高技术通讯 2013 年 第23 卷 第9 期:965 ~972

doi:10.3772/j.issn.1002-0470.2013.09.014

机器博弈风险分析及其估算方法的研究①

张加佳②　王　轩

(哈尔滨工业大学深圳研究生院　深圳 518055)

摘　要　提出了风险损失的概念,讨论了机器博弈问题中可能导致风险损失的几种因素。针对完备信息和非完备信息机器博弈问题的风险损失情况进行分析,提出了一个在博弈过程中估算机器博弈问题风险损失的方法。通过对完备信息博弈问题和非完备信息多人博弈问题的实验,论证了所提出的观点,并验证了提出的风险损失估算方法的有效性。最后提出了一种基于风险损失的博弈策略选择方法。

关键字　多重均衡,风险占优均衡,非完备信息博弈,风险损失

0　引　言

在机器博弈领域,完备信息二人机器博弈问题研究已取得很多成功,例如代表性的国际象棋博弈(深蓝,IBM)、西洋跳棋等[1]。在我国,该领域的主要工作集中在对中国象棋博弈问题的研究[2,3],其顶级的智能体博弈水平已经可以和职业象棋棋手相匹敌。这类问题研究的基本假设之一是博弈者采用相同的估值函数,博弈问题存在唯一的最优 Nash 均衡。但是当机器博弈扩展到更为复杂的非完备信息博弈问题时,经典的完备信息博弈算法并不能给出令人满意的表现。建立对手模型,采用多重均衡博弈理论及方法分析和这类复杂的机器博弈问题已经成为近年来机器博弈领域的研究热点[4]。很多学者对博弈多重均衡的直觉判断和分析的有关问题提出了不同观点,以 Nash 均衡作为博弈者的合理必然结果的策略选择模式并不总能成立。博弈论大师 Harsanyi 和 Selten 于 1988 年提出风险优势(risk dominance)的概念,以作为收益优势(payoff dominance)的补充[5]。库珀(Cooper,1999)的研究认为风险占优均衡实际上提供了正式处理博弈信息不确定性的方法[6]。对博弈问题中风险因素的研究,已成为博弈均衡选择理论的一部分。博弈中均衡的多重性在很多情况下来源于博弈者面临的博弈信息的不确定性,风险占优均衡已经成为多重均衡策略选择的一个重要方法。随后,围绕多重均衡策略选择问题的应用研究在不断深入。相关研究主要集中在基于 max-n 算法的策略选择模型的建立方法,其中以 Nathan 的对手模型研究为代表[7]。但是,系统地针对机器博弈问题中风险因素的分析及估算方法方面的研究并不多见,笔者初步建立了一套简单的通过抽样方法评估策略风险的模型[8]。2008 年,Zinkevich 基于损失最小化(regret minimization)思想提出了 CFR(counterfactual regret)概念及相关算法[9],Nick Abou Risk 等人进一步将 CFR 算法扩展到多人博弈的问题当中[10]。CFR 算法的本质是机器学习的方法,通过不断训练,无限逼近最优策略,这就限定了该算法目前只能处理 10^{12} 世界数规模的博弈问题。

本文对机器博弈问题中的风险因素展开了研究,通过提出风险损失的概念,对不同条件下的机器博弈问题的风险因素性质进行了分析。在此基础上,提出了一套系统地分析和评估机器博弈策略选择中的风险损失的方法,并通过对完备信息博弈和非完备信息博弈问题进行分析,证明了该方法的合理性。

1　机器博弈问题中的风险分析

风险优势和收益优势是 Nash 均衡博弈中的两个基本概念。Harsanyi 和 Selten 的研究提出了这样

① 国家重大科技专项(2011ZX03002-004-01)资助项目。

② 男,1984 年生,博士生;研究方向:人工智能,机器博弈;联系人,E-mail:ashrumscg@163.com

(收稿日期:2012-12-05)

的观点：当一个 Nash 均衡相对于其他均衡具有 Pareto 优势时，收益占优策略将会被采用。在这种情况下，所有的博弈者都会遵循收益占优策略，因为它提供了优于其他均衡的收益预期。但是当博弈者所处的博弈环境被赋予较大的不确定性因素时，风险占优策略将会成为更为合理的 Nash 均衡。在 Straub[11] 的研究中，通过对人类实验者的在非完备信息条件下博弈实验的测试，同样验证了这样的结论。

上述研究中，风险占优策略决定 Nash 均衡的实验环境和机器博弈领域的非完备信息博弈问题非常相似。与完备信息博弈问题相比较，非完备信息博弈问题的博弈者面对的博弈环境信息往往是不完全或者不可信的，经典的解决完备信息博弈问题的方法和理论在这种情况下无法取得好的表现。其主要原因之一就是经典的完备信息机器博弈算法往往是基于收益占优策略指导。由此，结合非完备信息机器博弈问题的特点，引入风险占优策略思想，建立一个有效地发现和评估机器博弈问题策略风险的模型是有意义的。

为形式化研究机器博弈问题风险模型，本文首先引入风险损失（博弈者预期收益与实际收益的差值）概念。

博弈者对自身策略产生的预期收益的计算过程如下：首先，博弈者通过对自身策略及对手策略的计算，产生自身和对手顺序执行的可能策略序列集合，该集合中序列的长度取决于当前算法的搜索深度。然后，根据该策略序列计算其会产生的未来局面。最后，对这些局面估值，并据此反向得出当前的最优策略序列，对最优策略序列产生局面的收益估值即为预期收益。

图 1 描述了一个简单的博弈树搜索过程。博弈者 A 的策略集为 $\{A_1, A_2\}$，博弈者 B 的策略集为 $\{B_1, B_2\}$，假设搜索深度为 2，那么博弈者 A 可以计算当前情况下的策略序列集合为 $\{A_1, B_1\}$，$\{A_1, B_2\}$，$\{A_2, B_1\}$，$\{A_2, B_2\}$。估值函数对以上 4 个策略序列产生的局面，即图 1 中的叶节点进行估值，选择双方最优策略序列（图中以 $\{A_1, B_1\}$ 为例）。此时，局面估值 V_{A1} 即为博弈者 A 选择策略 A_1 时的预期收益。

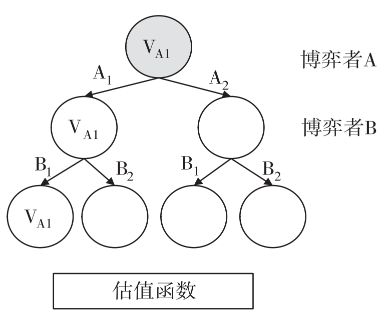

图 1　机器博弈问题策略选择过程示意

在非完备信息机器博弈问题中，博弈者面临的博弈环境是不可见或不可信的。在上述过程执行之前，通常采用抽样方式将非完备信息博弈条件转化为完备信息博弈条件，这一过程称为完备信息蒙特卡洛（perfect information Monte Carlo，PIMC）抽样。该方法在非完备信息博弈领域已得到广泛应用[12]。

如图 2 所示，集合 W 代表了非完备信息博弈环境 I 的所有可能的情况的集合，W 中的每个元素 w_i 都代表了 I 的一个可能的完备信息状态，I 的真实状态是 W 中的某一个 w_i。这里引入世界的概念：一个世界是非完备信息博弈的一个可能状态。

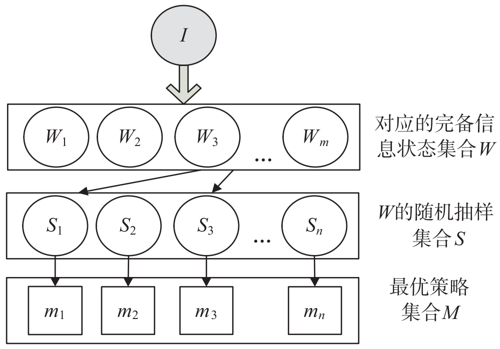

图 2　非完备信息博弈过程

W 是当前博弈状态的世界集，S 是 W 的抽样集，$S \subseteq W$。PIMC 方法的基本过程是，采用随机方法抽样出 W 的子集 S，对其中的每个完备信息世界 s_i 进行计算，统计分析每个 s_i 的最优解 m_i，最后在 M 中选择最终的最优策略序列。其中，从 s_i 中选择 m_i 的过程与图 1 所示过程基本相同。

在上述的策略选择过程中，至少有两个步骤存在不确定性。（1）博弈者 A 预计策略序列 $\{A_1, B_1\}$ 为当前双方的最优策略序列，也就是说，当博弈者 A 选择策略 A_1 时，预计博弈者 B 会选择策略 B_1，而实际上博弈者 B 可能会采取其他的策略导致局面演化到其他方向。（2）对最终局面的估值，博弈者 A 根据自己的估值函数认为策略的预期收益为 V_{A_1}，然而在非完备信息条件下，图 1 中顶点代表的博弈状态并非是真实状态，而是抽样出的一个可能世界，由

此对经过$\{A_1,B_1\}$衍生出的叶节点与真实状态衍生出的叶节点的估值很可能产生偏差。以上两点是博弈者 A 对于策略 A_1 的预期收益和实际收益产生偏差的原因,也是本文研究的风险损失的来源。

在本文中,将机器博弈问题策略选择算法的不确定性归结为两类风险损失。

I 型风险损失及其计算方法:

由估值函数的对世界估值的不准确性造成的风险损失称为 I 型风险损失。假设世界 w 的最优策略序列为 m,则此时 m 的 I 型风险损失计算方法如下:

$$L_{wI} = E_w^m - E_I^m \tag{1}$$

式中,E_w^m 代表估值函数对世界 w 下采取策略序列 m 的收益估值,E_I^m 代表真实世界采取策略序列 m 时的收益估值。

II 型风险损失及计算方法:

由对对手最优策略判断的不准确性造成的风险损失称为 II 型风险损失,策略序列 m 的 II 型风险损失计算方法如下:

$$L_{mII} = E_I^m - E_I^{m'} \tag{2}$$

式中,E_I^m 是估值函数对真实世界 I 采取策略序列 m 的收益估值。$E_I^{m'}$ 是真实世界 I 下博弈双方的实际策略序列 m'的收益估值。

图 3 再次展示了 I、II 型风险损失的区别。估值函数对世界 w 和真实世界 I 经过策略序列 m 的预期收益的估值差为 I 型风险损失,图中为 L_{wI},真实世界 I 中,策略序列 m 和实际策略序列 m'的预期收益差为 II 型风险损失,图中为 L_{mII}。由此,定义世界 w 下采用策略序列 m 的风险损失为

$$L_{wm} = L_{wI} + L_{mII} \tag{3}$$

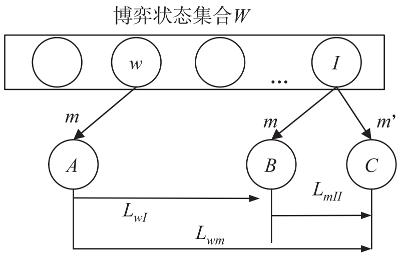

图 3　I、II 型风险损失示意图

式(3)给出了策略序列 m 在世界 w 的风险损失计算方法。当博弈者面对所有可能的世界集合 $W\{w_1,w_2\cdots w_n\}$时,定义策略序列 m 的综合风险损失的计算方法如下:

$$L_{Wm} = \frac{1}{n}\sqrt{\sum_{i=1}^{n} L_{w_i m}^2} = \frac{1}{n}\sqrt{\sum_{i=1}^{n}(L_{w_i I} + L_{mII})^2} \tag{4}$$

2　风险损失出现概率

本文提出两个风险损失的命题:

命题 1:完备信息条件下,若博弈问题中的智能体采用相同的估值函数,则它们面临的风险损失的概率为 0。

证明:在完备信息博弈条件下,只存在唯一一个已知的可能世界。在式(1)中,可以理解为 $I=w$,因此 $P\{L_{wI}=0\} = P\{E_w^m - E_I^m = 0\} = P\{E_I^m - E_I^m = 0\} = 1$。

如果此时博弈问题中的博弈者遵循同样的估值函数,则由此计算出的各方的策略序列是一致的。也就是说,博弈者可以正确计算出最优策略序列,在式(2)中,可以理解为 $w=w'$,由 II 型风险损失的定义可知 $P\{L_{mII}=0\} = P\{E_I^m - E_I^{m'} = 0\} = P\{E_I^{m'} - E_I^{m'} = 0\} = 1$。

由于此时只有唯一一个世界,即可以理解为式(4)中的 $n=1$。此时策略 m 的综合风险损失可以计算如下:$L_{Wm} = \frac{1}{n}\sqrt{\sum_{i=1}^{n} L_{w_i m}^2} = L_{wm} = (L_{wI} + L_{mII}) = 0$。证毕。

命题 2:在机器博弈问题中,若某一局面下世界集 W 经过策略序列 m 产生的节点有 n 个,估值函数对这些节点有 v 个可能的不同估值,则博弈者面临风险损失的概率满足:

$$P\{L_{Wm} \neq 0\} \geqslant 1 - \left(\frac{n-v+1}{n}\right)^n \tag{5}$$

证明:在完备信息条件下,同一状态只有一个可能的世界,经过策略序列 m 产生的节点也是唯一的,故 $n=v=1$,由命题 1:完备信息机器博弈问题中,博弈者面临风险损失的概率为 0,此时式(5)左右都为 0,等式成立。非完备信息条件下,如图 3 所示,风险损失为 0 的情况是节点 A 和 C 估值相同。在可能的节点为 n 个,不同估值为 v 个时,可以算出 A 和 C 估值相同的最大概率为$(n-v+1)/n$,即 $P\{L_{wm}=0\} \leqslant \frac{n-v+1}{n}$,所以 $P\{L_{Wm} \neq 0\} = P\{\frac{1}{n}\sqrt{\sum_{i=1}^{n} L_{w_i m}^2} \neq 0\} = 1 - P\{\sum_{i=1}^{n} L_{w_i m}^2 = 0\} = 1 - \prod_{i=1}^{n}$

$P\{L_{w_im}=0\}\geqslant 1-(\frac{n-v+1}{n})^n$ 即为式(5)。证毕。

在非完备信息机器博弈中,由于博弈信息的不确定性,任一局面的世界集 W 中的世界数 n 是非常大的。由命题 1、2,可以看到,博弈者在进行策略选择时几乎必然会面临风险损失。研究基于风险占优思想的博弈均衡选择策略,在非完备信息机器博弈问题的研究中具有重要的意义。

3　非完备信息条件下风险损失计算及风险占优策略选择方法

在非完备信息条件下的机器博弈算法策略选择过程中,上述计算方法会遇到两个问题:(1)通常的非完备信息问题的世界数非常巨大,使得系统无法一一计算世界集 W 中的所有元素,式(4)中的求和运算是无法真正实现的。(2)参与博弈问题的博弈者无法预先获知如图 3 中的真实世界 I 和实际策略序列 m',因此,式(1)、(2)中的 E_I^m 和 $E_I^{m'}$ 无法被精确计算。因此,本文给出一个近似化计算风险损失的方法,其基本思想是计算抽样集 S 中的预计收益的均值来代替世界集 W 中 I 的真实收益。

假设博弈者对当前状态的世界集为 W,元素数为 n,W 的抽样集为 S,元素数为 t,M 为 W 的所有合法策略序列集合,元素数为 k。首先给出此时的平均收益计算方法。

定义: $\overline{E}_s$ 为抽样集 S 的平均收益。计算方法为

$$\overline{E}_s=\frac{1}{tk}\sum_{i=1}^{t}\sum_{j=1}^{k}E_i^j(i\in S,j\in M)\qquad(6)$$

基于式(6),可得策略序列 m 的综合风险损失近似化计算方法公式如下:

$$\begin{aligned}L_{Wm}&=\frac{1}{n}\sqrt{\sum_{i=1}^{n}L_{w_im}{}^2}=\sqrt{\sum_{i=1}^{n}(L_{w_iI}+L_{mII})^2}\\&=\frac{1}{n}\sqrt{\sum_{i=1}^{n}(E_{w_i}^m-E_I^m+E_I^m-E_I^{m'})^2}\\&=\frac{1}{n}\sqrt{\sum_{i=1}^{n}(E_{w_i}^m-E_I^{m'})^2}\\&\approx\frac{1}{t}\sqrt{\sum_{i=1}^{t}(E_{w_i}^m-\overline{E}_s)^2}\qquad(w_i\in S)\qquad(7)\end{aligned}$$

式中,约等号所连接处即为使用 $\overline{E}_s$ 以及抽样集 S 进行近似计算的过程。

基于以上方法,博弈者进行策略选择过程中的每个候选策略的风险损失都可以被计算。经典的策略选择遵循的是收益占优原则,即在如图 2 所示的候选策略集 M 中,选择预期收益最大的候选策略作为最优策略。本文尝试同时考虑候选策略的预期收益和风险损失,提出一种基于风险占优原则的策略选择方法,即使用风险占优策略作为博弈者的最优策略。

Harsanyi 和 Selten 认为风险占优原则下的策略选择,即风险占优策略的选择是一个复杂的回溯过程,Selten 在研究[13]中提出了一个双人博弈中可行的计算风险占优策略的方法。本文以此为基础,提出了非完备信息博弈条件下的风险占优策略选择方法。

表 1 举出了两个非完备信息博弈问题中的典型场景,博弈者 1 面临着 A 和 B 两个候选策略,E_m 代表博弈者对于策略 m 的预期收益。L_m 代表策略 m 的风险损失。

表 1　两个典型的非完备信息博弈场景

场景 1:博弈者 1	E_m	L_m	E_m-L_m
A	8	4	4
B	6	3	3
场景 2:博弈者 1	E_m	L_m	E_m-L_m
A	8	8	0
B	6	2	4

直观地观察表 1 中的两种情况可以发现,当风险损失较小时(场景 1 所示情况),风险占优策略选择结果应该与收益占优是一致的,即选择 A 作为博弈者 1 的最优策略。但是当风险损失较大时(场景 2 所示情况),风险占优策略会做出和收益占优不同的选择。本文采用以下方法量化两个策略基于风险占优的优劣判断。

假设有策略 A,B。E_A 和 E_B 分别代表博弈者对于策略 A、B 的预期收益。L_A 和 L_B 代表策略 A 和 B 的风险损失。则策略 A、B 的优劣判断规则如下:

(1)若策略 A、B 满足 $u_A-L_A>u_B$,则 A 优于 B,反之,若满足 $u_B-L_B>u_A$,则 B 优于 A。

(2)否则,用下式:

$$R=\log[\frac{E_A-(E_B-L_B)}{E_B-(E_A-L_A)}]\qquad(8)$$

如果 $R>0$,则 A 优于 B,若 $R<0$,则 B 优于 A,若 $R=0$,则 AB 等优,系统可做随机选择。

由以上方法,可对当前博弈者的所有候选策略

进行排序,排序最优的策略作为当前的风险占优策略,也即是博弈者的最优策略。

4 实验分析

本节将针对分别由中国象棋和四国军棋代表的完备信息机器博弈问题和非完备信息机器博弈问题展开实验,分析其中的各种风险因素。中国象棋和四国军棋是在中国普及率较高的两种棋类,其吃子和移动规则由于篇幅限制本文不详细介绍,可参见文献[14]。

实验1:对完备信息和非完备信息机器博弈过程中的风险损失情况进行比对。表2至表5列出了中国象棋实验系统和四国军棋实验系统分别与自身及人类玩家对局的实验结果,实验对每10局的结果进行统计,表中数据是10局数据的平均值。为了方便对比,本文的实验调整了两个实验平台的估值函数,将其估值范围映射到0~100的范围内。表中p代表仅有I型风险损失出现的策略在全局中的比率,p'代表I、II型风险损失同时出现的比率。

表2 中国象棋100局对战实验数据(对战机器)

对局	L	p	L_I	p'	L_I+L_{II}
1—10	0	0	0	0	0
11—20	0	0	0	0	0
21—30	0	0	0	0	0
31—40	0	0	0	0	0
41—50	0	0	0	0	0

表3 中国象棋100局对战实验数据(对战人类)

对局	L	p	L_I	p'	L_I+L_{II}
1—10	1.04	0	0	0.08	13
11—20	0.66	0	0	0.06	11
21—30	1.44	0	0	0.12	12
31—40	0.55	0	0	0.05	11
41—50	0.18	0	0	0.02	9

表4 四国军棋100局对战实验数据(对战机器)

对局	L	p	L_I	p'	L_I+L_{II}
1—10	29	0.12	49	0.88	26
11—20	33	0.08	40	0.92	32
21—30	34	0.14	23	0.86	36
31—40	42	0.08	65	0.92	40
41—50	39	0.12	35	0.88	40
总计	35.4	0.108	42.4	0.892	34.8

表5 四国军棋100局对战实验数据(对战人类)

对局	L	p	L_I	p'	L_I+L_{II}
1—10	35	0.10	21	0.90	36
11—20	15	0.10	23	0.90	14
21—30	46	0.06	36	0.94	47
31—40	26	0.16	30	0.84	24
41—50	27	0.08	21	0.92	28
总计	29.8	0.10	26.3	0.90	29.8

从上表中我们可以看出,完备信息博弈时风险损失几乎为0,与计算机的对弈实验结果与命题1相同,只有在与人类的对局中出现风险损失,风险损失主要体现为II类风险损失,即对对手策略的不正确预期,这主要是来自于象棋博弈在布局阶段的多变性,不同的博弈者有着自身的布局策略和风格,针对该现象的研究即为对手建模[8]。在这种情况下,对手采用与计算机博弈者不同的估值函数,因此实验结果与命题1并不相悖。在非完备信息博弈的四国军棋中,初始信息空间大概为3.58×10^{58}[15],不同局面间的估值几乎互不相同,如命题2中定义的$w\approx v\approx3.58\times10^{58}$。由命题2可知,该博弈问题的策略风险概率接近为1。实现结果显示:一方面每一步都出现了风险损失,与命题2的结论相符;另一方面,风险损失也对局面产生了较为重要的影响。从表3、表4中可以看到,在100局试验中,风险损失的平均值在总分100的情况下,分别达到了35.4和29.8,风险损失对博弈结果产生了重要的影响。

实验2:对上一节提出的风险损失估算方法进行验证。表6—表8列出了随机抽取的3次对局中风险损失的估算值与真实值,图4—图6给出了随着走步,估算损失和真实损失的曲线图。可以看出,风险估值曲线与对局面风险曲线趋势基本相符,同时对主要波动点的估值进行了放大。由此,式(7)所示的风险损失估算方法是基本可行和有效的。

表 6 四国军棋对局 1 中风险损失估值与真实值对比

走步	估值	实际	走步	估值	实际
1	10	22	24	41	27
3	47	28	27	38	22
6	42	26	30	43	22
9	38	24	33	19	17
12	49	25	36	32	20
15	19	29	39	21	20
18	32	20	42	0	18
21	25	24	45	12	16

表 7 四国军棋对局 2 中风险损失估值与真实值对比

走步	估值	实际	走步	估值	实际
1	9	19	24	79	25
3	16	30	27	88	25
6	28	34	30	26	20
9	80	35	33	34	21
12	53	29	36	6	18
15	36	35	39	38	17
18	53	29	42	66	44
21	46	16	45	53	20

表 8 四国军棋对局 3 中风险损失估值与真实值对比

走步	估值	实际	走步	估值	实际
1	0	21	24	30	49
3	3	28	27	23	47
6	7	30	30	18	15
9	16	25	33	19	14
12	19	34	36	18	20
15	25	35	39	18	17
18	79	48	42	26	20
21	55	48	45	17	14

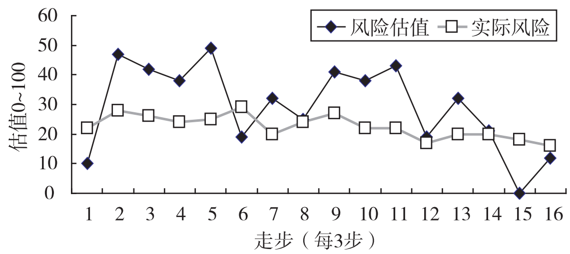

图 4 实验对局 1 风险估值结果

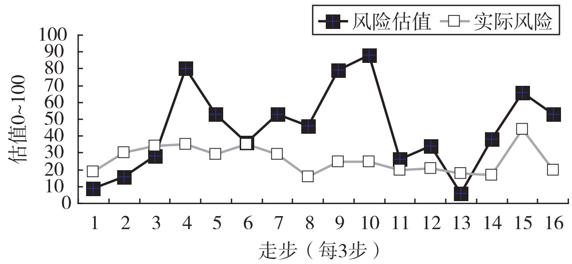

图 5 实验对局 2 风险估值结果

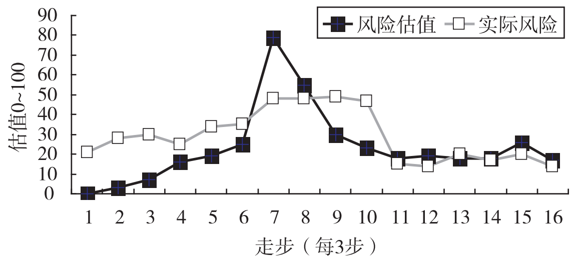

图 6 实验对局 3 风险估值结果

图 4—图 6 显示了 3 局对战中对于风险损失的估计值和真实风险损失的对比。图中的浅色曲线代表对局中风险损失的真实变化，深色曲线代表采用本文方法计算的风险损失的估计值的演化。可以看出，受到抽样次数和估值函数效率的制约，基于抽样集的风险损失估算方法还无法完全拟合真实风险损失，但是还是可以基本上反映风险损失的波动趋势。风险损失的估算精度可以通过扩大抽样集和优化估算方法来得到进一步提高，这也是本文下一步的研究目标之一。

实验 3：对博弈过程中的风险对博弈局势的影响进行进一步的分析，并对收益占优策略选择方法和本文提出的风险占优策略选择方法进行对比。实验选取了 2 个对局中造成局面下滑最大的三个策略点进行分析。在对局过程中，博弈者在这三个点的策略选择使其遭到了巨大损失。图 7 和图 8 列出了这些损失点的真实损失和估计损失，括号内前面的数字代表本文所提方法对风险损失的估值，后面的是数字风险损失的真实值。可以看出，在这些点上，风险损失的真实值和估算值都给出了正确的结果，远大于全局均值。尤其是图 8 中的第三个下滑点，系统通过对风险损失的计算，准确地预计了本次损失的发生。

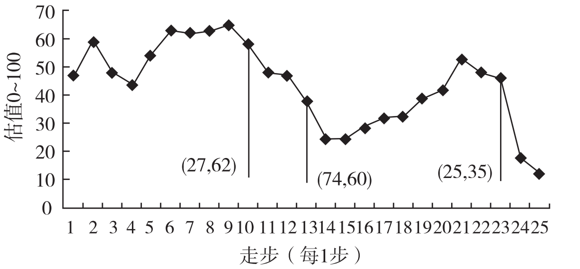

图7　实验对局4中3处最大损失点的风险估值结果

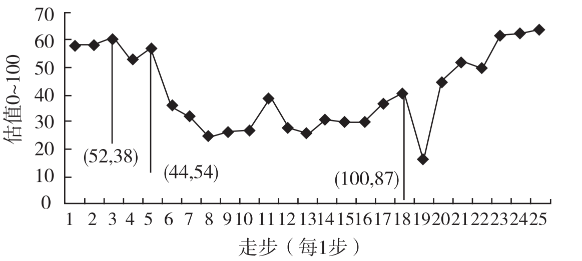

图8　实验对局5中3处最大损失点的风险估值结果

表9采用了本文提出的风险占优策略选择方法,对上述两个对局进行了进一步的分析。第1列是博弈者选择的策略在局中的3个关键点以及全局均值的实际收益。其策略是基于收益占优原则在实验中20个候选策略中选出的最优策略。第2列列出了如果这些最优策略采用风险占优方法分析,在候选策略中的排名。可以看到,收益占优策略不一定是风险占优策略,从全局的平均统计分析,收益占优策略一般排在风险占优策略排名的第2—5名左右。第3列列出了实验采用风险占优策略作为最优策略时对局中的实际收益并通过第4列做出了对比。可以看出,如果采用风险占优策略,两个实验对局的平均得分将提升3到5个百分点,尤其是在几个风险损失较大的关键点处,会有较为明显的提升。

实验3表明,本文提出的风险损失估算方法及风险占优策略选择方法是基本合理和有效的。如果博弈者在策略选择时能将风险因素考虑进来,那么这些点上的损失很有可能会被规避,也就是从另一个侧面说明了风险占优策略选择方法在机器博弈问题中的重要作用。

表9　收益占优策略选择与风险占优策略选择方法结果对比

实验对局4	策略得分	风险占优候选排名	风险占优策略得分	对比
1	49	7	62	+13
2	25	4	42	+17
3	19	9	28	+9
全局平均	43.88	2.97	48.85	+4.97

实验对局5	策略得分	风险占优候选排名	风险占优策略得分	对比
1	53	1	53	0
2	37	3	55	+18
3	17	6	23	+6
全局平均	41.84	3.10	45.59	+3.75

5　结　论

多重均衡博弈策略选择已经成为机器博弈领域研究的新方向,风险占优均衡在博弈论的研究中正是处理博弈信息不确定性的重要方法。因此,在机器博弈问题特别是非完备信息机器博弈问题的研究中,对风险因素的研究可以为博弈者的策略选择提供新的理论支持。本文从风险损失的概念入手,将机器博弈问题的风险损失分解为I类和II类不同来源的损失类型,尝试分析机器博弈问题中风险损失产生的原因及相关性质,提出了计算机器博弈问题决策过程中风险损失的方法,以及在非完备信息条件下博弈者对风险损失的估算方法,最后,提出了一个基本的风险占优策略选择方法。为验证提出的上述观点和方法,本文基于中国象棋和四国军棋两个平台,通过3组实验进行验证。基于实验结果,得出以下结论:(1)相对于完备信息博弈,非完备信息机器博弈问题中博弈者的策略选择面临着更为频繁、影响更大的风险损失。(2)本文提出的风险损失的计算方法、非完备信息条件下风险损失估算方法以及风险占优策略选择方法是有效和可行的。(3)风险损失对机器博弈问题的局势变化有着重要的影响,特别是在非完备信息机器博弈问题中,应该成为博弈者策略选择过程中的重要参考依据之一。实验也揭示了本文工作的下一步目标。首先是进一步完善和改进非完备信息条件下的风险损失估算方法,

使之可以更好地拟合风险损失的真实值。其次是与现有的机器博弈策略选择算法相结合，与不同机器博弈问题的特点相结合，进一步探索和完善更为合理和有效的风险占优策略选择方法。

参考文献

[1] Van den Herik H J, Uiterwijk J W, Van Rijswijck J. Games solved: how and in the future. *Artificial Intelligence*, 2001, 134(1): 277-311

[2] 徐心和，王骄. 中国象棋计算机博弈关键技术分析. 小型微型计算机系统，2006，27(6)：961-969

[3] 苏攀，王熙照，李艳. 基于不平衡学习的分类器博弈模型及其在中国象棋中的应用. 计算机研究与发展，2011，48(5)：841-847

[4] Sturtevant N R. Current challenges in multi-player game search. In: Proceedings of the 4th International Conference on Computers and Games, Ramat-Gan, Israel, 2004. 285-300

[5] Harsanyi J C, Selten R. A general theory of equilibrium selection in games. *Discrete Applied Mathematics*, 1990, 26(1), 126-127

[6] Cooper R. Forward induction in coordination games. *Economics Letters*, 1992, 40(2): 167—172

[7] Sturtevant N R, Bowling M H. Robust game play against unknown opponents. In: Proceedings of the 5th International Joint Conference on Autonomous Agents and Multiagent Systems, Hakodate, Japan, 2006. 713-719

[8] Wang X, Zhang J J, Xu X X, et al. Risk dominance strategy in imperfect information multi-player military chess game. In: Proceedings of International Swaps and Derivatives Association, Gaoxiong, China, 2008. 596-601

[9] Zinkevich M, Johanson M, Bowling M, et al. Regret minimization in games with incomplete information. In: Proceedings of 21st Annual Conference on Neural Information Processing Systems, Vancouver, Canada, 2007. 1729-1736

[10] Abou Risk N, Szafron D. Using counterfactual regret minimization to create competitive multiplayer poker agents. In: Proceedings of the 9th International Conference on Autonomous Agents and Multiagent Systems, Toronto, Canada, 2010. 159-166

[11] Straub P G. Risk dominance and coordination failures in static games. *Quart Rev Econ Finance* 35, 1995, 35(4): 339-363

[12] Long J, Sturtevant N R, Buro M. Understanding the success of perfect information monte carlo sampling in game tree search. *Association for the Advancement of Artificial Intelligence AAAI*-10, 2010, 1: 134-140

[13] Selten R. An axiomatic theory of a risk dominance measure for bipolar games with linear incentives. *Games and Economic Behavior*, 1995, 8(1): 213-263

[14] Xia Z Y, Hu Y, Wang J, et al. Analyze and guess type of piece in the computer game intelligent system. In: Proceedings of the 2nd International Conference on Fuzzy Systems and Knowledge Discovery, Changsa, China, 2005. 1174-1183

[15] 马骁，王轩，王晓龙. 一类非完备信息博弈的信息模型. 计算机研究与发展，2010，47(12)：2100-2109

Analysis and estimation of machine game risks

Zhang Jiajia, Wang Xuan

(Shenzhen Graduate School, Harbin Institute of Technology, Shenzhen 518055)

Abstract

The concept of risk lost is put forward, and several factors that possibly bring risk lost in computer games are discussed. The risk lost in perfect and imperfect information games are analyzed, and a method for estimation of potential risk lost in the game process is proposed. Based on the results of the examinations on perfect and imperfect information games, the proposed standpoints are proved and the efficiency of the proposed estimating method for estimation of risk lost is tested. In the end, a game strategy decision method based on risk lost is presented.

Keywords: multi-equilibrium, risk dominance, imperfect information game, risk lost