

对抗搜索

Outline

- ❖ 博弈的概念
- ❖ 极大极小法
- ❖ α - β 剪枝技术

一、博弈的概念

❖ 下棋、打牌、战争等一类竞争性智能活动称为**博弈**

◆ 一般来说，博弈包括一系列的**玩家**、**动作**、**策略**和最终的**报酬**

❖ **玩家**：参与博弈的理性主体。如

- ◆ 拍卖中的竞标者
- ◆ 玩石头剪刀布的玩家
- ◆ 参加选举的政治家等



❖ **报酬**：所有玩家在达到某种结果时得到的回报

- ◆ 可以是积极的，也可以是消极的
- ◆ 每个主体都是自私的，希望得到最大化的报酬

“深蓝”

1997年5月11日，IBM开发的“深蓝”击败了国际象棋冠军卡斯帕罗夫。

卡氏何许人也？

- 1980年他获得世界少年组冠军
- 1982年他并列夺得苏联冠军
- 1985年22岁的卡斯帕罗夫成为**历史上最年轻**的国际象棋冠军。积分是2849，这一分数是有史以来最高分，远远领先于第二位的克拉姆尼克的2770



1997年纽约，与IBM深蓝电脑终局对弈

一个时代的结束

❖ 围棋被认为人类「对抗」计算机的最后壁垒

❖ 2016.3月：AlphaGo 4:1 战胜李世石

感到惊讶——无话可说——令人绝望

❖ 棋盘游戏作为AI进步衡量标尺的时代宣告结束



游戏AI的发展历程

The Development of Gaming AI

非完全信息游戏难度比较

Difficulty of imperfect information games

游戏	信息集数目	信息集平均大小
两人德州扑克 (限注)	10^{14}	10^3
两人德州扑克 (无限注)	10^{162}	10^3
桥牌	10^{67}	10^{15}
麻将	10^{121}	10^{48}



博弈论中的纳什均衡

❖ 囚徒困境



Alan

Ben


		Silent	Confess
Alan	Silent	A:-1, B:-1	A:-15, B:0
	Confess	A:0, B:-15	A:-10, B:-10

囚徒困境



		Ben confesses	
		Silent	Confess
Alan	Silent	A: -1, B: -1	A: -15, B: 0
	Confess	A: 0, B: -15	A: -10, B: -10

		Ben stays silent	
		Silent	Confess
Alan	Silent	A: -1, B: -1	A: -15, B: 0
	Confess	A: 0, B: -15	A: -10, B: -10



We go to prison
for 10 years because
we both confessed. If
we hadn't confessed,
we would each have
gone to jail for only
one year.

Yeah!
But if I told
you that I wouldn't
confess, you still would
have confessed to avoid
prison. Then I would
have gone to jail for 15
years. I'm glad I
confessed.

“二人零和、全信息、非偶然”博弈

- ❖ 对垒双方(A、B)轮流走步，结果只有三种：A胜B败、A败B胜、双方平局。二人获得分数的代数和必为零，称为“二人零和”。
- ❖ 对垒过程中任何一方都了解当前格局及过去的历史。
- ❖ 任何一方都要根据当前情况，分析得失，选取对自己最有利而对对方最不利的对策，而不存在“碰运气”的偶然因素。即双方都是很理智地决定自己的行动。



❖ 以某一方的立场把双人完备信息博弈过程用图表示出来，就得到一棵与或树。描述博弈过程的与或树称为**博弈树**。

❖ 博弈树的特点：

- ◆ 博弈的初始格局是初始节点。
- ◆ 在博弈树中，“或”节点和“与”节点逐层交替出现。自己一方扩展的节点之间是“或”关系，对方扩展的节点之间是“与”关系。双方轮流地扩展节点。
- ◆ 所有能使自己获胜的终局都是本原问题，相应的节点是可解节点；所有使对方获胜的终局都是不可解节点。

二、极大极小方法（Minimax algorithm）

- ❖ 设博弈的双方中一方为**A**,另一方为**B**。然后为其中的一方(例如**A**)寻找一个最优行动方案。
 - ◆ 考虑每一方案实施后对方可能采取的所有行动, 并计算可能的得分。定义一个估价函数, 用来估算当前博弈树端节点的得分。估算出的得分称为**静态估值**。
 - ◆ 由端节点估值推算出父节点的得分——**倒推值**
 - “或”节点, 选其子节点中最大的得分作为父节点的得分, 立足最好;
 - “与”节点, 选其子节点中最小的得分作为父节点的得分, 立足最坏。
 - ◆ 如果一个行动方案能获得**较大的倒推值**, 则它就是当前最好的行动方案。

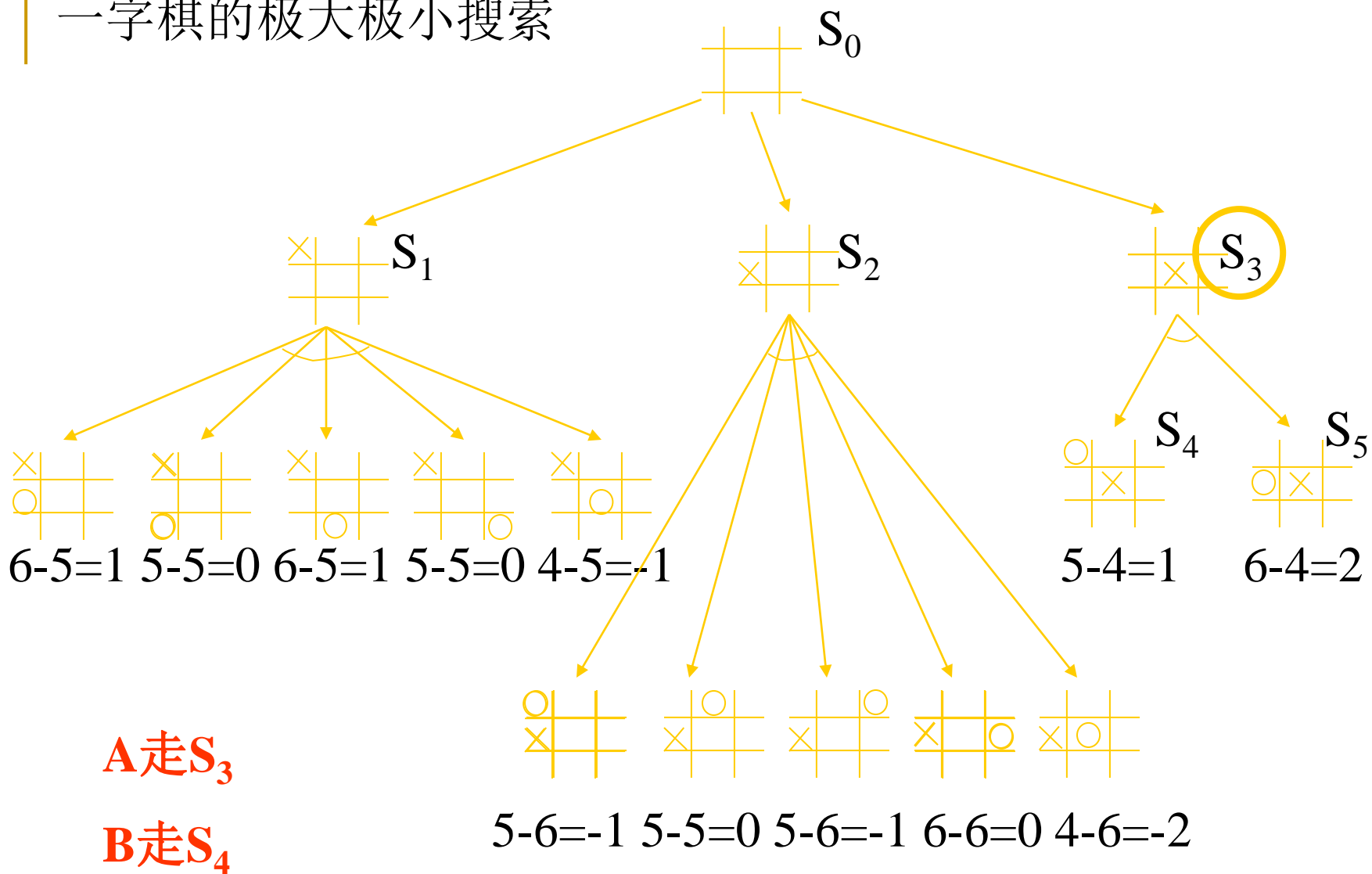
例：一字棋游戏

设有3x3的九个空格, 由A, B二人对弈, 轮到谁走棋谁就往空格上放一只自己的棋子, 先使自己的棋子构成“三子成一线”者胜利。

解：设棋局为 P , 定义估价函数为 $e(P)$:

- ✓ 若 P 是A必胜的棋局, 则 $e(P)=+\infty$;
 - ✓ 若 P 是B必胜的棋局, 则 $e(P)=-\infty$;
 - ✓ 若 P 是胜负未定的棋局, 则 $e(P)=e(+P)-e(-P)$, 其中 $e(+P)$ 表示棋局 P 上有可能使a成为三子成一线的数目; $e(-P)$ 表示棋局 P 上有可能使b成为三子成一线的数目。
- 具有对称性的棋盘认为是同一棋盘。

一字棋的极大极小搜索

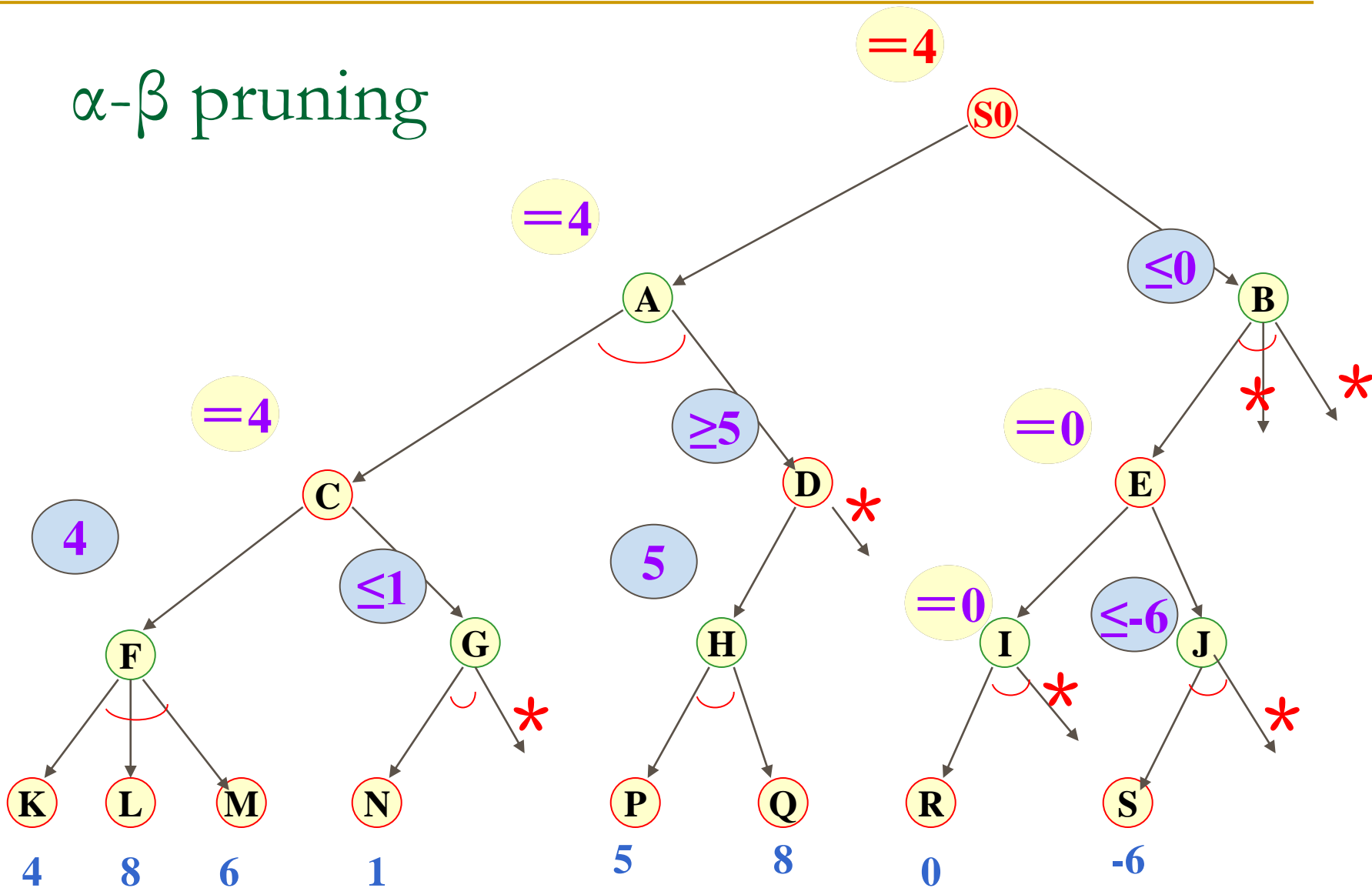


三、 α - β 剪枝 (α - β pruning)

- ❖ 极大极小搜索法实际上是先生成一棵博弈树，然后再计算其倒推值，这样做的缺点是效率较低。
- ❖ α - β 剪枝技术的基本思想:边生成博弈树边计算评估各节点的倒推值，并且根据评估出的倒推值范围，及时停止扩展那些已无必要再扩展的子节点，即相当于剪去了博弈树上的一些分枝，从而节约了机器开销，提高了搜索效率。

- ❖ 对于 “或” 节点，为了剪除某些分枝，取其子节点中的**最大倒推值**作为当前下界的参考，称此值为 **α 值**；
- ❖ 对于 “与” 节点，应取其子节点中的**最小倒推值**作为当前上界的参考，称此值为 **β 值**。
- ❖ 剪枝技术的一般规律：
 - ◆ 任何或节点 n 的 α 值，如果不能降低其父节点的 β 值，则对节点 n 以下的分支可停止搜索，并使 n 的倒推值为 α 值。
 - ◆ 任何与节点 n 的 β 值，如果不能升高其父节点的 α 值，则对节点 n 以下的分支可停止搜索，并使 n 的倒推值为 β 值。
 - 对于一个或节点，如果估值最高的节点最先生成，或者对于一个与节点，估值最低的子节点最先生成，则被剪的节点数最多，搜索的效率最高，称为**最优 α - β 剪枝**。

α - β pruning



AlphaGo

❖ 组成部分:

- ◆ 1. 走棋网络（Policy Network），给定当前局面，预测/采样下一步的走棋。
- ◆ 2. 快速走子（Fast rollout），目标和1一样，但在适当牺牲走棋质量的条件下，速度要比1快1000倍。
- ◆ 3. 估值网络（Value Network），给定当前局面，估计是白胜还是黑胜。
- ◆ 蒙特卡罗树搜索（Monte Carlo Tree Search, MCTS），把以上这三个部分连起来，形成一个完整的系统。

AlphaGo/Zero 的核心组件

- ❖ 蒙特卡洛树搜索——内含用于树遍历的 PUCT (Upper Confidence Bound applied to trees) 函数的某些变体
- ❖ 残差卷积神经网络——其中的策略和价值网络被用于评估棋局，以进行下一步落子位置的先验概率估算
- ❖ 强化学习——通过自我对弈进行神经网络训练

课程部分材料来自他人
和网络，仅限教学使用
，请勿传播，谢谢！