或合理的方案、着法。这个过程的外在表现就是一个博弈行为序列,其实质是博弈行为选择问题。因此,双方完备信息博弈问题具有3个特点,即双方对弈,严格控时,共同产生了博弈行为时间序列。比如象棋的甲1步→乙1步→甲1步→乙1步……单步博弈行为序列,或如六子棋的甲1步→乙2步→甲2步→乙2步……多步博弈行为序列。信息完备是指博弈各方能获得相同的博弈信息。零和博弈即不存在博弈各方均得利或无利的博弈行为,博弈结果只能是某一方胜、另一方负或和局的状况,且必须是其一状况。据此特点,可以进一步提炼如下博弈规则:

- ① 双方轮流完成博弈行为;
- ② 博弈过程有时间限制;
- ③ 博弈过程是寻找对己方有利的博弈行为的时间序列过程;
- ④ 各方不能干预对方选择;
- ⑤ 比赛结果是赢、输、平局之一,且只能是之一。

假设我方为 $W$ 、对方为 $E$ 并为先手。从上面的分析可知:如果是六子棋博弈的多步博弈行为序列,那么经过 $n$ 步对弈后,就形成如下博弈行为时间序列 $S$ :

$$S = \{s_{E1}, s_{w1}, s_{E21}, s_{E22}, s_{w21}, s_{w22}, \dots, s_{En1}, s_{En2}, s_{wn1}, s_{wn2}\} \quad (1)$$

在式(1)中, $W$ 和 $E$ 之间任何一轮次切换都为双方留下发现、规划自认为有利或合理着法的时空。此处着法可以是单步,也可以是多步所构成的博弈行为序列。因此,通过模仿人类博弈过程,

本文构建了如图1所示的博弈机器人行为过程结构。

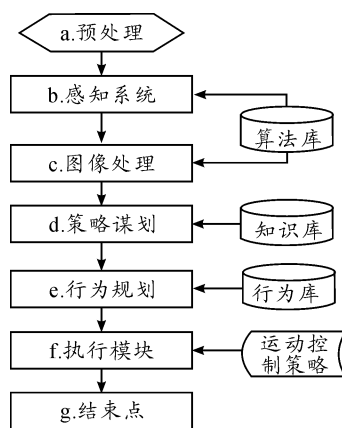


图1 博弈机器人的行为过程结构

在图1中,利用数据库作为机器人的知识源。采用分层递阶的模块化控制思想,将机器人博弈过程分解为相对独立的、耦合度比较低的工作模块,从而形成了机器人的博弈行为过程。其中:模块a完成博弈初始化状态;模块b完成博弈棋盘、博弈现场情况(如是人-机博弈,还需采集人的情感信息等)的规范化处理工作;模块c完成感知模块b的数字化图像处理、分析和判读,形成可支持判断棋盘状态的数据,并借助知识库形成初步的策略,比如攻、防,或吃、兑、弃、马后炮、双迫着、单迫着等比较高层的谋划;模块e针对模块d的策略,借助行为库,规划、选择具体的实施行为,如“马后炮”、“三·三”等定式,规划具体行为或行为序列;模块f完成前述模块的具体动作,是具体的执行机构;模块g主要完成博弈过程的结束扫尾工作,比如保存棋谱、执行机构归位等。

## 2 博弈行为体系

智能是感知、理解、判断、推理、规划和学习的一种能力,其中策略、规划的制定与行为选择过程密切相关。通常将行为选择分为竞争型和合作型2种类型,前者是以最大行为驱动力大于一定阈值为目标的行为选择;后者是以合作行为驱动力强度趋大为原则的行为选择<sup>[3]</sup>。竞争性的博弈行为还具有比较明显的特征,如目的性强、时效性突

出、对抗激烈。博弈各方必须在有限时间内选择、产生合适的博弈着法,为己方博弈进程贡献最大利益值。该利益值通常由相应的评估函数估算<sup>[4]</sup>。因此,着法产生就是博弈过程的核心步骤。而任何着法的产生都源于对当前局势的判断、评估,理论上可以建立一棵完全博弈树,对该博弈树进行完全搜索,即可产生最佳的博弈着法。但现实是目前运算速度最快的计算机都无法满足博弈规则对时间严格限制的需要,这样人们就尝试利用推理技术,化繁为简,构建机器人的推理能力。

通过分析发现:在博弈机器人的棋类博弈活动中,推理过程大多数采用定式型联想式推理方法,也有少部分采用规则型搜索式推理方法。但这些推理方式都需要比较完备的、形式化的行为体系来支撑,这就是本文构建博弈行为库,研究博弈行为规划的重要原因。

从仿人、仿生角度出发,本文采用系统科学的分层分类方法分析和模仿人类的博弈行为,对机器人博弈行为进行表1所示的分层,即将机器人的博弈行为划分为优先顺序从低到高的4个层次:基本行为、技术行为、战术行为和战略行为。

表1 博弈机器人的博弈行为分层体系列表

序号	行为层次	摘要
1	基本行为	博弈机器人的基本动作
2	技术行为	博弈机器人的技术性动作
3	战术行为	博弈机器人的战术性行动
4	战略行为	博弈机器人的战略性行动

1) 基本行为:指博弈机器人的基本动作,基本行为依据不同的棋类博弈规则会有所不同。如吃子、落子、移动等。

2) 技术行为:指博弈机器人在基本行为基础之上所形成的有一定技术含量的技巧性进攻、防守性行动。如象棋中的将、吃、兑等,围棋中的打、提、征、枷、扳、(找)眼位、(计算)气,六子棋的连、挡、断等。

3) 战术行为:指具有一定知识含量、可复用的一些组合动作,是一种行动。此外,失败行为也以案例形式归入本类,以避免历史性悲剧的再次

发生。如象棋的马后炮、双将,围棋中的劫、大飞、双灯,六子棋的连五、连四、连三等。

4) 战略行为:为实现某个特定战略意图,由2个或2个以上技术行为、战术行为组成的动作序列是一种行动方案。如象棋的仙人指路、各残局定式,围棋的点三、三、大场、终局定式,六子棋的双迫着、单迫着等。

技术行为是以基本行为为基础,战术行为又是以技术行为为基础,战略行为则是建立在技术行为和战术行为基础之上的。因此,4类行为的层次就如堆积台阶一样,逐层向上堆积,低层行为是上层行为实现的保障,最终组成博弈行为体系。

### 3 博弈行为规划

博弈行为是博弈机器人智能的外在表现,博弈行为选择是博弈机器人内在智能的表现形式之一。实际上,人类的高级博弈策略,如“以退为进”、“得寸进尺”、“欲擒故纵”、“围魏救赵”等,都是在经验、先验知识基础上进行规划、布局的产物。

在文献[3]中,笔者所在团队建立了人工生命体行为选择模型。该模型是一个基于优先度的行为选择分级模型。它能较好地体现动物在特别饥饿时,冒着被捕猎的危险实施“顺手牵羊”的进食行为,体现出动物的高级智能。这个模型利用优先度方法,从另外角度提出了一种可行的行为选择机制。因此,应用优先度思想,对表1的行为进行规划,就能减少搜索的盲目性,提高博弈行为搜索的准确性和搜索效率。这样,获取优先度数值就成为首要问题。

实际上,要获取表1中各具体行为的优先度值,通常可以采用2种方法:一是利用人类博弈的先验知识,在此基础上,利用各类算法进行优选、调整,这也是机器博弈领域的主要方法。如1997年IBM“更深的蓝”打败世界国际象棋大师卡斯帕罗夫,就得到了1名国际特级大师、4名电脑专家的先验知识支持;二是利用机器学习技术,让机器人在不断的博弈实战中积累博弈经验,逐渐形成自己的博弈先验知识及其优先度值,这种方法也

是不少学者常选的方法,但其不确定性因素太多,目前表现的效果较差。另外,不少学者融合应用2种方法,取长补短,也取得了实战效果。

### 3.1 博弈行为的形式化

由于目前计算机结构基本上还是二进制的冯·诺依曼体系,因此,无论采用什么方法获取行为优先度,都必须对式(1)所代表的博弈问题进行形式化、数字化。通过分析,可以将其描述为如下形式:

定义1 基于四元函数的博弈过程

$$S = \langle \Omega_c, \Omega_o, c^{(o)}, c^{(g)} \rangle \quad (2)$$

其中:  $\Omega_c = \{c\}$  表示博弈空间,即棋局或博弈状态空间,是一个二维状态空间;  $\Omega_o = \{o\}$  表示算子空间,即操作或规则的集合,是图1知识库的主体内容;  $c^{(o)} \in \Omega_c$  表示当前博弈状态; 式(1)中开始时的博弈初始状态,也称开局状态;  $c^{(g)} \in \Omega_c$  表示博弈目标、结果,  $\Sigma c^{(g)}$  是所追求博弈目标的集合。

定义2 博弈行为集合  $A$ 。假设  $A$  为博弈机器人所有可能行为所构成的集合,其取值包含基本行为  $Ab$ 、技术行为  $At$ 、战术行为  $Aw$  和战略行为  $As$  共4个子集,如  $Ab = \{Fc - \text{落子}, Ec - \text{吃子}, Mc - \text{移动}\}$ ,则  $A$  取值为

$$A = \{Ab, At, Aw, As\} \quad (3)$$

在实施中,进一步细分上述行为,如基本行为子集  $Ab$  中的  $Mc$  包含图2中8个方位的移动动作,即从左到右、从前到后构成为  $\{LF, F, RF, L, R, LB, B, RB\}$ 。

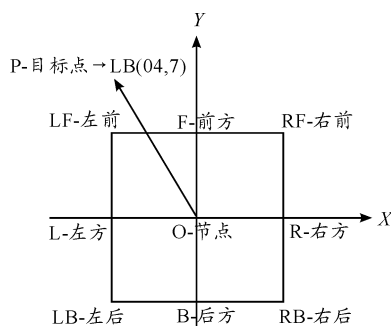


图2 棋盘上节点周围的8个方位图

由于受具体行棋规则所限,8个方位并不是各类博弈都拥有的行为。为便于使用,在实施中,进一步按表2进行8位编码。

表2 博弈行为编码列表

博弈行为	编码	摘要
Mc-LF	10 000 010	向左前方移动
Mc-F	10 000 110	向前方移动
Mc-RF	10 001 010	向右前方移动
Mc-L	10 001 110	向左方移动
Mc-R	10 010 010	向右方移动
Mc-LB	10 010 110	向左后方移动
Mc-B	10 011 010	向后方移动
Mc-RB	10 011 110	向右后方移动

在编码规则表2中,第1位表示行为类型、第2~6位表示行为、第7位表示棋种类、第8位为标示位。

需要注意的是,对于不同棋类,集合  $A$  的取值范围大不相同,算子空间  $\Omega_o$  的运算规则也是极不相同。比如,在象棋中,有吃子着法,棋盘上的棋子随棋局发展会越来越少,而诸如五子棋、六子棋的  $K$ -子棋就没有吃子着法,随着棋局推进和不断落子,棋盘上的棋子数目会越来越多。这些具体差异需要通过算子、编码予以规范,通过第7、8位编码的不同值来体现不同棋类及其运算规则。

### 3.2 博弈行为规划示例

下面以六子棋“迫着”算法为例,说明博弈行为的规划过程。“迫着”就是甲方需要下  $t$  颗子来避免乙方连  $k$ ,则称该乙方有“迫着”。当  $t=2, k=6$  时,就是“双迫着”,即甲方要下2颗子才能避免乙方连六的“迫着”。当  $t=1, k=6$  时,就是“单迫着”,即甲方需要下1颗子才能避免乙方连六的“迫着”。通过“迫着”,强迫对方只能在指定点下棋,否则将输棋,这极大地限制了对方的走棋,从而使对方处于被动挨打的地步。

为实施战略性行为,进行行为规划如下:

① 搜索六子棋博弈空间,发现  $W, E$  双方的连四、连五技术行为(棋形)。

在六子棋中,表1战术行为中,理论上连四行为有32种,连五行为有8种。搜索棋盘状态空间可能存在的上述战术行为,并标记和单独存储。

② 计算  $E$  方“必堵点”  $\text{PlugChessN}^{[5-6]}$

$$\text{PlugChessN} = \frac{(6 - \text{type}) \text{typeN} - \text{typeReTimes}}{6 - \text{type}}$$

必堵点数 PlugChessN 就是需要立即堵住对方特殊棋形否则我方必输所需的棋子数; type 表示战术行为的子类,此处为某方棋子连续成形的棋形,如连五、连四等,此处取值为 4 或 5; TypeN 表示对应 type 数的棋形数目; typeReTimes 表示对应 type 数棋形交叉点的重复数量。

③ 依据 PlugChessN 执行相应行为。

If PlugChessN = 0 then plan-behavior-1

plan-behavior-1: 执行己方最大优先度值的指定行为

If PlugChessN = 1 then plan-behavior-2

plan-behavior-2: 在必堵点落 1 子再执行己方最大优先度值的指定行为

If PlugChessNum = 2 then plan-behavior-3

plan-behavior-3: 在必堵点落 2 子

If PlugChessNum > 2 then failure (认输)

注: 六子棋博弈基本规则是在  $19 \times 19$  围棋棋盘上,双方交替落子,先手方第 1 手只能下 1 颗棋子,此后各方交替每次落子 2 颗,当某方率先连六成功时为胜,如果棋盘下满无胜负即为和。

上述研究成果已应用于重理工骑士队的六子棋博弈项目,并参加全国大学生锦标赛,获得了亚军的优异成绩。这从一个侧面证明了本文研究思想、成果的有效性。

## 4 结束语

本文以博弈机器人的完备信息博弈行为作为研究对象,为机器人利用联想推理技术,实现博弈树的剪枝奠定基础。本文的研究对其他机器人行为的研究具有直接的借鉴意义。

## 参考文献:

- [1] 钟义信. 人工生命由分立走向和谐 [M]. 北京: 科学出版社, 2006.
- [2] Tyrrell T. Computational mechanisms for action selection [D]. Edinburgh: University of Edinburgh, 1993.
- [3] 李祖枢, 谢汝林, 张小川. 人工生命体行为选择及其进化研究 [J]. 模式识别与人工智能, 2005, 18(3): 49-55.
- [4] 吕艳辉, 宫瑞敏. 计算机博弈中估值算法与博弈训练的研究 [J]. 计算机工程, 2012, 38(11): 163-166.
- [5] 张世强. 基于六子棋机器博弈平台的智能决策系统的研究与设计 [D]. 重庆: 重庆理工大学, 2009.
- [6] 陈东立, 姚杰, 史艳维. 基于纯策略的区间数矩阵博弈模型的研究 [J]. 重庆理工大学学报: 自然科学版, 2012, 26(2): 107-111.

(责任编辑 刘 舸)