第四部分： 多Agent决策

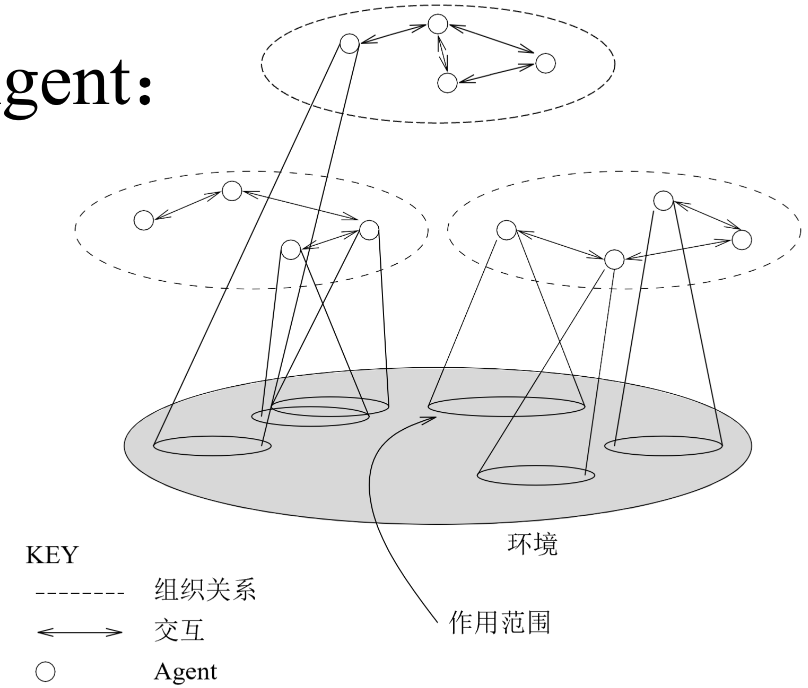
章宗长

2023年4月11日

内容安排

多Agent交互

- 一个多Agent系统包含一些Agent：
  - 通过通信相互交互
  - 可以在环境中动作
  - 有不同的“作用范围”
    - 可以控制、至少是影响环境的不同部分
  - 影响的范围可能重叠
    - 产生Agent间的依赖关系
      - 如：两个Agent都可以通过一道门，但不能同时通过
  - 其他关系
    - 如：“权力”关系



多Agent系统的标准结构

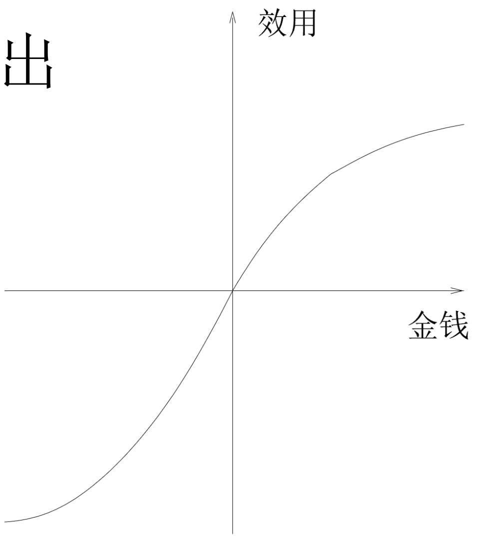
本节最重要的内容：理解发生在这些Agent之间的交互类型

多Agent交互

- **效用和偏好**
- 多Agent相遇
- 解的概念和性质
- 竞争与零和交互
- 囚徒困境
- 其他的对称2×2交互
- 多Agent系统的依赖关系

效用（Utility）

- 个体在交互时需要考虑什么？
  - 理性假设：什么样的结果对自己最有利，即**效用**
- 效用
  - 来源：经济学中的概念，是**价值**的抽象。
  - 表示**结果**对某个**个体**的**价值**，能够反映出一个个体的**意图**
- 效用函数的定义
  - Agent：$Ag = \{i, j\}$
  - 结果$\Omega = \{\omega_1, \omega_2, \ldots\}$
  - 效用函数：$u_i: \Omega \to \mathbb{R}$，$u_j: \Omega \to \mathbb{R}$



金钱与效用之间的关系

偏好（Preference）

- 效用函数使得结果能够进行**比较**
  - 偏好关系：$u_i(\omega) \geq u_i(\omega') \Leftrightarrow \omega \succcurlyeq_i \omega'$
  - 严格偏好关系：$u_i(\omega) > u_i(\omega') \Leftrightarrow \omega \succ_i \omega'$

- 性质：偏好关系是全序关系
  - 自反性：$\omega \succcurlyeq_i \omega$
  - 传递性：如果$\omega_1 \succcurlyeq_i \omega_2$和$\omega_2 \succcurlyeq_i \omega_3$，则$\omega_1 \succcurlyeq_i \omega_3$
  - 完全性：对于任意$\omega_1, \omega_2 \in \Omega$，有$\omega_1 \succcurlyeq_i \omega_2$或$\omega_2 \succcurlyeq_i \omega_1$

- 很多时候，人们并不关心结果的具体效用是多少，只关心偏好

多Agent交互

- 效用和偏好
- **多Agent相遇**
- 解的概念和性质
- 竞争与零和交互
- 囚徒困境
- 其他的对称2×2交互
- 多Agent系统的依赖关系

环境

- 环境
  - Agent：$\mathrm{Ag} = \{i, j\}$
  - 动作$Ac = \{C, D\}$
    - C：合作（cooperate），D：不合作（defect）
  - 结果$\Omega = \{\omega_1, \omega_2, \dots\}$
  - 状态转移函数τ: $\underbrace{Ac}_{\text{Agent } i\text{的动作}} \times \underbrace{Ac}_{\text{Agent } j\text{的动作}} \rightarrow \Omega$
- 如何拓展这样的环境？
  - 引入更多的Agent
  - 动作加入更多的选择
  - 设计更复杂的状态转移函数
  - 在状态转移中引入随机性

环境函数的例子（一）

- 每个个体的行为都会对结果产生影响
  - $\tau(D,D)=\omega_1, \tau(C,D)=\omega_2, \tau(D,C)=\omega_3, \tau(C,C)=\omega_4$

注：形式为(*j*’s action, *i*’s action)

	i **Defects**	*i* **Cooperates**
j **Defects**	ω_1	ω_3
j **Cooperates**	ω_2	ω_4

这个环境对每个Agent执行的动作是敏感的

环境函数的例子（二）

- 个体的行为对结果没有影响
  - $\tau(D,D)=\omega_1, \tau(D,C)=\omega_1, \tau(C,D)=\omega_1, \tau(C,C)=\omega_1$

	i **D**efects	*i* **C**ooperates
j **D**efects	ω_1	ω_1
j **C**ooperates	ω_1	ω_1

在这个环境中，不论Agent做什么动作，结果都是相同的

环境函数的例子（三）

- 只有一个个体的行为对结果有影响
  - $\tau(D,D)=\omega_1, \tau(D,C)=\omega_1, \tau(C,D)=\omega_2, \tau(C,C)=\omega_2$

	i **Defects**	*i* **Cooperates**
j **Defects**	ω_1	ω_1
j **Cooperates**	ω_2	ω_2

在这个环境中，结果只依赖于Agent *j*执行的动作

效用函数

- 把环境与效用结合起来，有意思的事情就发生了
- 假设Agent的效用函数如下：
  - Agent i：$u_i(\omega_1) = 1, u_i(\omega_2) = 1, u_i(\omega_3) = 4, u_i(\omega_4) = 4$
  - Agent j：$u_j(\omega_1) = 1, u_j(\omega_2) = 4, u_j(\omega_3) = 1, u_j(\omega_4) = 4$

	i **Defects**	i **Cooperates**
j **Defects**	ω_1	ω_3
j **Cooperates**	ω_2	ω_4

Agent i：$u_i(D,D) = 1, u_i(D,C) = 1, u_i(C,D) = 4, u_i(C,C) = 4$

Agent j：$u_j(D,D) = 1, u_j(D,C) = 4, u_j(C,D) = 1, u_j(C,C) = 4$

理性的动作

Agent i：$u_i(D,D)=1, u_i(D,C)=1, u_i(C,D)=4, u_i(C,C)=4$
Agent j：$u_j(D,D)=1, u_j(D,C)=4, u_j(C,D)=1, u_j(C,C)=4$

- Agent i对于可能的结果的偏好
  - $(C,C) \succcurlyeq_i (C,D) \succ_i (D,C) \succcurlyeq_i (D,D)$

在这种情况下，Agent i应该选择什么动作？

- Agent i更愿意选择通过合作产生的结果
  - 因此，Agent i的理性动作是C

收益矩阵（Pay-off Matrix）

	i **Defects**	i **Cooperates**
j **Defects**	ω_1	ω_3
j **Cooperates**	ω_2	ω_4

- 假设Agent的效用函数如下
  - Agent i：$u_i(\omega_1) = 1, u_i(\omega_2) = 1, u_i(\omega_3) = 4, u_i(\omega_4) = 4$
  - Agent j：$u_j(\omega_1) = 1, u_j(\omega_2) = 4, u_j(\omega_3) = 1, u_j(\omega_4) = 4$

收益矩阵：

	i **Defects**	i **Cooperates**
j **Defects**	1 1	4 1
j **Cooperates**	1 4	4 4

- 本节的剩余部分将使用收益矩阵表示双人博弈
  - 正则形式的博弈：采用这种矩阵形式表示的博弈

多Agent交互

- 效用和偏好
- 多Agent相遇
- **解的概念和性质**
- 竞争与零和交互
- 囚徒困境
- 其他的对称2×2交互
- 多Agent系统的依赖关系

正则形式的博弈

- 回顾
  - 效用：个体的偏好或目的
  - 环境：个体之间是如何交互的
  - 正则形式的博弈：多Agent交互最基本的模型

下一步是什么？

- 博弈论最基本的问题
  - 个体应该怎么做，才能得到最大的收益？

正则形式的博弈

- 优势策略
- 纳什均衡
- 帕累托最优
- 最大化社会福利

不同类型的解

最优反应

- s_i： Agent i的策略
- $s_{1:n}$： 所有n个Agent的策略组合（strategy profile）
- s_{-i}： 所有Agent中除去Agent i的策略组合

- $U_i(s_{1:n})$或$U_i(s_i, s_{-i})$： 给定一个策略组合$s_{1:n}$， Agent i的效用

- Agent i对策略组合s_{-i}的一个最优反应（best response）是一个策略s_i^*，满足：

  对所有策略s_i，有$U_i(s_i^*, s_{-i}) \geq U_i(s_i, s_{-i})$

- 一般地，给定s_{-i}，可以有多个不同的最优反应

优势策略

- 在一些博弈中，存在一个s_i，它是所有可能s_{-i}的一个最优反应，称s_i为一个优势策略（dominant strategy）

优势策略均衡

	i **Defects**	*i* **Cooperates**	*i*的最优反应
j **Defects**	4 4	1 3	**Defect**
j **Cooperates**	3 1	2 2	**Defect**

假设Agent j采取策略D，那么Agent i的**最优反应**是D
假设Agent j采取策略C，那么Agent i的**最优反应**是D

D是Agent i的**优势策略**

- 当所有Agent都使用优势策略，称它们的组合为优势策略均衡（dominant strategy equilibrium）

优势策略的性质

- 好消息：优势策略让决策变得非常简单

	i **Defects**	*i* **Cooperates**
j **Defects**	4 / 4	1 / 3
j **Cooperates**	3 / 1	2 / 2

- 坏消息：优势策略不一定存在

	i **Defects**	*i* **Cooperates**	*i*的最优反应
j **Defects**	4 / 4	3 / 3	Defect
j **Cooperates**	1 / 1	2 / 2	Cooperate

*j*不合作，*i*的最好反应是不合作

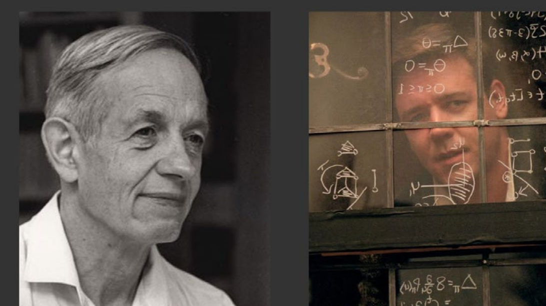
*j*合作，*i*的最好反应是合作

纳什均衡

- 优势策略并不适用于所有的问题，因此需要更**通用**的解决方法
- 纳什均衡（Nash equilibrium）
  - 策略s_i和s_j处在**纳什均衡**，当且仅当
    - 当Agent i采取策略s_i，Agent j没有比s_j更好的策略
    - 当Agent j采取策略s_j，Agent i没有比s_i更好的策略
  - 或者说，Agent i和Agent j的策略s_i, s_j互为**最优反应**

John Forbes Nash

(Nobel Laureate in Economics)



Portrayed by Russel Crowe in the film "A Beautiful Mind"

纳什均衡的性质

- 理性个体思考的方式：
  - 如果对方选择a_1，那么我应该选择b_1，则对方会选择a_2，那么我应该选择b_2……则对方会选择a'，那么我应该选择b'，则对方会选择a'，那么我应该选择b'
  - 那么a', b'处在纳什均衡
- 当策略s_i, s_j处在纳什均衡时，Agent没有动力来改变原来的策略，因此非常稳定

纳什均衡的性质

- 即使博弈问题没有优势策略，但还是可能存在纳什均衡

	i **Defects**	i **Cooperates**	i的最优反应
j **Defects**	4 4	3 3	Defect
j **Cooperates**	1 1	2 2	Cooperate

在这种情况下，没有优势策略，但存在纳什均衡(D, D)

纯策略的纳什均衡

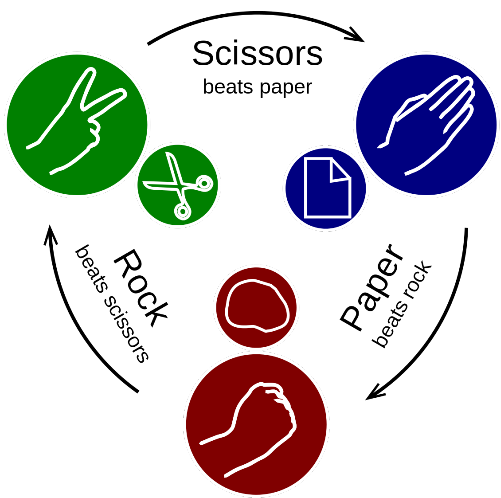
- 一个Agent的策略（strategy）分为：
  - 纯策略：行动的选择是确定性的
    - 如：以1的概率选择不合作
  - 混合策略：行动的选择是概率性的
    - 如：[不合作: 0.7; 合作: 0.3]

纯策略是混合策略的一种特殊情形

- 两个重要的结论：
  - 并不是每个交互的情形都有纯策略的纳什均衡
  - 有些交互的情形有一个以上的纯策略的纳什均衡

纯策略的纳什均衡（续）

- 在石头-剪子-布博弈中，纯策略的纳什均衡**不存在**



	i 石头	i 剪刀	i 布
j 石头	0 0	-1 1	1 -1
j 剪刀	1 -1	0 0	-1 1
j 布	-1 1	1 -1	0 0

游戏规则：
石头克剪刀，剪刀克布，布克石头

纯策略的纳什均衡（续）

- 在猎鹿博弈中，存在多个纯策略的纳什均衡

猎鹿博弈（Stag hunt）
在这一场景下，两名猎人一起去打猎，他们可以猎取鹿，也可以猎取野兔。鹿需要两个人合作才能获取，野兔一个人就可猎得，但猎鹿所得的收益大于猎野兔所得的收益。

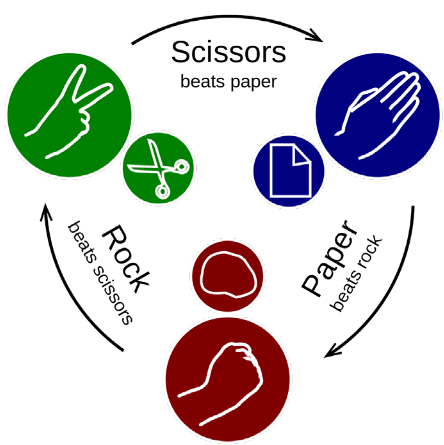
	i 鹿	i 兔
j 鹿	3, 3	0, 2
j 兔	2, 0	1, 1

纯策略的纳什均衡为：两个猎人都去打鹿，或者都去打兔

混合策略的纳什均衡

- 纳什定理：一个动作空间有限的博弈问题，**总是存在**至少一个纳什均衡解。

- 例子：石头-剪子-布博弈没有纯策略的纳什均衡，但是有混合策略的纳什均衡
  - 以相同的概率选取石头，剪刀和布中的一个



	i 石头	i 剪刀	i 布
j 石头	0 0	-1 1	1 -1
j 剪刀	1 -1	0 0	-1 1
j 布	-1 1	1 -1	0 0

优势策略均衡 vs. 纳什均衡

优势策略均衡 ⊆ 纯策略的纳什均衡 ⊆ 混合策略的纳什均衡

- 优势策略均衡：自己的一个动作，是对方**所有**动作的最优反应
- 纯策略的纳什均衡：自己的动作与对方的动作，**互为最优反应**

在策略中引入随机性

- 混合策略的纳什均衡：自己的策略与对方的策略，互为最优反应

帕累托最优（Pareto Optimality）

- 帕累托最优性（也称帕累托效率，Pareto efficiency）

结果ω为帕累托最优 $\Longleftrightarrow$ 如果想要提升一个个体的效用，必须以其它个体效用的减少作为代价

结果ω不是帕累托最优 $\Longleftrightarrow$ 存在一个结果ω'，使得至少一个个体的效用提高，同时不会使其余个体的效用减少

例1：

	i **Defects**	i **Cooperates**
j **Defects**	5 3	1 2
j **Cooperates**	0 2	0 1

帕累托最优解为(D, D)

例2：

	i **Defects**	i **Cooperates**
j **Defects**	2 2	0 5
j **Cooperates**	5 0	3 3

帕累托最优解为(C, D), (D, C), (C, C)

社会福利（Social welfare）

- 一个结果ω的社会福利，是所有Agent效用的总和

$$\sum_{i \in Ag} u_i(\omega)$$

- 在**合作型**的多Agent系统中，目标往往是最大化社会福利

例1：

		i	
		Defect	**Coop**
j	**Defect**	2 2	1 1
	Coop	3 3	4 4

例2：

		i	
		Defect	**Coop**
j	**Defect**	2 2	1 1
	Coop	3 3	7 0

(C, C)最大化社会福利

多Agent交互

- 效用和偏好
- 多Agent相遇
- 解的概念和性质
- **竞争与零和交互**
- 囚徒困境
- 其他的对称2×2交互
- 多Agent系统的依赖关系

竞争

- 对于一个Agent，它与其他Agent存在**竞争**，如果
  - 当它的效用最大时，其它Agent的效用并不是最大
- 严格竞争性
  - 对于环境中的任意两个结果$\omega, \omega' \in \Omega$，满足

$$\omega \succ_i \omega' \Leftrightarrow \omega' \succ_j \omega$$

Agent i和Agent j的偏好相互处于完全对立的位置

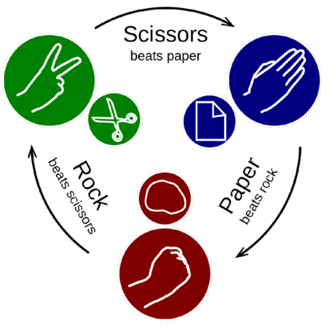
- **零和博弈**满足严格竞争性
  - 竞争最“激烈的”博弈形式

零和博弈

- 零和博弈（Zero-sum game）
  - 对于任意结果$\omega \in \Omega$，两个Agent的效用之和为零：

$$u_i(\omega) + u_j(\omega) = 0$$

- 零和博弈的例子
  - 下棋
  - 石头-剪子-布



- 现实世界中，真正的零和博弈是非常罕见的
  - 在很多情形下，有相互利益合作的可能

多Agent交互

- 效用和偏好
- 多Agent相遇
- 解的概念和性质
- 竞争与零和交互
- **囚徒困境**
- 其他的对称2 × 2交互
- 多Agent系统的依赖关系

囚徒困境

- 有两个囚徒（Agent）被隔离审讯
  - 一方可以揭发（即不合作，D）另一方，也可以保持沉默（即合作，C）
  - 如果i揭发j，j保持沉默，则i被释放，j判10年
  - 如果i和j相互揭发，则i和j各判5年
  - 如果i和j都保持沉默，则各判1年
- 假设两个Agent都知晓右图的收益矩阵

	i 揭发（D）	i 沉默（C）
j 揭发（D）	-5 -5	-10 0
j 沉默（C）	0 -10	-1 -1

- 两个Agent需要在不知晓另一方行动的前提下，同时行动

囚徒困境（续）

- 对于理性的Agent *i*来说：
  - 当Agent *j*选择揭发时，最优反应是**揭发**
  - 当Agent *j*选择沉默时，最优反应是**揭发**
- 因此，相互揭发不仅是**纳什均衡**，**也是优势策略均衡**

	i 揭发（D）	*i* 沉默（C）
j 揭发（D）	-5 -5	-10 0
j 沉默（C）	0 -10	-1 -1

囚徒困境（续）

	i 揭发（D）	*i* 沉默（C）
j 揭发（D）	-5 -5	-10 0
j 沉默（C）	0 -10	-1 -1

- 直觉告诉我们，**双方保持沉默**是比相互揭发更好的选择
- 然而，如果两个Agent都做**理性的推理**，会做出揭发的选择
- 因此，这被称作是一个**困境**（Dilemma）

这一悖论是多Agent交互的一个根本问题

囚徒困境（续）

- 囚徒困境问题在真实世界中普遍存在
  - 如：两个国家达成核武器条约的问题
    - 合作（销毁核武器）还是不合作（保留核武器）？

为什么人类社会在面对这些问题时，有时不会选择完全不合作的方式呢？

- 人类交互时并不总是寻找最大化自己利益的行为
  - 具有牺牲精神，享受被信任，不会不考虑他人

重复进行的囚徒困境

- 囚徒困境的变种：重复进行的囚徒困境
  - 想法：对局进行多次，每一次对局称为“一轮”
  - 假设每个Agent都可以看到其对手前一轮的动作

- 重复进行的囚徒困境有两种类型：
  - 重复固定轮数的囚徒困境
  - 重复无限轮数的囚徒困境

这两种情形下，Agent的理性动作分别是什么？

重复固定轮数的囚徒困境

- 在固定轮数的囚徒困境中，不合作是理性的行为！
  - 假设双方都知道对局的轮数为n
- 反向归纳法（backwards induction）
  - 第n轮事实上是一次性囚徒困境对局，理性的行为是不合作
  - 但是，这又意味着“真正的”最后一轮是第$n-1$轮，同样的推理导致这一轮事实上也会像一次性囚徒困境对局，理性的行为是不合作
  - 继续这样的反向归纳过程，得到结论：**在固定的、事先双方都知道轮数的囚徒困境问题中，不合作是理性的**

重复无限轮数的囚徒困境

- 在无限轮数的囚徒困境中，合作是理性的行为！
- 两个原因：
  - 如果你现在不合作，对手将来可以通过不合作惩罚你
    - 这种惩罚不可能发生在一次对局的囚徒困境问题中
  - 如果你一开始通过合作“试试深浅”，并且在第一轮得到欺骗的回报，那么因为你们要无限期地进行对局，这次的效用损失会在将来的轮次中“分期偿还”
    - 相比合作可能带来的期望效用而言，这个损失更小

阿克塞尔罗德（Axelrod）竞赛

- 虽然个体之间交互的过程中会遭受其它个体的背叛，但仍然有**很大概率**遇见愿意合作的个体
- 阿克塞尔罗德竞赛
  - 一个个体，与**多个个体**进行多轮囚徒困境博弈
  - 被用于研究合作的行为，在由理性个体组成的**群体**中，是如何产生的

Robert Axelrod

阿克塞尔罗德竞赛

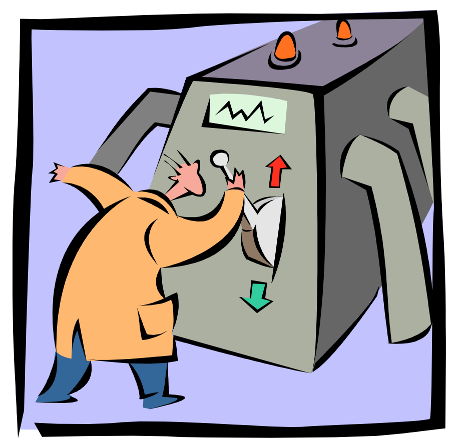
不同领域的专家被邀请参加比赛，设计程序来进行多轮囚徒困境博弈。参与比赛的程序两两之间进行一场比赛，对弈5局，每局由200轮的囚徒困境组成。比赛的最终胜利者是在所有程序中总体表现最好的。

一些参赛的程序

- 随机：总是以相同的概率选择合作（沉默，C）或背叛（揭发，D）
- ALL-D：无论对手之前选择什么，总是背叛
- 以牙还牙（TIT-FOR-TAT）：
  - 在首轮选择合作
  - 在第一轮后，重复对手上一轮的动作
- Joss：90%的时间采取以牙还牙的策略，10%的时间采取背叛
- Tester：在第1轮通过不合作检验对手。如果对手以不合作报复，那么接着采用以牙还牙的策略。如果对手合作，那么它重复进行两轮合作对局，然后不合作

思考

- 基于到目前为止已经了解的，特别是有关重复固定轮数的囚徒困境问题的结论，你认为哪个策略是总体上最好的？
- 如果你参加比赛，你会采取什么策略？



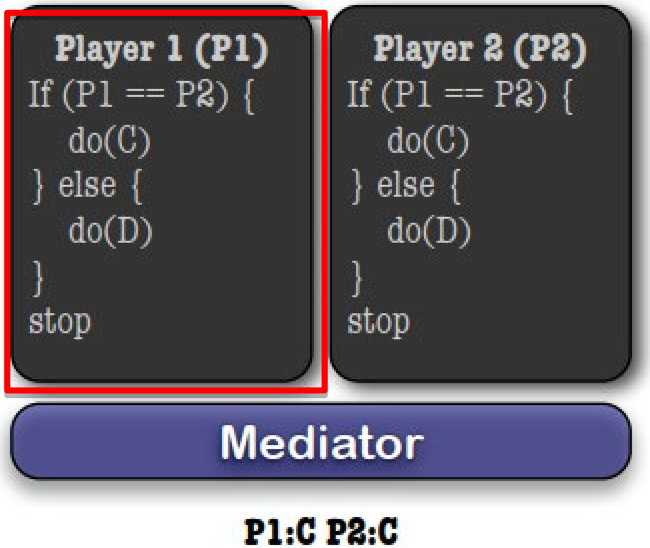
获胜者：以牙还牙（TIT-FOR-TAT）

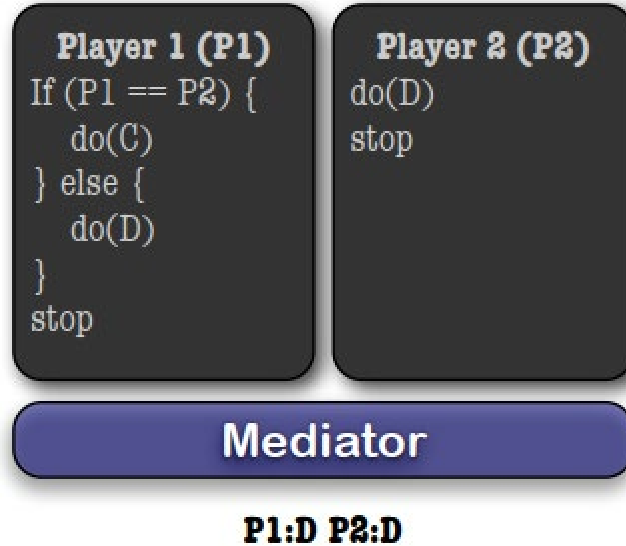
- 意想不到的是，以牙还牙策略最终取胜
  - 以牙还牙策略并不能够在多轮囚徒困境中，赢下所有对手
  - 当有足够多的机会遇到愿意合作的Agent，这种策略还是足够好的
- 获胜的启示
  - 不要过分想赢
  - 不要首先背叛
  - 正确回应合作和背叛行为
  - 不要过分聪明

程序均衡（Program Equilibria）

- 传统的策略直接给出动作，能否采取更加精细化的策略？
  - 例如：策略“如果你合作，那么我就合作”
- 新的规则
  - 每个Agent提交**程序**给仲裁（mediator），仲裁再同时执行提交的程序
  - 每个程序可以**基于对方给出的程序**，来决定采取什么动作

程序均衡（续）

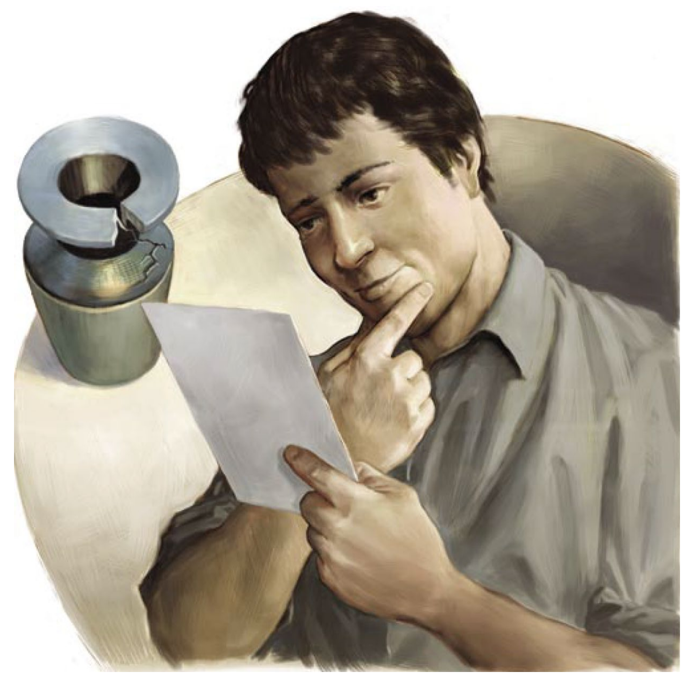




- 如果对方提交左边的程序，那么**最优反应**也是提交相同的代码
- 双方提交相同的程序形成**纳什均衡**

旅行者困境

- 航空公司弄丢了两个旅行者（Lucy和Pete）的两个相同的箱子
- 让旅行者写下他们箱子的价值a_i 和a_{-i} ，范围是$2和$100之间



旅行者困境的效用函数

- 效用函数

$$U_i(a_i, a_{-i}) = \begin{cases} a_i & \text{if } a_i = a_{-i} \\ a_i + 2 & \text{if } a_i < a_{-i} \\ a_{-i} - 2 & \text{otherwise} \end{cases}$$

例1：Lucy报价\$100，Pete报价\$100
赔偿给Lucy和Pete各\$100

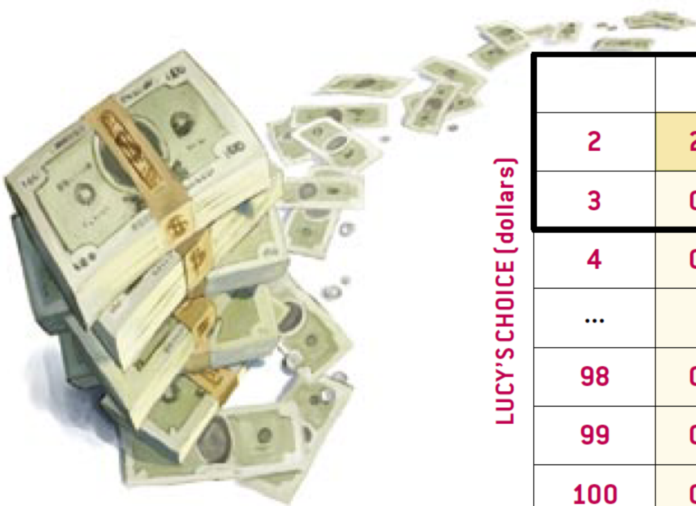
例2：Lucy报价\$10，Pete报价\$90
以\$10为基准，
赔偿给Lucy：\$10 + \$2 = \$12
赔偿给Pete：\$10 − \$2 = \$8

- **问**：大家倾向写多少钱？

- 大多数人倾向写\$97和\$100之间

旅行者困境的收益矩阵和纳什均衡

- 收益矩阵（又称效用矩阵）



PETE'S CHOICE (dollars)

LUCY'S CHOICE (dollars)

	2	3	4	...	98	99	100
2	2 2	4 0	4 0	...	4 0	4 0	4 0
3	0 4	3 3	5 1	...	5 1	5 1	5 1
4	0 4	1 5	4 4	...	6 2	6 2	6 2
...	...	...	...		...	...	...
98	0 4	1 5	2 6	...	98 98	100 96	100 96
99	0 4	1 5	2 6	...	96 100	99 99	101 97
100	0 4	1 5	2 6	...	96 100	97 101	100 100

事实上，旅行者困境问题只有唯一一个纳什均衡：

($2, $2)

多Agent交互

- 效用和偏好
- 多Agent相遇
- 解的概念和性质
- 竞争与零和交互
- 囚徒困境
- **其他的对称2 × 2交互**
- 多Agent系统的依赖关系

对称2×2交互

- 在囚徒困境问题中，Agent i偏好的排序：

$$(D,C)\succ_i (C,C)\succ_i (D,D)\succ_i (C,D)$$

- 如果把注意力限制在两个Agent的交互上，每个Agent有两个可能的动作（C或者D），这种情况是对称的，那么总共有$4!=24$个偏好排序
- 合作占优势的博弈
  - $(C,C)\succ_i (C,D)\succ_i (D,C)\succ_i (D,D)$
  - $(C,C)\succ_i (C,D)\succ_i (D,D)\succ_i (D,C)$
- 合作不占优势的博弈
  - $(D,D)\succ_i (D,C)\succ_i (C,C)\succ_i (C,D)$
  - $(D,D)\succ_i (D,C)\succ_i (C,D)\succ_i (C,C)$
- 留下了两个有意思的情况需要研究：
  - 猎鹿博弈（stag hunt）：$(C,C)\succ_i (D,C)\succ_i (D,D)\succ_i (C,D)$
  - 小鸡博弈（game of chicken）：$(D,C)\succ_i (C,C)\succ_i (C,D)\succ_i (D,D)$

猎鹿博弈

- 猎鹿博弈的收益矩阵如下：

	i **Defects**	*i* **Cooperates**
j **Defects**	1 1	0 2
j **Cooperates**	2 0	3 3

猎鹿博弈（Stag hunt）
在这一场景下，两名猎人一起去打猎，他们可以猎取鹿，也可以猎取野兔。鹿需要两个人合作才能获取，野兔一个人就可猎得，但猎鹿所得的收益大于猎野兔所得的收益。

  - “猎鹿”是合作行为，“猎兔”是不合作行为

- 有两个纳什均衡解
  - 分别是(C,C)和(D,D)
- 帕累托最优解只有(C,C)
  - 理性的Agent会排除解(D,D)

小鸡博弈

- 小鸡博弈的收益矩阵如下：

	i **Defects**	*i* **Cooperates**
j **Defects**	0 / 0	1 / 3
j **Cooperates**	3 / 1	2 / 2

小鸡博弈

两个局中人通过驾驶他们的汽车向悬崖高速驶去，两者中最不勇敢的人（“小鸡”）是第一个绕开悬崖的人。胜者是在车上坚持时间最长的人。

这里，合作表示绕开悬崖，不合作表示呆在车上。

- 有两个纳什均衡解：(D,C)和(C,D)
- 有三个帕累托最优解：(D,C)，(C,D)，(C,C)
- 有三个社会福利最优解：(D,C)，(C,D)，(C,C)

多Agent交互

- 效用和偏好
- 多Agent相遇
- 解的概念和性质
- 竞争与零和交互
- 囚徒困境
- 其他的对称2 × 2交互
- **多Agent系统的依赖关系**

独立关系 & 单向依赖

- 独立（independence）：个体要实现目标，不需要对方的帮助

	i **Defects**	*i* **Cooperates**
j **Defects**	1 1	0 1
j **Cooperates**	1 0	0 0

- 单向（unilateral）：一方实现目标需要对方帮助，另一方不需要

	i **Defects**	*i* **Cooperates**
j **Defects**	0 1	0 1
j **Cooperates**	1 0	0 0

互惠依赖 & 共同依赖

- 互惠（reciprocal）：双方实现目标，都需要对方的帮助
  - 双方的目标可以不同

	i **Defects**	*i* **Cooperates**
j **Defects**	0 / 0	1 / 0
j **Cooperates**	1 / 1	1 / 1

- 共同（mutual）：互惠依赖的一种形式
  - 双方有相同的目标

	i **Defects**	*i* **Cooperates**
j **Defects**	0 / 0	1 / 1
j **Cooperates**	1 / 1	2 / 2

小结

- 博弈
  - 效用：定义环境中不同结果对于个体的**价值**
  - 环境：定义个体的行为会带来什么结果
  - 策略：纯策略、混合策略
  - 解：优势策略、纳什均衡、帕累托最优、社会福利最优
- 对称2 × 2交互
  - 囚徒困境
    - 个人最佳选择并非团体最佳选择
    - 变种：重复进行（重复固定轮数、无限轮数的）囚徒困境
  - 猎鹿博弈、小鸡博弈……

课后作业4-1

- 考虑下列交互的情形：

	i **Defect**	*i* **Coop**
j **Defect**	-5 -5	-10 0
j **Coop**	0 -10	-1 -1

	i **Defect**	*i* **Coop**
j **Defect**	-5 -5	-4 0
j **Coop**	0 -4	-1 -1

	i **Descend**	*i* **Climb**
j **Descend**	-4 -4	-1 0
j **Climb**	0 -1	-5 -5

对于每种情形，写出所有的（纯策略）纳什均衡解、帕累托最优解和社会福利最优解。

注：形式为(*j*'s action, *i*'s action)

课后作业4-2

- 考虑下列交互的情形：

i

		defect	coop
j	defect	1 1	2 4
	coop	4 2	3 3

i

		defect	coop
j	defect	-1 1	1 -1
	coop	1 -1	-1 1

i

		defect	coop
j	defect	5 3	1 2
	coop	0 2	0 1

对于每种情形，写出所有的（纯策略）纳什均衡解、帕累托最优解和社会福利最优解。

注：形式为(j’s action, i’s action)

课后作业4-3

- 解释为什么在只有一轮的囚徒困境问题中，得到的程序均衡解可以是相互合作的行为?