博弈

2019年度南京大学“专创融合”特色示范课程培育项目

高 阳

http://cs.nju.edu.cn/rl, 2019.11.19

博 弈

从极小极大到博弈均衡

高　阳

http://cs.nju.edu.cn/gaoy, 2019.11.19

大 纲

极小极大搜索

α-β剪枝

帕里托优和纳什均衡

协商(投票，拍卖和谈判)

大 纲

极小极大搜索

α-β剪枝

帕里托优和纳什均衡

协商(投票，拍卖和谈判)

极小极大过程

强化学习：与环境的交互

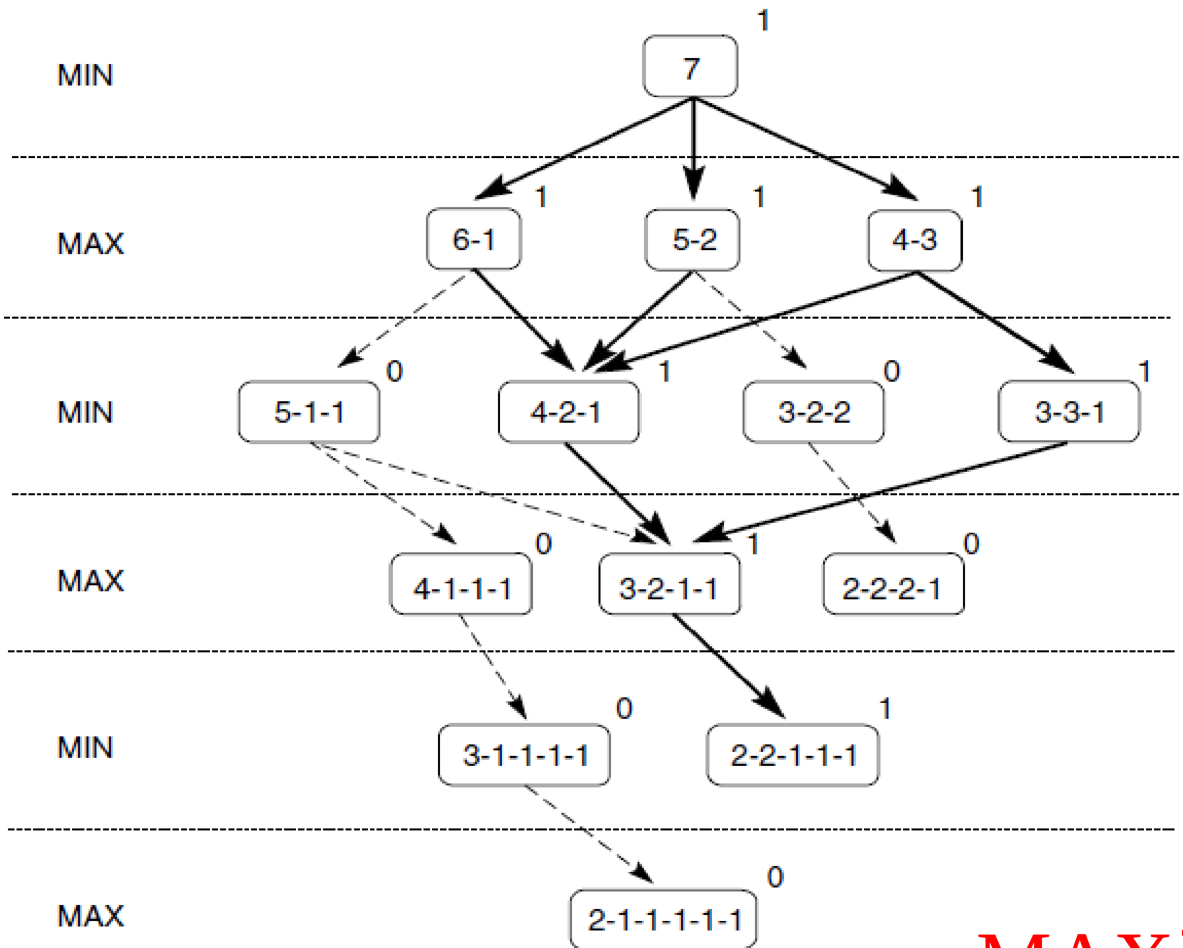
博弈：对对手行为/策略的建模

MAX：代表一方玩家，最大化其收益(赢得博弈)

MIN：代表对手，最小化MAX的收益

MIN总是移动到使MAX收益最坏的状态

余一棋



MIN走第一步

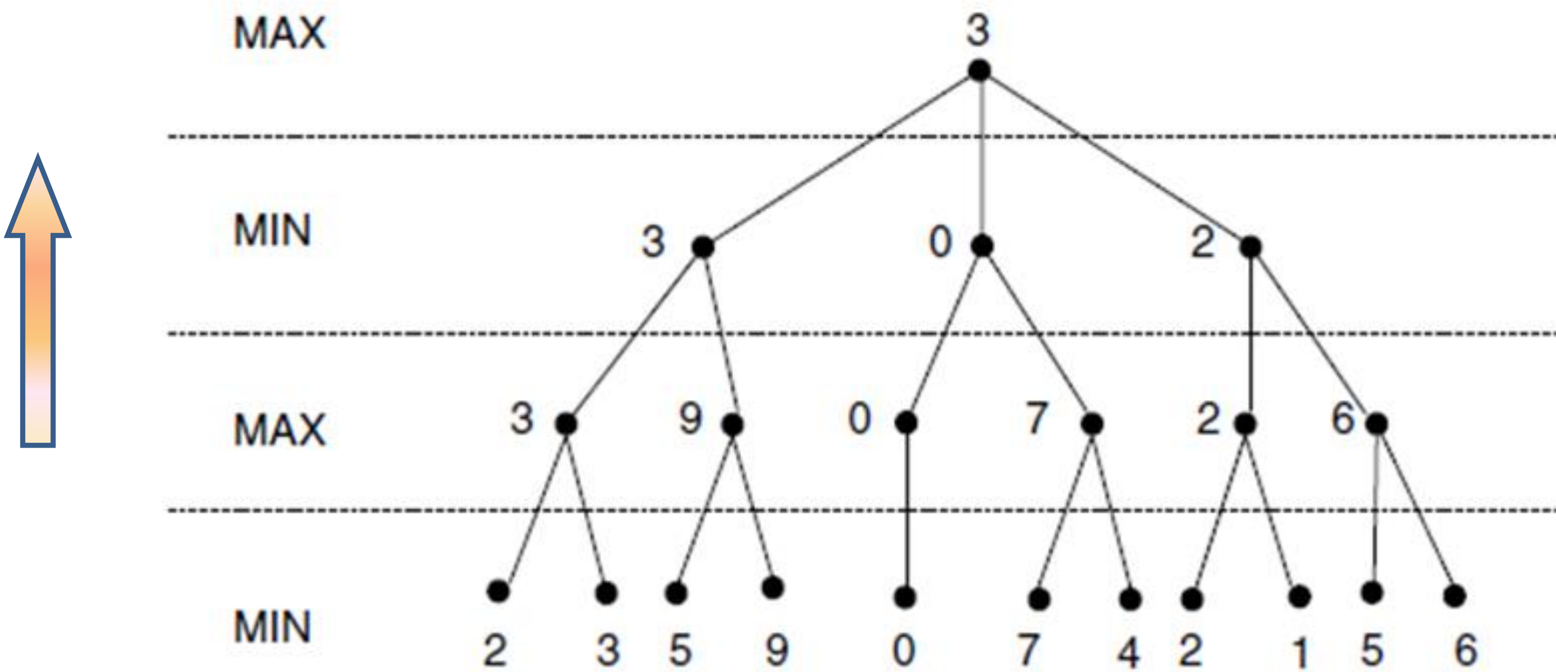
MAX走第二步

MAX获胜：值函数为1

MIN获胜：值函数为0

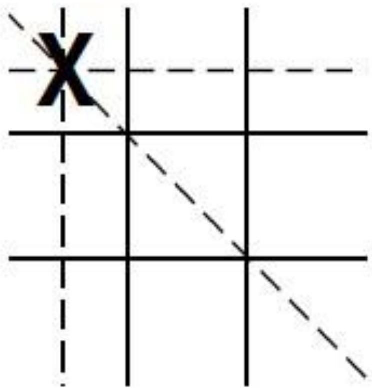
规则：将其分为两个不相等的两堆

极小极大搜索

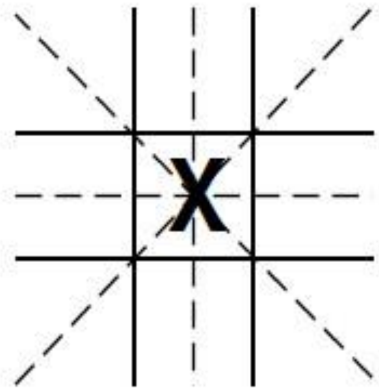


- **将启发值自底向上传播**
  - ✓ 如果父状态是MAX节点，将孩子节点中最大值传给它
  - ✓ 如果父状态是MIN节点，将孩子节点中最小值传给它

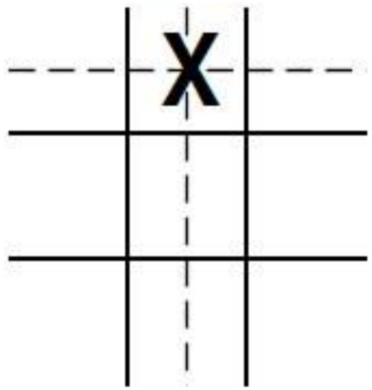
固定层深的极小极大过程



Three wins through
a corner square

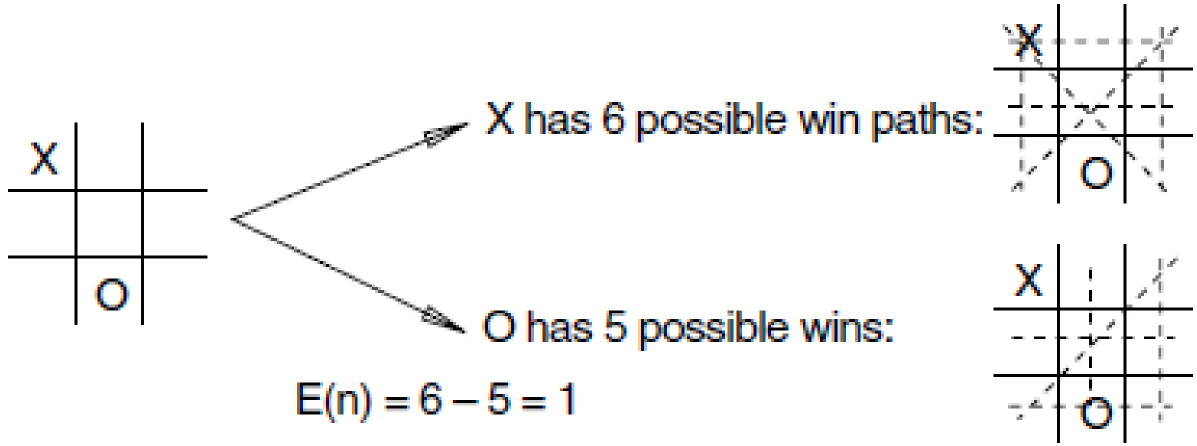


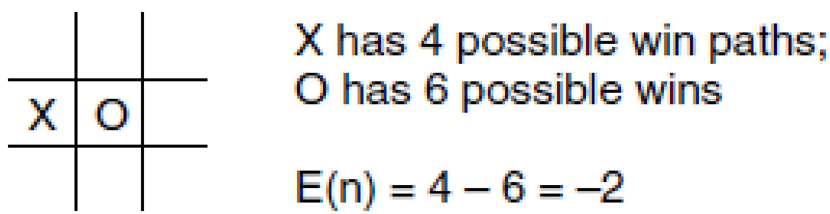
Four wins through
the center square

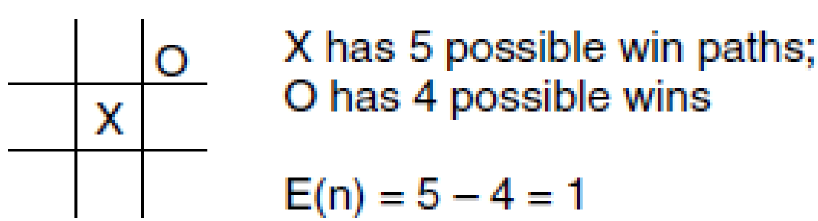


Two wins through
a side square

定义启发式函数

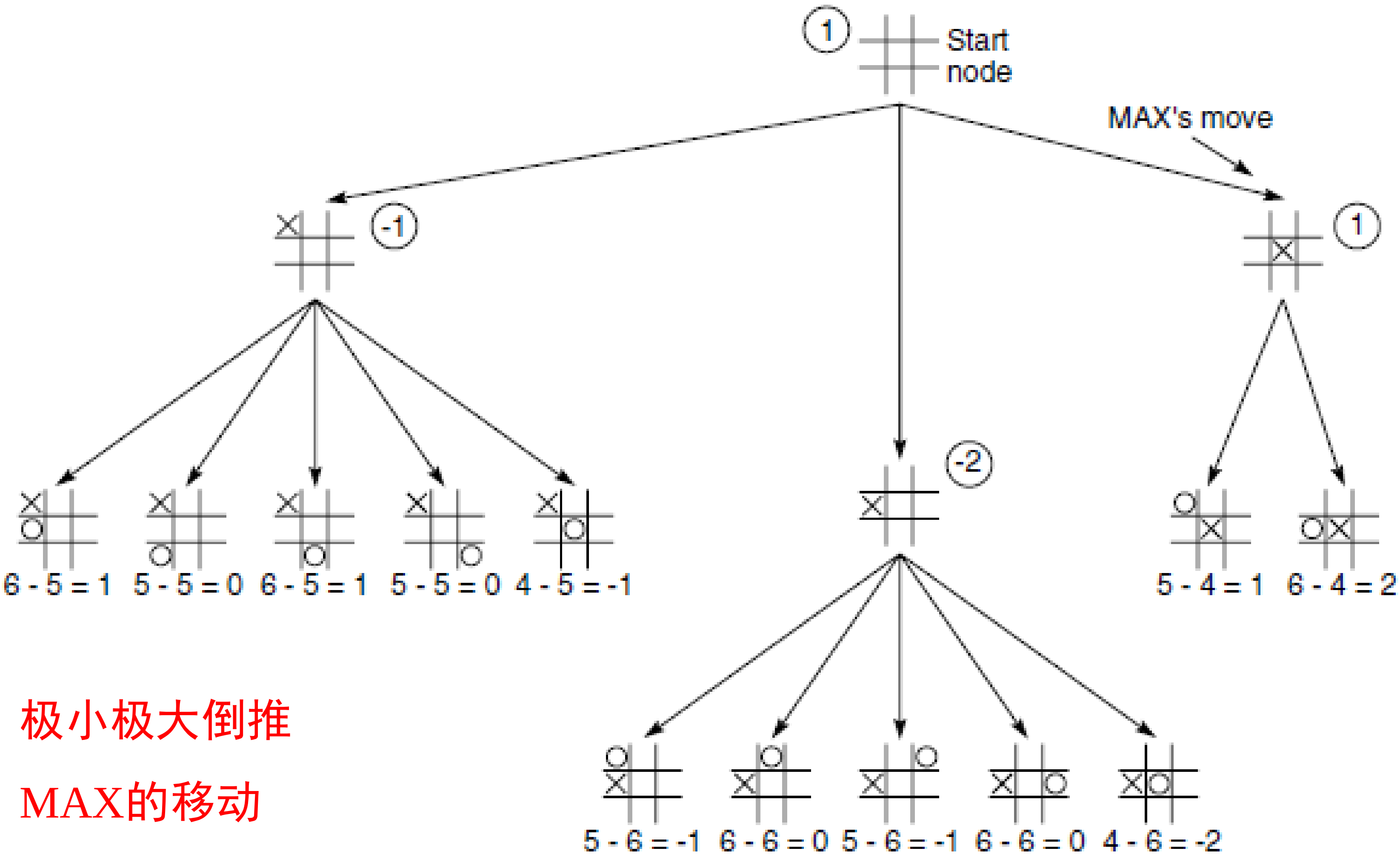






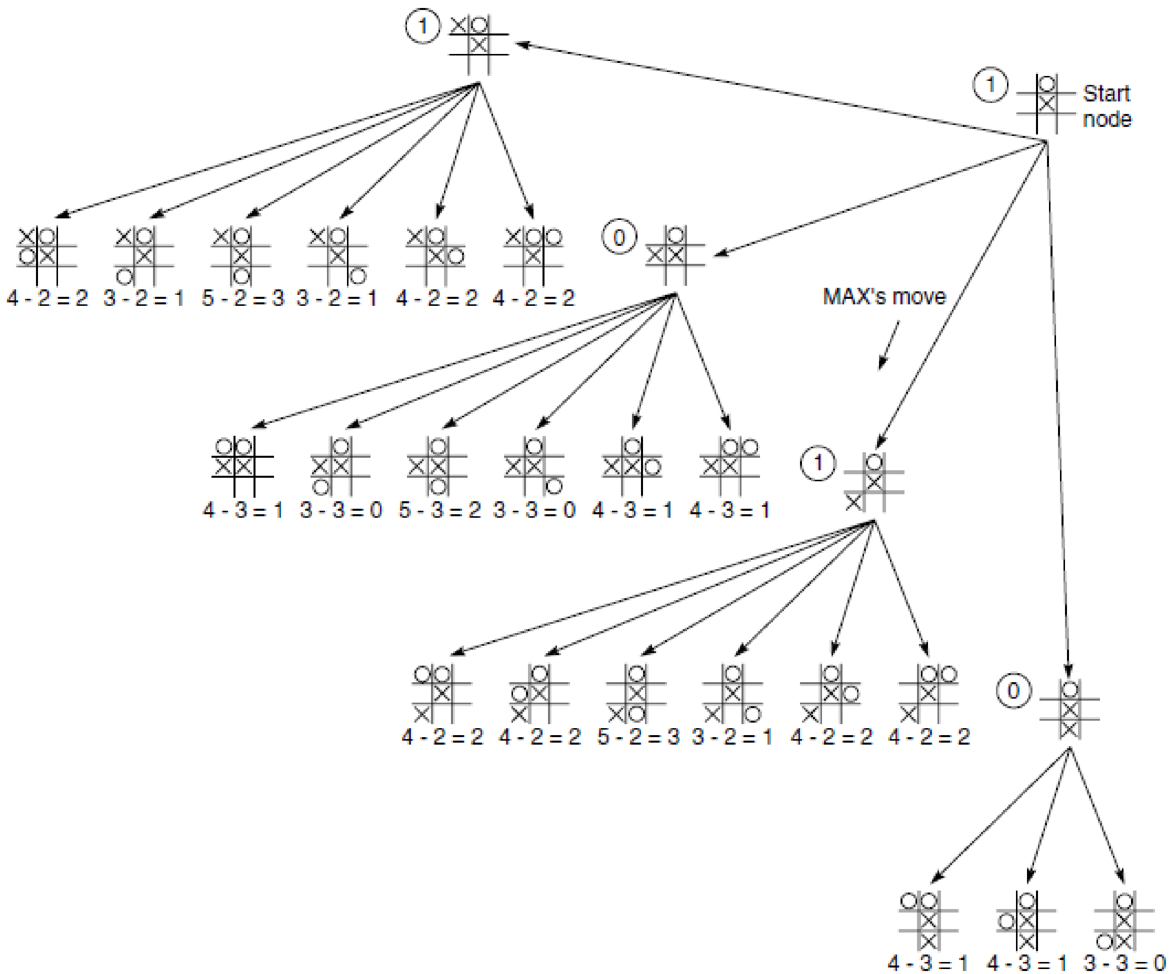
- 启发式评估：E(n)=M(n)-O(n)
  - ✓ M(n)是当前玩家可能获胜的行数
  - ✓ O(n)是对手可能获胜的行数

例：MAX开局移动



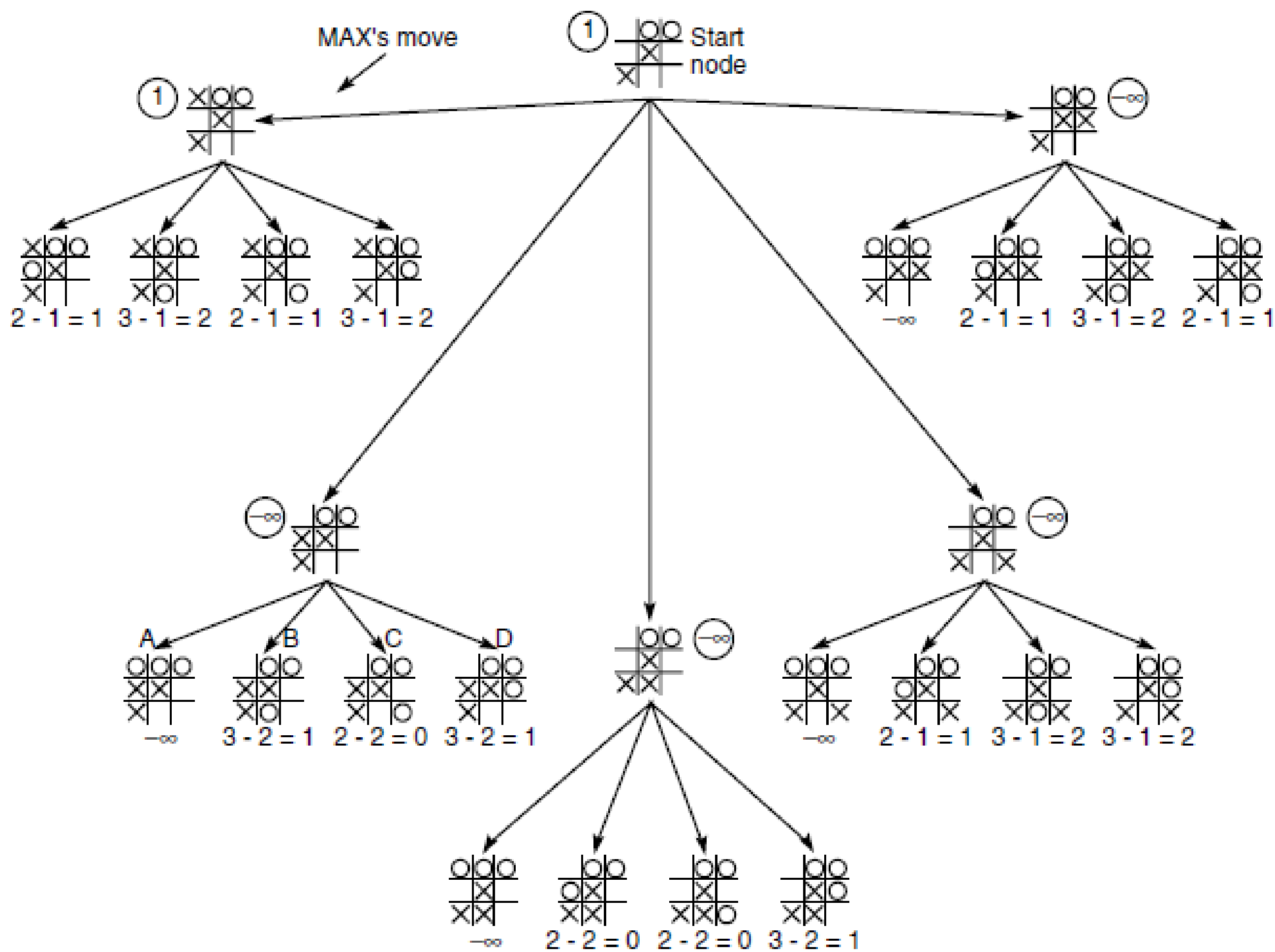
- 极小极大倒推
- MAX的移动

例：MAX第二步移动



1
Start node
MAX's move
0
4 - 2 = 2
3 - 2 = 1
5 - 2 = 3
3 - 2 = 1
4 - 2 = 2
4 - 2 = 2
4 - 3 = 1
3 - 3 = 0
5 - 3 = 2
3 - 3 = 0
4 - 3 = 1
4 - 3 = 1
4 - 2 = 2
4 - 2 = 2
5 - 2 = 3
3 - 2 = 1
4 - 2 = 2
4 - 2 = 2
4 - 3 = 1
4 - 3 = 1
3 - 3 = 0

例：MAX第三步移动


MAX's move
1
Start
node
1
−∞
2 - 1 = 1
3 - 1 = 2
2 - 1 = 1
3 - 1 = 2
−∞
2 - 1 = 1
3 - 1 = 2
2 - 1 = 1
−∞
−∞
−∞
A
B
C
D
−∞
3 - 2 = 1
2 - 2 = 0
3 - 2 = 1
−∞
2 - 1 = 1
3 - 1 = 2
3 - 1 = 2
−∞
2 - 2 = 0
2 - 2 = 0
3 - 2 = 1

大 纲

极小极大搜索

α-β剪枝

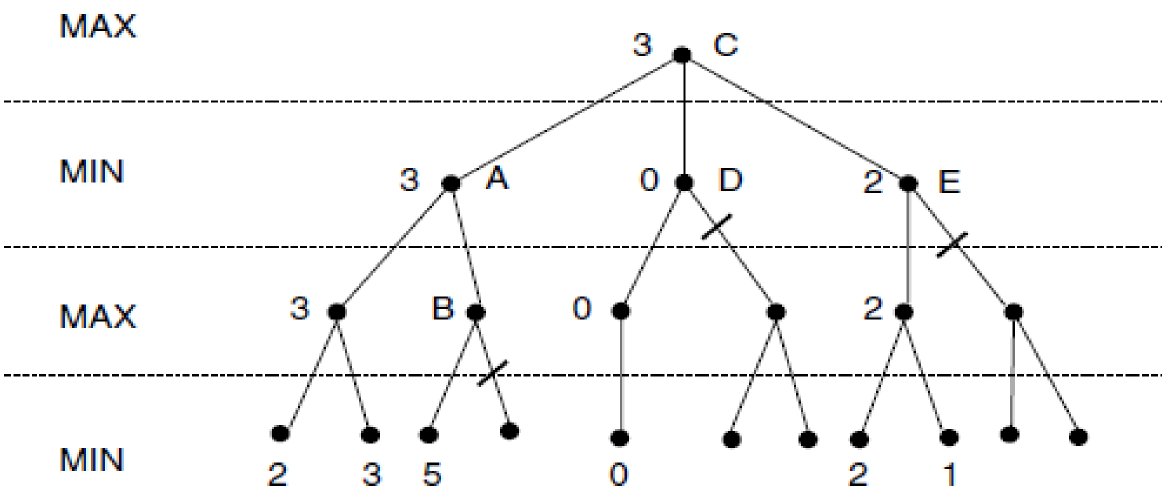
帕里托优和纳什均衡

协商(投票，拍卖和谈判)

α-β过程

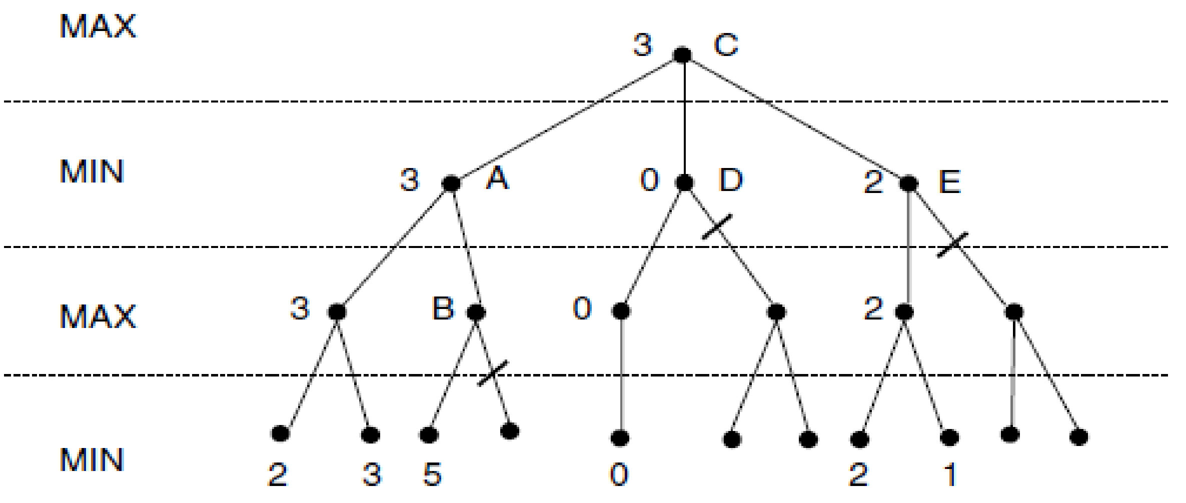
- 极小极大过程
  - ✓ 在预判层应用启发式评估
  - ✓ 展开所有的后继分支
  - ✓ 沿树向上传播评估值
- α-β剪枝
  - ✓ 当确定是一个dead end时，停止展开其后继节点
  - ✓ 是对博弈树的深度优先搜索，且维护
  - ✓ Alpha: 与MAX节点关联，从不减小
  - ✓ Beta: 与MIN节点关联，从不增大

α-β过程



A has $\beta = 3$ (A will be no larger than 3)
B is β pruned, since $5 > 3$
C has $\alpha = 3$ (C will be no smaller than 3)
D is α pruned, since $0 < 3$
E is α pruned, since $2 < 3$
C is 3

α-β过程



A has $\beta = 3$ (A will be no larger than 3)

B is β pruned, since $5 > 3$

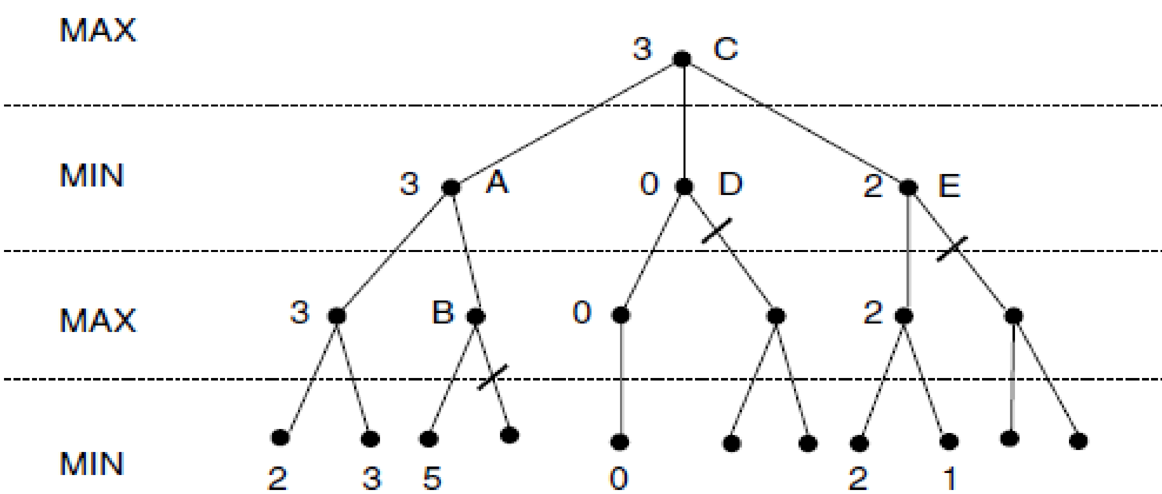
C has $\alpha = 3$ (C will be no smaller than 3)

D is α pruned, since $0 < 3$

E is α pruned, since $2 < 3$

C is 3

α-β过程



A has β = 3 (A will be no larger than 3)

B is β pruned, since 5 > 3

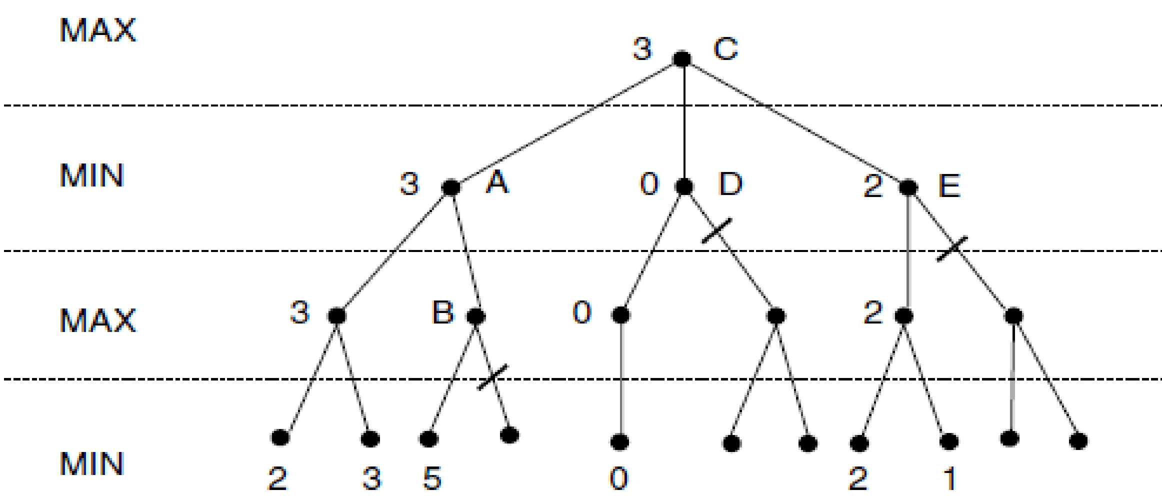
C has α = 3 (C will be no smaller than 3)

D is α pruned, since 0 < 3

E is α pruned, since 2 < 3

C is 3

α-β过程



A has β = 3 (A will be no larger than 3)

B is β pruned, since 5 > 3

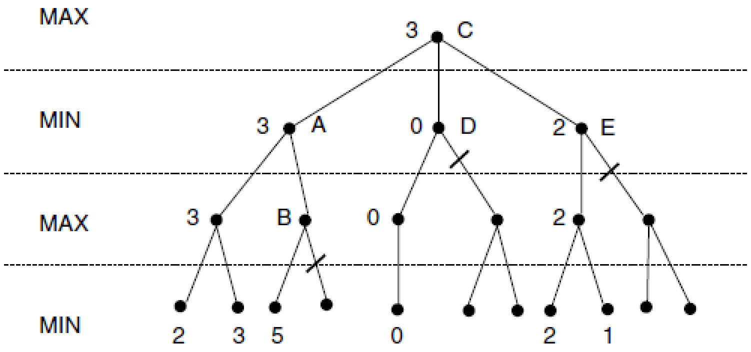
C has α = 3 (C will be no smaller than 3)

D is α pruned, since 0 < 3

E is α pruned, since 2 < 3

C is 3

剪枝规则



A has β = 3 (A will be no larger than 3)
B is β pruned, since 5 > 3
C has α = 3 (C will be no smaller than 3)
D is α pruned, since 0 < 3
E is α pruned, since 2 < 3
C is 3

剪枝规则

- ✓ Alpha剪枝: 任一MIN节点，如果其Beta值小于等于其祖先MAX节点的Alpha值，则停止搜索
- ✓ Beta剪枝: 任一MAX节点，如果其Alpha值大于等于其祖先MIN节点的Beta值，则停止搜索

大 纲

极小极大搜索

α-β剪枝

帕里托优和纳什均衡

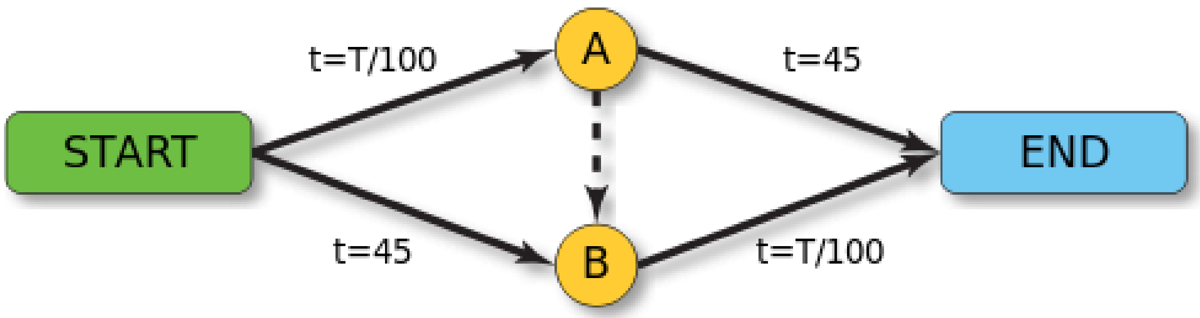
协商(投票，拍卖和谈判)

囚徒困境

		囚犯B	
		坦白	抗拒
囚犯A	坦白	(-5,-5)	(0,-10)
	抗拒	(-10,0)	(-1,-1)

- 每个囚徒如果只考虑自身的利益，则会选择‘坦白’行为；
- 而囚徒困境的最优策略是双方都选择‘抗拒’行为。

布雷斯悖论 Braess's paradox



- 考虑上图中的交通网，有4000辆车打算在其中路上通行。其中边上的数值表示通行时间，T表示边上的车辆数目。
- A到B的近路不存在
  - ✓ 2000辆从起点到A到终点，2000辆从起点到B到终点，65分钟
- A到B存在一条通行时间接近于0的近路
  - ✓ 所有司机都会选择从起点到A到B到终点，80分钟
  - ✓ 假如约定好不走近路？

最优策略

- 社会福利（Social Welfare）
  - ✓ 最大化所有参与者的收益和
- 帕里托优（Pareto Efficiency）
- 纳什均衡（Nash Equilibrium）
- 优超（Dominant）
  - ✓ 不依赖其他参与者

帕里托优

- 一个方案x是帕利脱最优，当且仅当不存在另一个方案 x’满足
  - $\exists$agent $ag: ut_{ag}(x') > ut_{ag}(x)$
  - $\forall$agent $ag': ut_{ag'}(x') \geq ut_{ag'}(x)$
- 帕利脱最优：不考虑跨Agent效益比较的情况下满足一个全局最优
- 帕利脱改善：在不减少一方利益的同时，通过改变现有的资源配置而提高另一方的利益
- 社会福利是帕利脱最优的一个子集
  - 一个agent要想提高自己的利益，必然存在其他agent的利益受损

纳什均衡

- Agent 的策略依赖于其他agent

  如果 $S^*_A=< S^*_1, S^*_2,\ldots\ldots, S^*_{|A|} >$ 为纳什均衡策略，当且仅当对agent i：S^*_i 对于 agent i 是最优策略当其他agent选择以下策略时 $< S^*_1, S^*_2,\ldots, S^*_{i-1}, S^*_{i+1} \ldots, S^*_{|A|}>$

- 没有参与者可以独自行动而增加收益

- 问题

  - ✓ 无纯Nash均衡解
  - ✓ 多个Nash均衡解

无纯策略NASH均衡解

		女孩		
		剪刀	石头	布
男孩	剪刀	(0,0)	(-1,1)	(1,-1)
	石头	(1,-1)	(0,0)	(-1,1)
	布	(-1,1)	(1,-1)	(0,0)

多个Nash均衡解

		女孩	
		球赛	电影
男孩	球赛	(2,1)	(0,0)
	电影	(0,0)	(1,2)

- 恋爱博弈问题
  - ✓ 男孩喜欢看比赛
  - ✓ 女孩喜欢看电影
  - ✓ 两人都希望一起度过周末

结婚后…

<table>
<tr><td colspan="2" rowspan="2"></td><td colspan="2">女孩</td></tr>
<tr><td>球赛</td><td>电影</td></tr>
<tr><td rowspan="2">男孩</td><td>球赛</td><td>(2,0)</td><td>(3,3)</td></tr>
<tr><td>电影</td><td>(-3,-3)</td><td>(0,2)</td></tr>
</table>

恋爱博弈问题

- ✓ 男孩喜欢看比赛
- ✓ 女孩喜欢看电影
- ✓ 两人倾向于周末分道扬镳…

不同准则下的最优策略

		囚犯B	
		坦白	抗拒
囚犯A	坦白	(-5,-5)	(0,-10)
	抗拒	(-10,0)	(-1,-1)

- **社会福利：** <抗拒，抗拒>
- **帕里托优：** 除了<坦白，坦白>之外的其他情况
- **纳什均衡：** <坦白，坦白>
- **优超：** <坦白，坦白>

非共享支付的博弈

(A,B)	b_1	b_2
a_1	(0, 0)	(1, 1)
a_2	(2, 2)	(3, 3)

Original Game Matrix

值表在多智能体间共享，为公共知识

✓透露太多信息，不安全

✓适用范围有限，分布决策环境中难适用

✓空间复杂度高

信息分布环境下的博弈形式

Distributed Game Matrix

(A,B)	b_1	b_2
a_1	(0, ?)	(1, ?)
a_2	(2, ?)	(3, ?)

(A,B)	b_1	b_2
a_1	(?, 0)	(?, 1)
a_2	(?, 2)	(?, 3)

非共享支付的博弈

分布式决策下的博弈，绝对理性的纳什均衡未必最适用

(D,D)比(C,C)更有利

(D,D)比(C,C)更有利

智能体A

A	Confess	Deny
Confess	-9	0
Deny	-10	-1

智能体B

B	Confess	Deny
Confess	-9	-10
Deny	0	-1

策略组(*D*,*D*) 帕里托优超(Pareto dominates)于纳什均衡策略(*C*,*C*)。

一个有趣的博弈

(A,B)	b_1	b_2	b_3
a_1	**(20,40)**	(4,22)	**(29,30)**
a_2	(18,9)	**(36,19)**	(7,4)
a_3	(17,26)	(15,38)	**(27,38)**

$(\boldsymbol{a_1}, \boldsymbol{b_3})$: $(\mathbf{29}, \mathbf{30})$

$(\boldsymbol{a_3}, \boldsymbol{b_3})$: $(\mathbf{27}, \mathbf{38})$

分布式决策下的博弈矩阵

A	b_1	b_2	b_3
a_1	20	4	29
a_2	18	36	7
a_3	17	15	27

智能体A

B	b_1	b_2	b_3
a_1	40	22	30
a_2	9	19	4
a_3	26	38	38

智能体B

(a_1,b_3) (a_3,b_3) (a_1,b_1)

(a_1,b_3) (a_3,b_3) (a_2,b_2)

大 纲

极小极大搜索

α-β剪枝

帕里托优和纳什均衡

协商(投票，拍卖和谈判)

协商(投票)

投票机制

- ✓ Agents给予一个投票机输入，投票机结果作为Agents的解决方案
- ✓ 设A为所有Agent的集合，O为所有投票结果的集合
- ✓ 一个agent i的投票结果可以被描述为$\succ_i \subseteq O \times O$

阿罗不可能定理！

投票机制的六原则

- 对所有可能的输入组合，都存在一个社会偏序$\succ^*$
- $\succ^*$对任意候选人的二元组$o, o' \in O$都有定义
- $\succ^*$在O上是非对称且传递的
- 结果满足帕利脱最优，即$\forall i \in A, o \succ_i o'$, 则$o \succ^* o'$
- 投票方案对不相关的候选人是独立的
- 没有Agent可以是独裁的

多数投票

陈水扁39%

宋楚瑜36%

连战为22%

- **投票协议**
  - ✓ 所有候选人同时进行比较，得票最高者获胜
  - ✓ 不满足无关方案独立原则
- 例：60%的agents支持$a \succ b$，40%的agents支持$b \succ a$

a胜出！

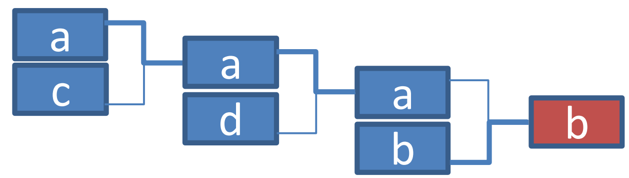
- 引入候选人c，30%的agents支持$a \succ c \succ b$， 30%的agents支持$c \succ a \succ b$， 40%的agents支持$b \succ a \succ$ c

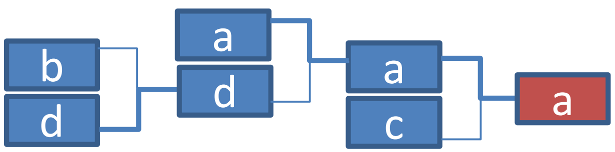
b胜出！

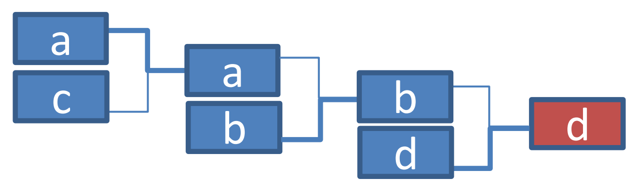
二叉投票

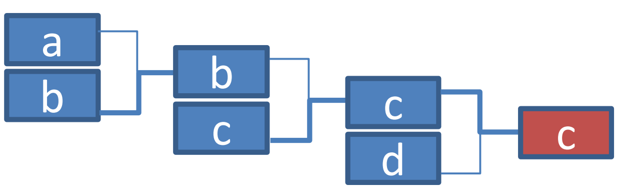
- 投票协议
  - ✓ 候选人成对PK
  - ✓ 胜者和其他候选人继续PK
  - ✓ 败者淘汰
- 不满足无关方案独立原则！
- 投票的结果依赖于比较的次序！

- 35% agents 支持$c \succ d \succ b \succ a$
- 33% agents 支持$a \succ c \succ d \succ b$
- 32% agents 支持$b \succ a \succ c \succ d$









计分投票

- 投票协议
  - ✓ 设置一个分值$|O|$
  - ✓ 排名第一的得$|O|$分，排名第二的得$|O|$-1分，依次类推
  - ✓ 累加所有候选人的得分
- 不满足无关方案独立原则！
- 投票的结果依赖于分值！

计分投票

Agent	投票	a b c d 得分
1	a > b > c > d	4 3 2 1
2	b > c > d > a	1 4 3 2
3	c > d > a > b	2 1 4 3
4	a > b > c > d	4 3 2 1
5	b > c > d > d	1 4 3 2
6	c > d > a > b	2 1 4 3
7	a > b > c > d	4 3 2 1
记分结果	c 20, b 19, a 18, d 13	
记分结果（排除d）	a 15, b 14, c 13	

协商(拍卖)

拍卖机制

- ✓ 投票机制的设计，其目的是使结果帕里托优
- ✓ 拍卖机制的设计，其目的是使拍卖者增加自己的利益

- ✓ 拍卖者尽可能高价卖出物品！
- ✓ 竞拍者尽可能使自己以低价获得物品！

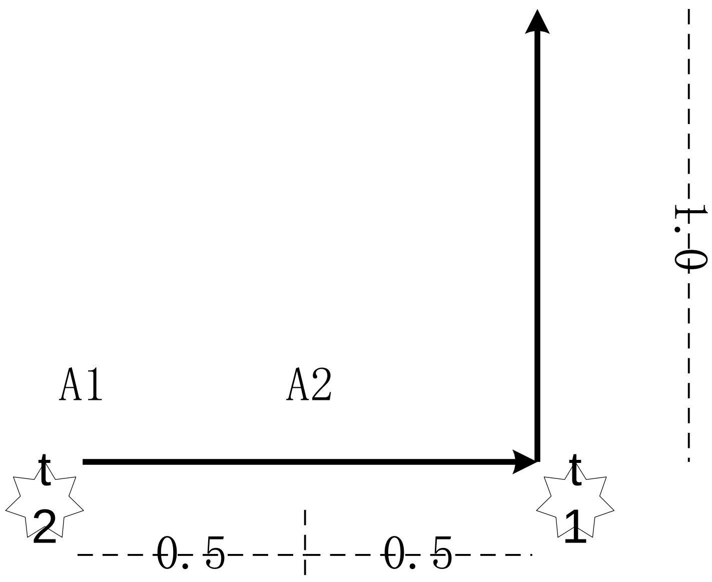
拍卖机制

☐ 拍卖协议

- ✓ 英格兰拍卖 first-price open-cry
- ✓ 密封拍卖 first-price sealed-bid
- ✓ 荷兰式拍卖
- ✓ Vickery拍卖second-price sealed-bid

通过博弈论计算其NASH均衡！

相关资源的拍卖



$c_1(t_1)=2$

$c_1(t_2)=1$

$c_1(t_1+t_2)=2$

$c_2(t_1)=1.5$

$c_2(t_2)=1.5$

$c_2(t_1+t_2)=2.5$

两个物流任务和两个骑手

相关资源的拍卖

- If a1 have t1, c1(t1+t2)-c1(t1)=2-2=0.   else c1(t2)=1
- If a2 have t1, c2(t1+t2)-c2(t1)=2.5-1.5=1   else c2(t2)=1.5

- So when a1 have t1, it bids t2 will get extra profit 1.5-0=1.5
- when a2 have t1, it bids t2 will get extra profit 1-1=0

- So when a1 bids t1, it will bid c1(t1)-extra profit=2-1.5=0.5
- when a2 bids t1, it will bid c2(t1)-extra profit=1.5-0=1.5

- A1 wins!

协商(谈判)

- 谈判机制
  - ✓ 公理谈判机制

    不变性、对称性、无关性、帕里托优

  - ✓ 策略谈判机制

    物品的折扣因素、谈判的代价因素

公理谈判机制

谈判机制

- ✓ 公理谈判机制

不变性、对称性、无关性、帕里托优

U1/U2?

策略谈判机制

谈判机制

- ✓ 策略谈判机制

有限轮博弈，假设折扣=0.9，输出依赖于谈判的轮次#

物品的折扣因素、谈判的代价因素

Round	I’ s share	2’ s share	Total value	Offerer
… n-3	… 0.819	… 0.181	… 0.9^{n-4}	… 2
n-2	0.91	0.09	0.9^{n-3}	1
n-1	0.9	0.1	0.9^{n-2}	2
n	1	0	0.9^{n-1}	1

策略谈判机制

◘ 谈判机制

无限轮博弈，假设Agent1折扣为δ1，Agent1折扣为δ2

✓ 策略谈判机制

Round	1's share	2's share	Offerer
…	…	…	…
t-2	$1-\delta_2(1-\delta_1\pi_1)$		1
t-1		$1-\delta_1\pi_1$	2
t	π_1		1
…	…	…	…

纳什均衡

$$\pi_1 = 1-\delta_2\left(1-\delta_1\pi_1\right) \Leftrightarrow \pi_1 = \frac{1-\delta_2}{1-\delta_1\delta_2}$$

策略谈判机制

- 谈判的代价(Bargaining Costs)
  - Agent 1 pays c1, agent 2 pays c2.
- Time t: 1 get p, 2 get 1-p;
- Time t-1: 2 thinks: 1 get p+c2, 2 get 1-p-c2;
- Time t-2: 1 thinks: 1get p+c2-c1, 2 get 1-p-c2+c1;
- Time t-2k: 1 thinks: 2 get 1-p-k(c2-c1).

> c1=c2: Any split is in Nash-equilibrium.
>
> c1<c2: Agent 1 gets all.
>
> c1>c2: Agent 1 gets c2, agent 2 gets 1-c2.

思考和讨论

1. 极小极大过程
2. ALPHA-BETA剪枝
3. 帕里托优和NASH均衡的区别
4. 进一步阅读拍卖和谈判理论

谢 谢 ！