

# 不完全信息博弈的机器人对抗决策

史晓茹, 侯媛彬, 张涛

(西安科技大学 电气与控制工程学院, 陕西 西安 710054)

**摘要:**针对机器人比赛时局势的动态变化给机器人对抗决策博弈局面带来的不完全性问题,提出了豪尔绍尼转换和贝叶斯均衡相融合的不完全信息博弈算法,该算法克服了博弈局势中对未知信息的盲目“猜测”。以机器人足球比赛时的数据为背景建立不完全信息博弈模型,研究机器人的决策对抗系统。仿真结果表明,不完全信息博弈算法可以使得机器人进行较优策略的选择,从而进一步提高机器人在比赛中的自主性和智能性。

**关键词:**不完全信息博弈;贝叶斯均衡;豪尔绍尼转换;对抗决策;机器人比赛

**中图分类号:** TP391 **文献标识码:** A **文章编号:** 1673-4785(2011)02-0147-05

## The decision-making system of robots based on an incomplete information game

SHI Xiaoru, HOU Yuanbin, ZHANG Tao

(Xi'an University of Science and Technology, Xi'an 710054, China)

**Abstract:** While considering the incomplete game situation of decision-making in robot competition caused by the dynamic changes of the situation, games of incomplete information fusing Harsanyi conversion and Bayesian equilibria were proposed. This enabled the elimination of blind guessing of unknown information in game circumstances. Based on the robot soccer competition, a game model of incomplete information was established and a decision-making robot system was researched. The simulation results showed that the incomplete information game algorithm can help robots to choose better strategies, further improving the independence and intelligence of robots in competition.

**Keywords:** incomplete information game; Bayesian equilibria; Harsanyi conversion; decision-making; robot competition

博弈论(game theory)作为一门现代科学体系,起源于20世纪初,在二战后发展成为一门完整而丰富的理论科学。博弈算法经过不断的研究、改进与提高,目前已经成为解决诸多动态复杂环境中决策问题的一种重要方法。以往的机器人决策方法主要采用了有限状态机法<sup>[1]</sup>和智能体强化学习法<sup>[2]</sup>。常规的方法都是建立在信息完全已知的前提条件下的。足球机器人在进行比赛时,虽然场地的信息及周围的环境信息是可以通过视觉系统完全已知,但是由于局势的动态变化导致敌方的信息无法直接获取。随着博弈理论的演化和发展,不完全信息博弈理论的应用对常规的方法是有很好的补充和优化。

近年来,博弈论在机器人方面的应用都有了一定的研究<sup>[3]</sup>,意大利的学者 N. Basilico 和 N. Gatti<sup>[4]</sup>在其论文中将博弈论理论应用于模拟机器人学,并提出了 Leader-Follower 平衡;美国耶鲁大学的专家 M. Beetz、S. Buck 等人<sup>[5]</sup>在其论文中将合作概率博弈应用于足球机器人策略选择中,提高了机器人应对复杂环境的能力;国内的学者柳长安等<sup>[6]</sup>在对已有仿真足球机器人协作防守战术分析研究的基础上,运用合作  $n$  人博弈理论对防守战术进行分析研究,提出了一种基于合作 4 人博弈的足球机器人协作防守对策,并建立了数学模型,最后通过仿真实验验证了其正确性和有效性;王云等<sup>[7]</sup>在协调博弈信念形成过程的基础上,利用智能体策略相似性,提出换位推理的协调博弈学习方法,通过信念修正模型将客观观察行为和主观预测行为结合在一起,从而能取得更好的协调性能;宋

收稿日期:2010-04-15.

基金项目:陕西省自然科学基金资助项目(2009JM8002).

通信作者:史晓茹. E-mail: sxiaoru\_girl@163.com.

明鑫等人<sup>[8]</sup>将典型的“囚徒困境”博弈模型应用在机器人比赛的策略对抗中,但是这些模型的应用都是以信息是完全已知为前提的.本文针对机器人比赛时信息的不完全性,探讨了将豪尔绍尼转换和贝叶斯均衡相融合的不完全信息博弈理论,该理论通过引入虚拟的参与者,使得机器人所选择的保守决策点范围缩小,从而可以有效地提高机器人比赛时对未知信息的推理能力.

## 1 不完全信息博弈理论

**定义1** 不完全信息.不完全信息博弈理论中的不完全信息专指一种博弈局势中,人对其他局中人(或者他自己)与该种博弈局势有关的事前信息了解不充分.这里的事前信息是指关于在博弈实际开始之前局中人所处地位或者状态的信息,这种地位与状态对于博弈局势会产生影响<sup>[9-10]</sup>.

**定义2** 贝叶斯均衡.局中人具有类型  $\theta_i \in \Theta_i$ , 策略  $s_i \in S_i$  及支付函数  $\mu_i$ , 类型上先验分布为  $p$  的不完全信息博弈中的纯策略贝叶斯均衡是一种“扩充”博弈的纳什均衡,这种扩充博弈中每个局中人  $i$  的纯策略空间是有  $\Theta_i$  到  $S_i$  的映射的集合  $S_i^{\Theta_i}$ .

豪尔绍尼转换的思想:在理论上,各类不完全信息在博弈分析中都可以转换为一种不完全信息情势,即局中人对其他局中人的支付函数的不完全了解.豪尔绍尼理论中,博弈的不完全信息表现为对博弈的基本的数学结构的了解不充分,在策略型博弈中,也就是对于3种组成部分,即局中人、策略和支付有着不完全了解.豪尔绍尼转换的主要思路是以类型概念构造对不完全信息的描述,在此基础上构造统一的概率模型来描述局中人在博弈中对不完全信息的处理,从而将不完全信息转换为不完全信息的完全信息博弈.

贝叶斯理论求解转换后的博弈局势,即首先按照概率分布  $P(\theta_1, \theta_2, \dots, \theta_n)$  随机抽取选择类型向量  $(\theta_1, \theta_2, \dots, \theta_n)$ , 每个局中人知道自己的实际类型,但不知道其他局中人的实际类型,每个局中人选择自己的策略,并且得到自己的支付.

## 2 基于贝叶斯均衡不完全信息博弈机器人对抗决策方法

在完全信息博弈中,全体参与人都知道博弈的规则,否则这一博弈就是一个不完全信息博弈.一个不完全信息博弈,经过豪尔绍尼转换就变成一个完全但不完美信息博弈,对这个转换形成的博弈就可以利用

完全信息博弈的相关理论与方法来进行分析.

### 2.1 不完全信息博弈模型建立

机器人的视觉系统可以感知到周围的场地环境,且假设感知得到的这些信息都是准确的,即能够保证局部信息完全可知.在实际的比赛中,机器人不仅仅要根据周围的场地信息,还要依据敌方机器人的决策动作信息,进行综合的分析推理后,从而得到自己的决策动作信息.

在某场足球机器人比赛中,假设视觉系统传递的场地环境信息是可靠的,且参与者为我方和敌方2队,即局中人  $N = \{1, 2\}$ , 并将某个局中人  $i$  之外的其他局中人称为“ $i$  的对手”,记为  $-i$ . 机器人比赛决策博弈局势:局中人1要根据自己对场上形势和局中人2的判断,决定自己是否踢球;局中人2要根据自己对场上形势和局中人1的判断,决定自己是否控球;局中人2的局势为在“进攻”状态下控球与不控球和在“防守”状态下控球与不控球;局中人1决定是否踢球,而同时局中人2决定是否控球.这里只分析局中人1该做出何种决策,对其分析得出如表1所示的不完全信息博弈局势模型.

表1 机器人比赛博弈模型

Table 1 The robot competition game model

局中人1		进攻支付矩阵		防守支付矩阵	
		踢球	不踢球	踢球	不踢球
局中人2	控球/%	(30 20)	(40 10)	(40 20)	(90 10)
	不控球/%	(70 20)	(50 50)	(70 20)	(70 5)

上述不完全信息引起了局中人1决策的困难,即是否进行踢球或者不踢球.利用划线法求得上述比赛博弈模型的纳什均衡点:如果局中人2在进攻状态下,那么惟一的纯策略纳什均衡是,局中人1不踢球,局中人2不控球,即(50% 50%);如果局中人2在防守状态下,那么惟一的纯策略纳什均衡是,局中人1踢球,局中人2不控球,即(70% 20%).所以,当机器人进行比赛时,2个局中人当时所处的状态对比赛的结果起着一定的调控作用,它们分别对应着不同的均衡结局.因此,局中人需要对自己所不能确知的任何信息做出主观判断,并在此基础上决定自己的应该采取的行为以避免盲目“猜测”而导致比赛的失败.在上述的不完全信息博弈中,并非所有人都知道同样的信息,除了均知道的公共信息外,局中人各自具有自己的私有信息.于是在进行策略选择的时候,局中人需要对其他局中人的私有信息进行判断,并依此进行相应的策略选择.

在不完全信息博弈模型中,局中人知道其他局

中人的实际类型为若干可能类型中的一种,但不知道究竟是哪一种,只能在猜测的基础上选择自己的策略。

为了描述这种主观判断,贝叶斯博弈理论利用贝叶斯理性原则来描述这种不确定情形下人们的理性行为。贝叶斯博弈理论假设局中人的类型  $\{\theta_i\}_{i=1}^n$  来自于一种类型上的联合概率分布  $p(\theta_1, \theta_2, \dots, \theta_n)$ , 这种联合概率分布是局中人的共有信息,且比赛双方对这种共有信息都是已知的。这种联合概率分布是已知的,然后各个局中人在此基础上形成对其他局中人实际类型的概率判断,即局中人  $i$  在知道自己的实际类型为  $\theta_i$  的情况下对对手类型形成的条件概率分布  $p(\theta_{-i}|\theta_i)$ , 按贝叶斯推断即为  $p(\theta_{-i}|\theta_i)/p(\theta_i)$ 。假设局中人 1 (我方) 和局中人 2 (对方) 均具有 2 种可能的类型:“防守”和“进攻”。现在引入上述不完全信息博弈的贝叶斯均衡概率模型,其中概率分布如表 2 所示。

表 2 联合概率分布

Table 2 Joint probability distribution %

类型	进攻	防守
进攻	30	20
防守	10	40

根据这一联合概率分布,就可以知道每个局中人在不同情况下对其他局中人实际类型的概率推断。如果局中人 1 为“进攻”类型,那么它对局中人 2 的类型判断依据贝叶斯推断原则有局中人 2 为“进攻”类型的概率为

$$\frac{p(\theta_i, \theta_{-i})}{p(\theta_i)} = \frac{30\%}{30\% + 20\%} = 60\%.$$

同理,局中人 2 为“防守”类型的概率为 20% / (30% + 20%) = 40%;而当局中人 1 为“防守”类型

时,对局中人 2 的主观判断为局中人 2 为“进攻”的概率为 20%,为“防守”的概率为 80%。类似的也可以有局中人 2 对局中人 1 类型的判断。在这种概率模型中,由于联合概率分布为各个局中人均知道的公有信息,这种形成主观判断的机制也为所以局中人了解,所以局中人知道其他局中人的主观判断的方式以及相应的结果。

2.2 不完全信息博弈模型的转换

在概率模型的基础上,就可以通过豪尔绍尼转换将不完全信息博弈局势转化为不完美信息博弈局势。其中引入“自然”这一虚拟局中人,所以局中人的实际类型均来自于“自然”根据类型上的联合概率分布进行的一种初始抽彩,局中人根据这种抽彩决定自己对其他局中人类型的主观判断,然后局中人进行实际博弈。豪尔绍尼转换将不完全信息博弈转化为不完美信息的完全信息博弈后,就可以利用完全信息博弈的处理方法。上述的足球机器人比赛不完全信息博弈的豪尔绍尼转换如图 1 所示。在此博弈树中,有一个初始节点先于其他节点,记为“O”,表示博弈的起点。为了博弈树的清晰起见,通过引入虚拟自然局中人的“行动”建立惟一的初始节点;并且,除了初始节点之外,博弈树中的每一个节点有且仅有一个直接前列节点;没有后序节点的节点被称为终止节点,表示博弈的可能结局,从而达到博弈的一种结局。非终止节点的节点被称为决策节点,非初始节点的决策节点用黑圆点表示。

豪尔绍尼转换将不完全信息博弈转换为了不完美的信息扩展博弈,针对图 1 所示的不完美信息博弈,利用贝叶斯均衡可以将其转化为策略型博弈,即求得不完全信息博弈树的解。

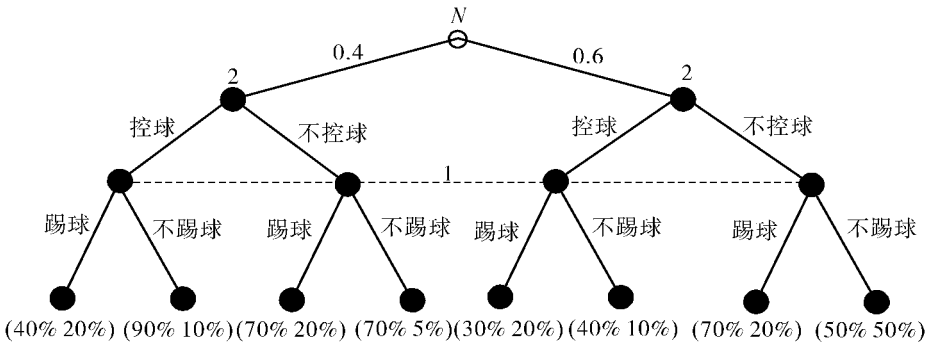


图 1 经豪尔绍尼转换后的机器人博弈树

Fig.1 Robot game tree after Harsanyi conversion

2.3 机器人博弈树求解

贝叶斯均衡是一种与类型有关的策略组合

$(s_1^*(\theta_1), s_2^*(\theta_2), \dots, s_n^*(\theta_n))$ , 其中每个局中人  $i$  在给定自己类型  $\theta_i$  和其他局中人策略  $s_{-i}^*(\theta_{-i})$  的

情况下最大化自己的期望效用函数. 即

$$s_i^*(\theta_i) \in \arg \max_{s_i} \sum_{\theta_{-i}} p(\theta_{-i} | \theta_i) \cdot u_i(s_i, s_{-i}^*(\theta_{-i}), (\theta_i, \theta_{-i})).$$

式中:  $u_i(\sigma) = \sum_{s \in S} [\prod_{j=1}^n \sigma_j(s_j)] u_i(s)$  为与类型组合有关的局中人支付函数;  $p(\theta_{-i} | \theta_i)$  为局中人对其其他局中人类型的概率判断. 利用上面的贝叶斯均衡可以得到豪尔绍尼转换后的不完全信息博弈树的解, 进而得到与不完全信息博弈局势等价的扩充博弈局势如表3所示.

表3 对应的扩充博弈局势

Table 3 The corresponding expansion game situation

局中人2	局中人1	
	踢球	不踢球
(控球 控球)	(0.34 <u>0.2</u> )	(0.6 0.1)
(控球 不控球)	( <u>0.58</u> 0.2)	(0.66 <u>0.34</u> )
(不控球 控球)	(0.46 <u>0.2</u> )	( <u>0.7</u> <u>0.2</u> )
(不控球 不控球)	(0.4 <u>0.4</u> )	(0.58 0.32)

如表3利用划线法可以求得经过豪尔绍尼转换并进行贝叶斯推理后的博弈局势的纯策略纳什均衡点为((不控球 控球)不踢球), 即(0.7 0.2). 可以看出, 这一贝叶斯博弈的纯策略贝叶斯均衡有一个, 即无论局中人2是处于“进攻”状态还是“防守”状态, 局中人1都选择不踢球, 这样的决策保证了局中人1不失利, 也就是说选择了最保守的决策方式.

这种建模和求解方式给出了局中人1在2种类型下的策略选择, 括号内前者是局中人1在实际类型为“进攻”时选择的策略, 后者为局中人1为“防守”时选择的策略. 通过这种概率机制的引入, 构造了局中人对其他局中人主观判断的不确定性, 形成了对不完全信息的完整描述.

## 2.4 2种决策方式比较

对于不完全信息给博弈局势带来的决策困难, 选择了豪尔绍尼转换和贝叶斯相融合的博弈处理方法, 这样的处理简化了博弈决策的结果. 直接对博弈状况进行分析的结果有2种博弈均衡, 就是对于局中人2不同的状况, 局中人1需进行不同的决策, 即踢球和不踢球. 而经过豪尔绍尼转换和贝叶斯推理后, 局中人1可以不用考虑局中人2的状况, 选择最保守的决策命令, 即选择不踢球.

## 3 仿真验证

博弈局势的建模和求解是一个复杂的过程, 即使是确定的双方, 在不同的比赛局势中的攻守策略也是不同的, 豪尔绍尼转换和贝叶斯推理相融合的方法为解决机器人比赛中局势的动态变化引起的信

息不确定性问题提供了一种解决方案.

本文利用 Matlab 软件对机器人的决策优化问题进行仿真. 基本的仿真环境设计: 在二维平面中, 系统运行的工作环境中的不同位置点表示不同的决策命令, 这些决策命令点是已知的, 用“\*”表示; “-”表示博弈选择的结果, 直线穿越的决策点就是机器人要选择的决策点, 即这些决策点都是对比赛的结果有利的决策点. 任务要求机器人根据采集到的周围环境及敌方的信息推理分析后, 从所有的决策命令集中选择出最有利的决策, 并传送给机器人执行命令. 图2是机器人某次比赛的博弈局势, 直接求其博弈局势的纳什均衡点, 选择出了4个决策点; 图3是对同样的博弈局势进行豪尔绍尼转换和贝叶斯相融合的不完全信息博弈转换推理后, 所得到的2个决策点.

比较图2和图3可知, 豪尔绍尼转换和贝叶斯相融合的不完全信息博弈算法在机器人比赛中的应用比传统直接求解方法减小了2个决策点, 提高了决策的精度. 如果在决策点更多时, 这种方法可以大大的缩小有利决策点的决策范围, 从而进一步提高在比赛中获胜的概率.

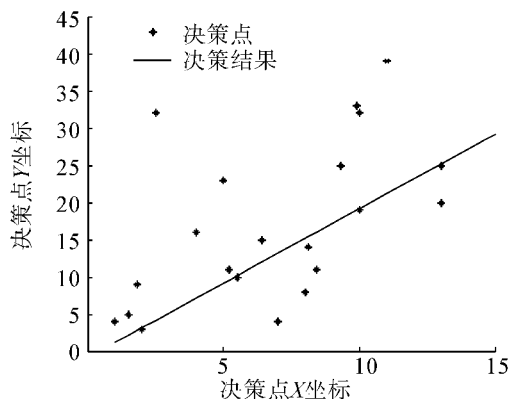


图2 未经过转换推理的博弈仿真

Fig.2 Game simulation diagram before convert reasoning

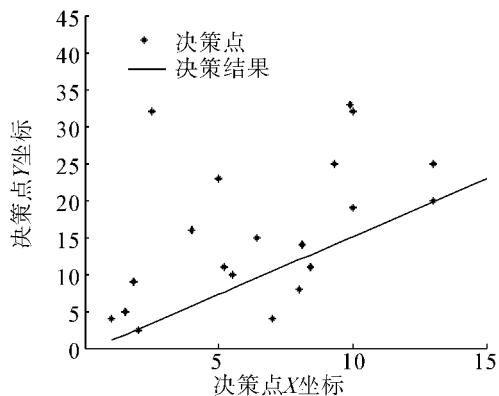


图3 转换推理前后仿真图比较

Fig.3 Game simulation diagram before convert reasoning

## 4 结束语

豪尔绍尼转换和贝叶斯博弈相融合的不完全信息博弈算法可以有效地拓宽有利决策点的选择范围,其实际仿真效果与理论相符合.存在的问题是在决策过程中的豪尔绍尼转换是否会使机器人的决策变得复杂,如何在程序中将这种转换程序化、简单化是需要进一步进行研究的.不完全信息动态博弈理论引入到机器人足球决策系统当中,具有很好的理论意义和实际应用价值.

## 参考文献:

- [1] 王志红. Robocup 中型组足球机器人决策系统的研究 [D]. 济南:山东大学, 2007: 4.  
WANG Zhihong. The research on the decision-making system of Robocup medium group soccer robot [D]. Ji'nan Shandong University, 2007: 4.
- [2] 施卫强. 基于强化学习的足球机器人决策系统设计 [D]. 长沙:中南大学, 2007: 5.  
SHI Weiqiang. The design on the decision-making system of robot soccer based on Q learning [D]. Changsha: Central South University, 2007: 5.
- [3] 徐心和, 邓志立, 王骄, 等. 机器博弈研究面临的各种挑战 [J]. 智能系统学报, 2008, 3(4): 288-296.  
XU Xinhe, DENG Zhili, WANG Jiao, et al. The challenges of machine game [J]. CAAI Transactions on Intelligent Systems, 2008, 3(4): 288-296.
- [4] BASILICO N, CATTI N, AMIGONI F, AMIGONI F. Leader-follower strategies for robotic patrolling in environments with arbitrary topologies [C]//Proc AAMAS. [S. l.], 2008: 57-64.
- [5] BEETZ M, BUCK S, HANEK R. The AGILO autonomous robot soccer team: computational principles, experiences, and perspectives [C]//Proc AAMAS, [S. l.], 2009: 805-813.
- [6] 柳长安, 王静, 刘春阳. 基于合作4人博弈的足球机器人协作防守模型研究 [J]. 系统仿真学报, 2009, 21(1): 132-134.
- LIU Changan, WANG Jing, LIU Chunyang. The research on the defensive support of soccer robot based on four cooperative game [J]. Journal of System Simulation, 2009, 21(1): 132-134.
- [7] 王云, 韩伟. 对称协调博弈问题的多智能体强化学习 [J]. 计算机工程与应用, 2008, 44(36): 230-234.  
WANG Yun, HAN Wei. A multi-agent reinforcement learning on symmetrical and reconciling game [J]. Computer Engineering and Applications, 2008, 44(36): 230-234.
- [8] 宋明鑫. 计算机技术在“囚徒困境”博弈中的应用 [D]. 天津:天津大学, 2006: 23-25.  
SONG Mingxin. The application of computer technology in prisoner's dilemma game [D]. Tianjin: Tianjin University, 2006: 23-25.
- [9] 黄涛. 博弈论教程 [M]. 北京:首都经济贸易大学出版社, 2004: 45-49.
- [10] 李光久. 博弈论基础 [M]. 镇江:首都经济贸易大学出版社, 2008: 98-100.

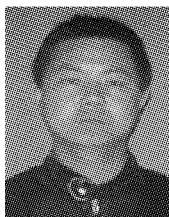
### 作者简介:



史晓茹,女,1987 生,硕士研究生,主要研究方向为控制理论与控制工程.



侯媛彬,女,1953 生,教授,博士生导师,博士,主要研究方向为控制理论、系统辨识、人工智能与模式识别等.



张涛,男,1985 生,硕士研究生,主要研究方向为控制理论与控制工程.