

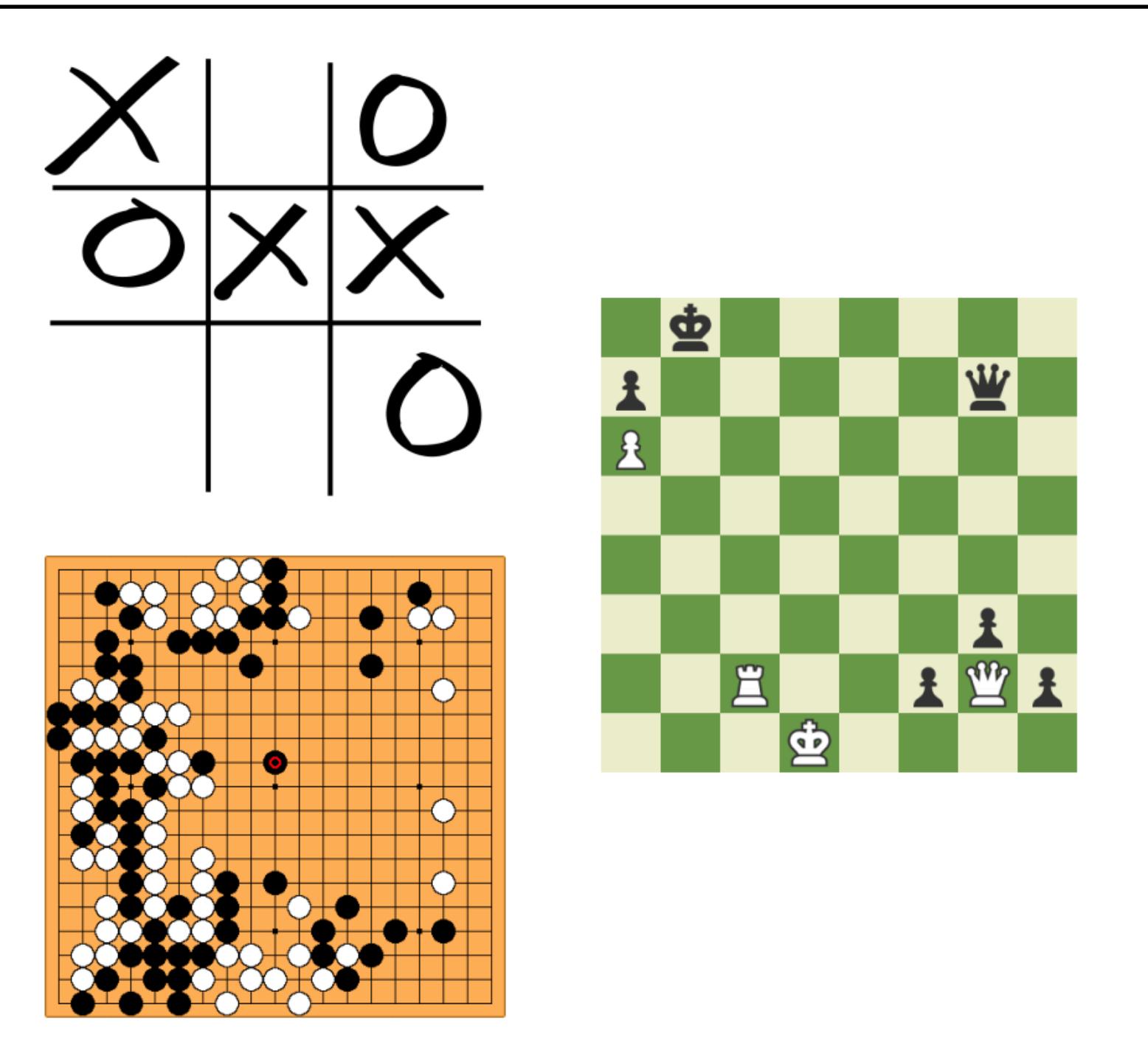
浅谈蒙特卡洛树搜索

五子棋AI

胡骁 2022.06.29

传统决策树方法

完全信息博弈&最大-最小树搜索算法



完全信息博弈：是指在任意时刻，双方玩家都知道游戏的全部状态（“完全信息”）。

最大-最小树搜索算法：每次决策时，希望找到对方让我方陷入最坏情况的各种策略中的较好的策略。

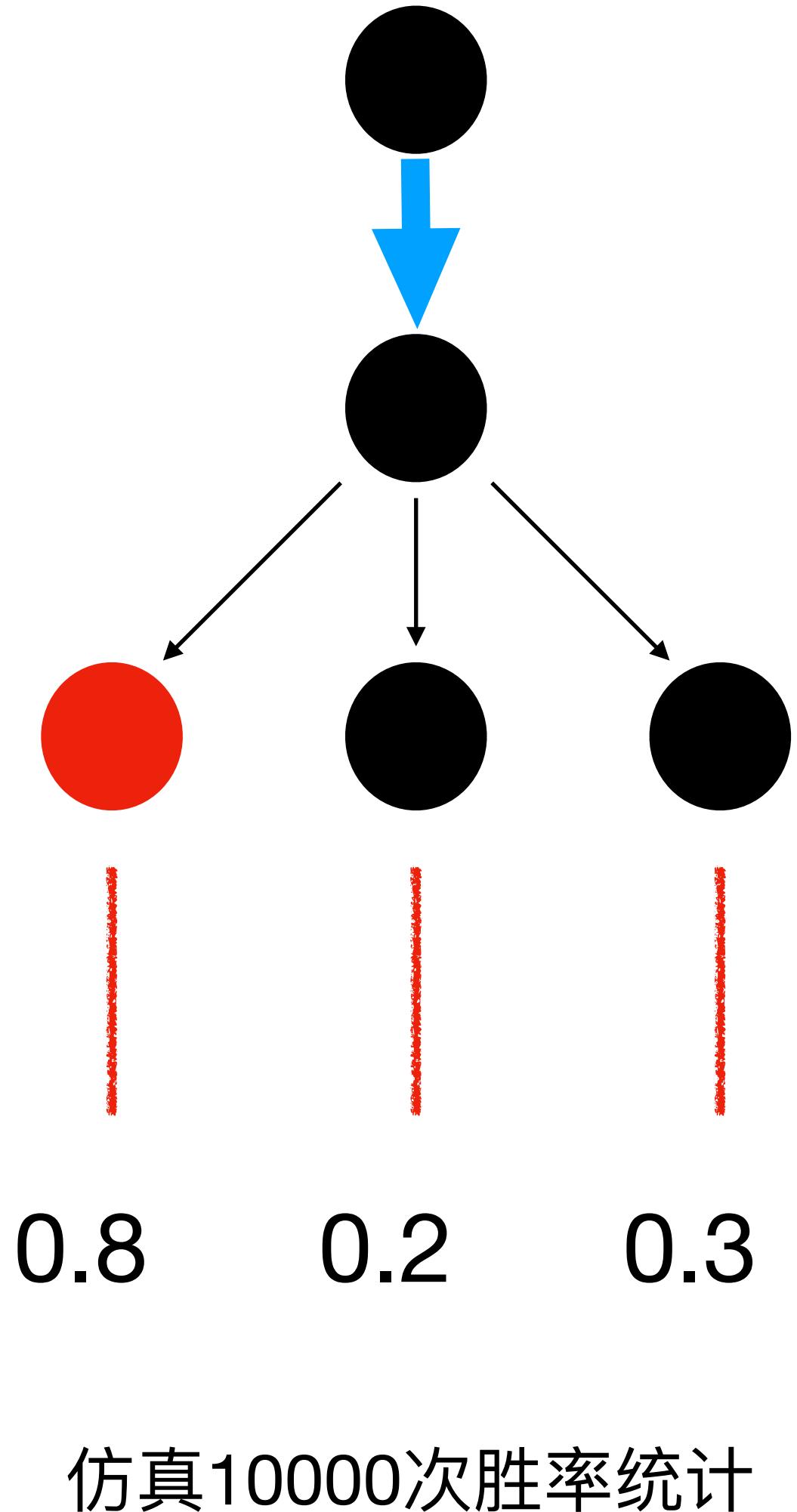
有一个好的heuristic的时候且决策链较短、分支数量较少的情况下比较高效。

评估策略

蒙特卡洛法&仿真

问题：当我们缺乏一个有效评估当前策略？

- 用随机策略决定下一步
- 从某个状态将游戏进行到终态后，标记胜负
- 重复这一过程
- 决策执行选择胜率最高的下一步

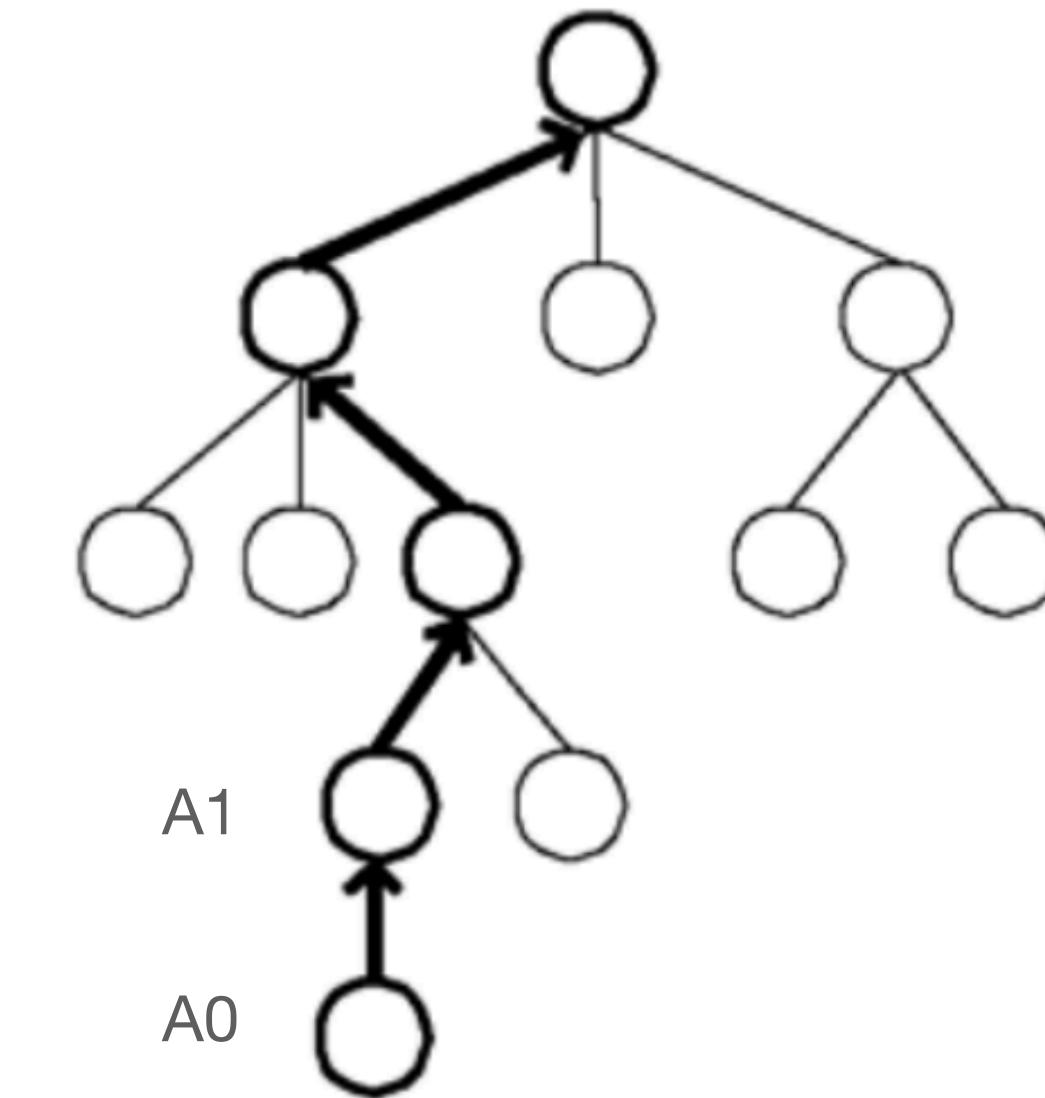


更精准的评价

马尔科夫链&反向传播

问题：游戏是一个连续决策的过程，每一步行动都影响了后续决策的可能性，由于仿真次数的限制，我们无法精准得到每一步的精准胜率，如何能使得胜率更加精准呢？

我们是否可以后续的仿真结果更新到前序节点—反向传播



A0和A1各做了100次仿真，A1的胜率是50/100，A0的胜率是70/100。于是将A0的仿真结果依次更新到他所经过了前序节点上，因此A1的胜率变成120/200

探索和利用的平衡

多臂老虎机

问题：完整的决策状态树异常巨大，我们是否能制定合理的策略只探索其中一部分节点？

经典问题：赌徒面前有N个老虎机,事先他不知道每台老虎机的真实盈利情况,他如何根据每次玩老虎机的结果来选择下次拉哪台或者是否停止赌博,来最大化自己的从头到尾的收益

A



$$P(A \text{ wins}) = 60\%$$

B



$$P(B \text{ wins}) = 55\%$$

C



$$P(C \text{ wins}) = 40\%$$

两难境地：

探索过多：牺牲掉我们本能多的利益

利用过多：陷入一个local-optimal

优化目标：minimize regret = loss

探索和利用的平衡

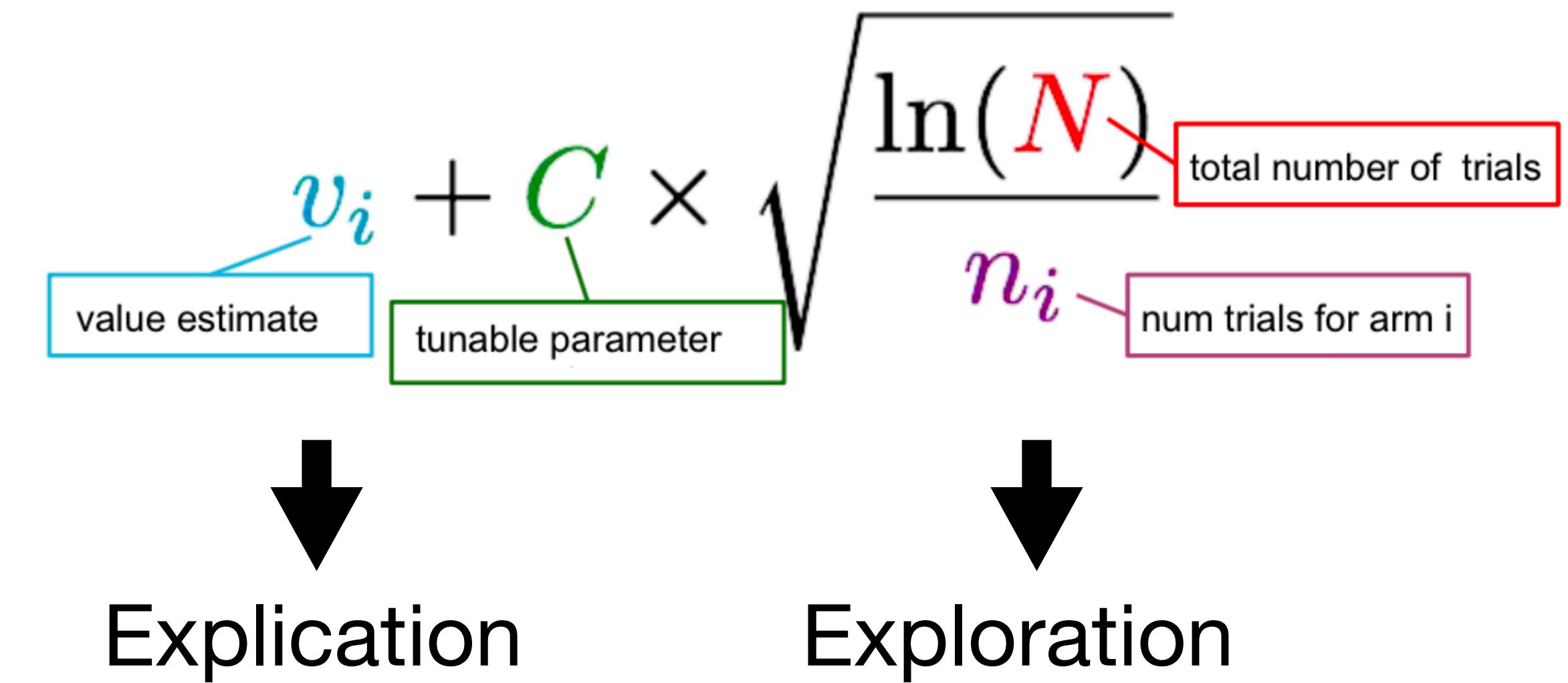
UCB&UCT

Upper Confidence Bound策略：

- 每一个老虎机都试一次
- 最大化UCB的公式

$$UCT = UCB + \text{反向传播}$$

UCT就作为合理的策略指导算法只探索其中一部分节点



常数C非常关键，它决定着我们最终是否能收敛到最优解。当 v_i 在 $[0,1]$, $C=2$ 被 Kocsis and Szepesvari 证明能收敛到最优解

蒙特卡洛树搜索

MCTS的行为描述

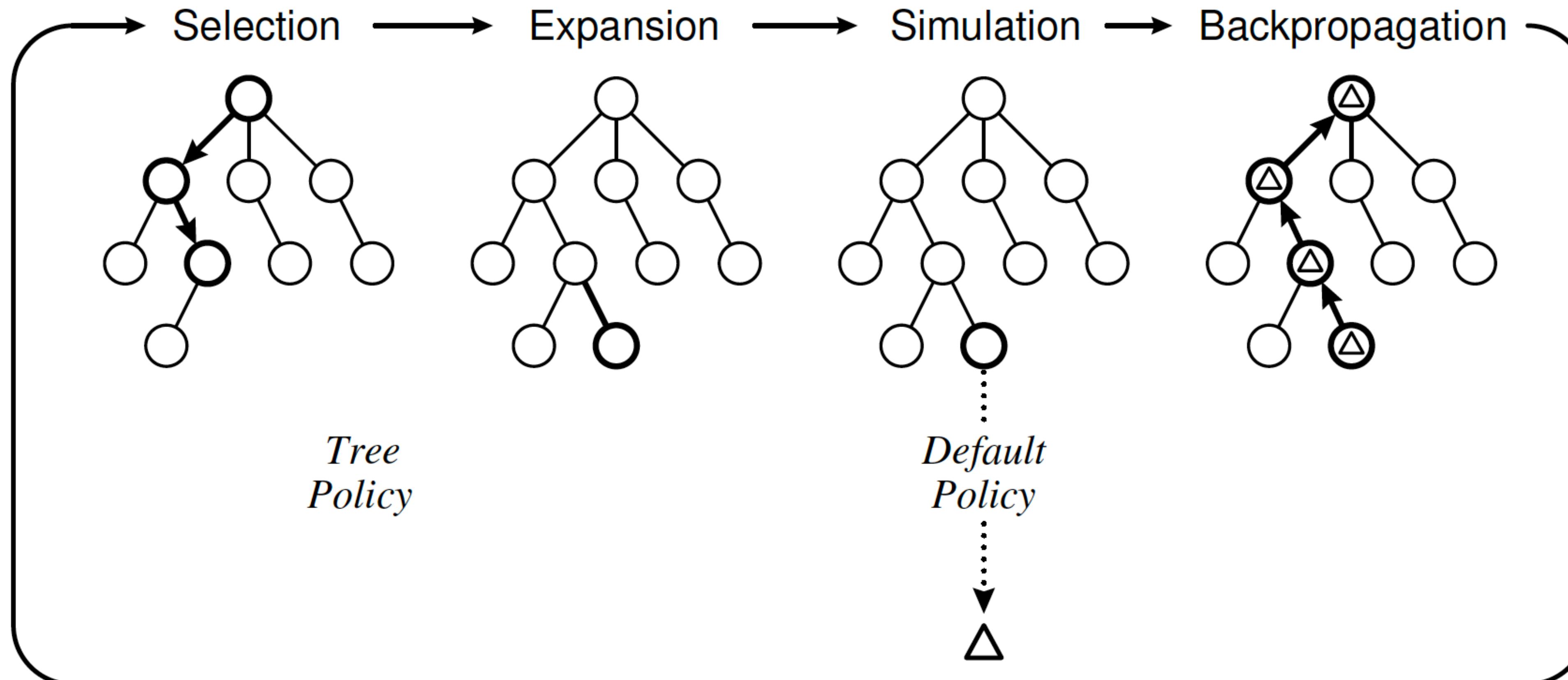


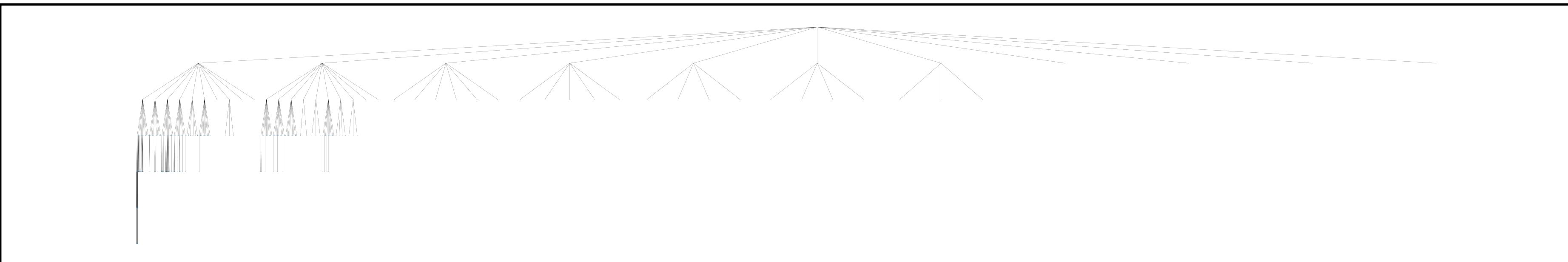
Fig. 2. One iteration of the general MCTS approach.

蒙特卡洛树搜索

MCTS的算法特性

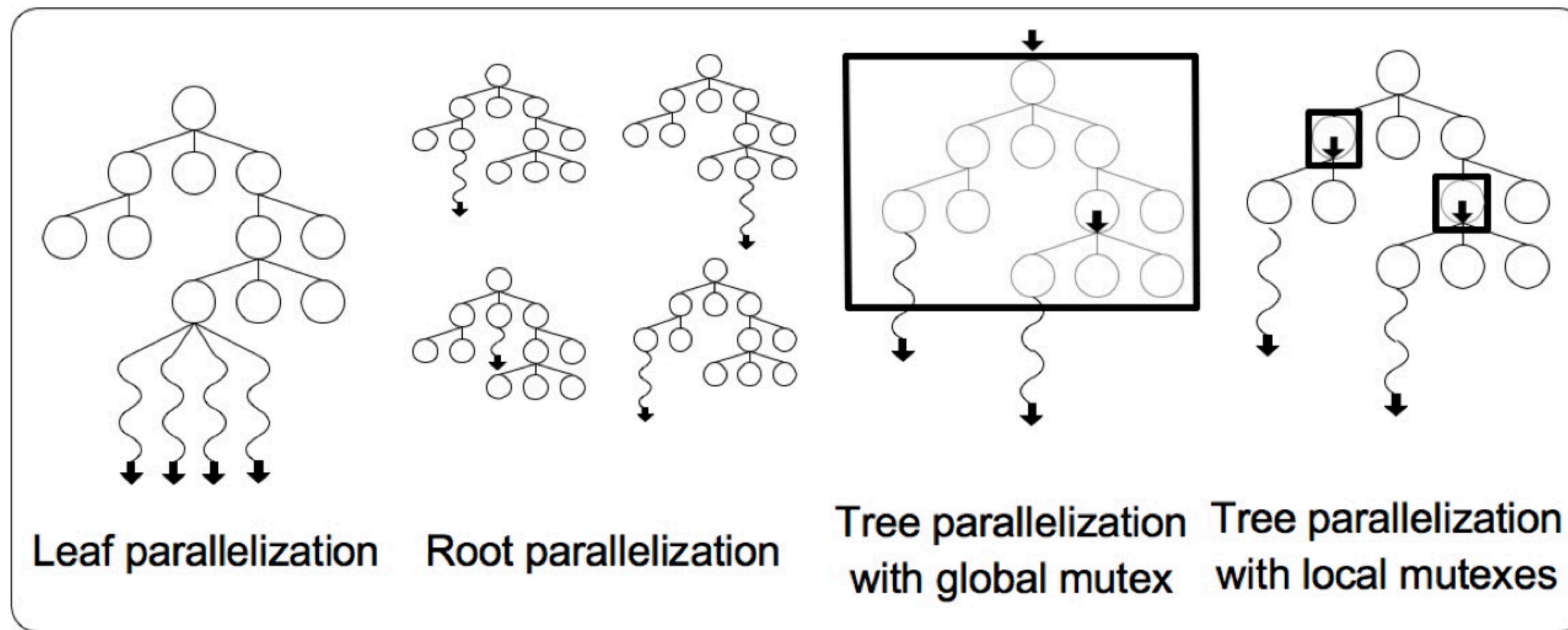
- 蒙特卡洛树构造了一颗基于数值的决策树（每个节点都有自己的Evaluation值），它是完整决策状态树的一部分
- 每个节点的Evaluation指导算法有方向探索最有价值的节点
- 每个节点的Reward是基于Simulation的结果

决策状态树视图 ->



蒙特卡洛树搜索

MCTS的并行



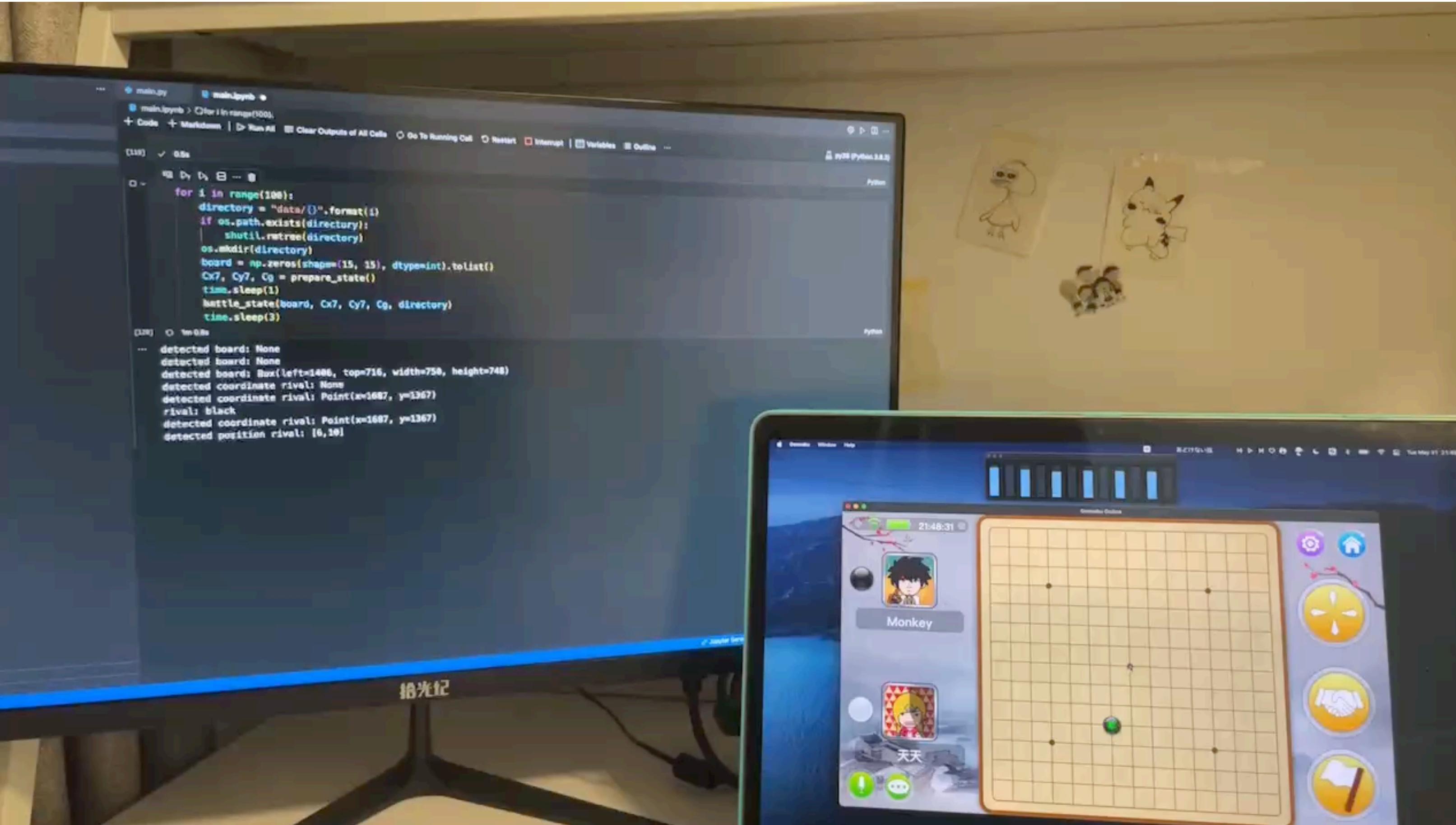
○ = Tree node

↓ = Thread location

□ = Locked memory section

{ = Tree-independent simulation

五子棋AI实战演示



项目代码地址：<https://github.com/Eggrolling-hu/MCTS>

MCTS+Learning

AlphaGo思路

AlphaGo Fan/Lee -> AlphaGo Zero -> AlphaZero

AlphaGo Fan/Lee:

估值网络和策略网络独立，策略网络预测人类的行为，
估值网络评估局势。训练集为高水平人类对局。

AlphaGo Zero:

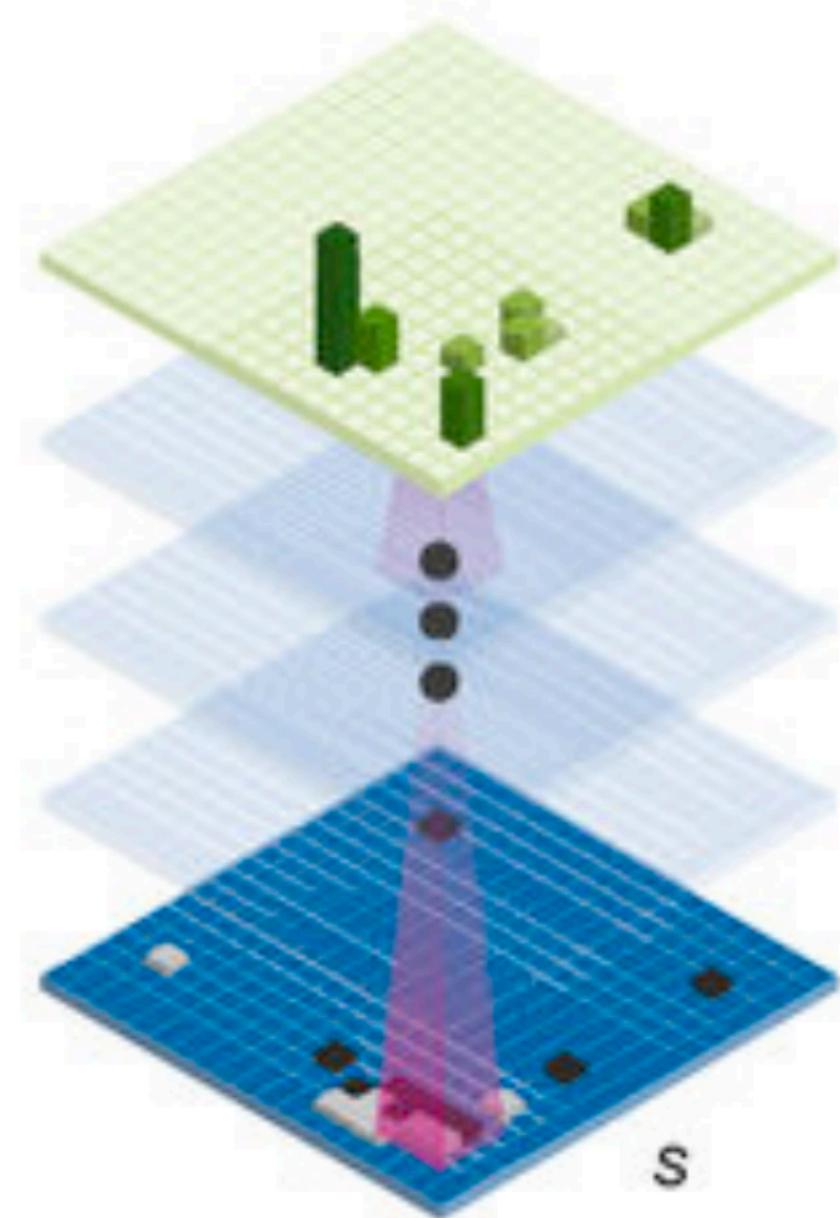
估值网络和策略网络合并，通过网络自我对抗产生训练
集。新网络与过去最好的网络对抗，最终留下一个。

AlphaZero:

通用性提升，只维护一个网络，新网络继承最新网络的
参数，引入部分随机性进行对抗。

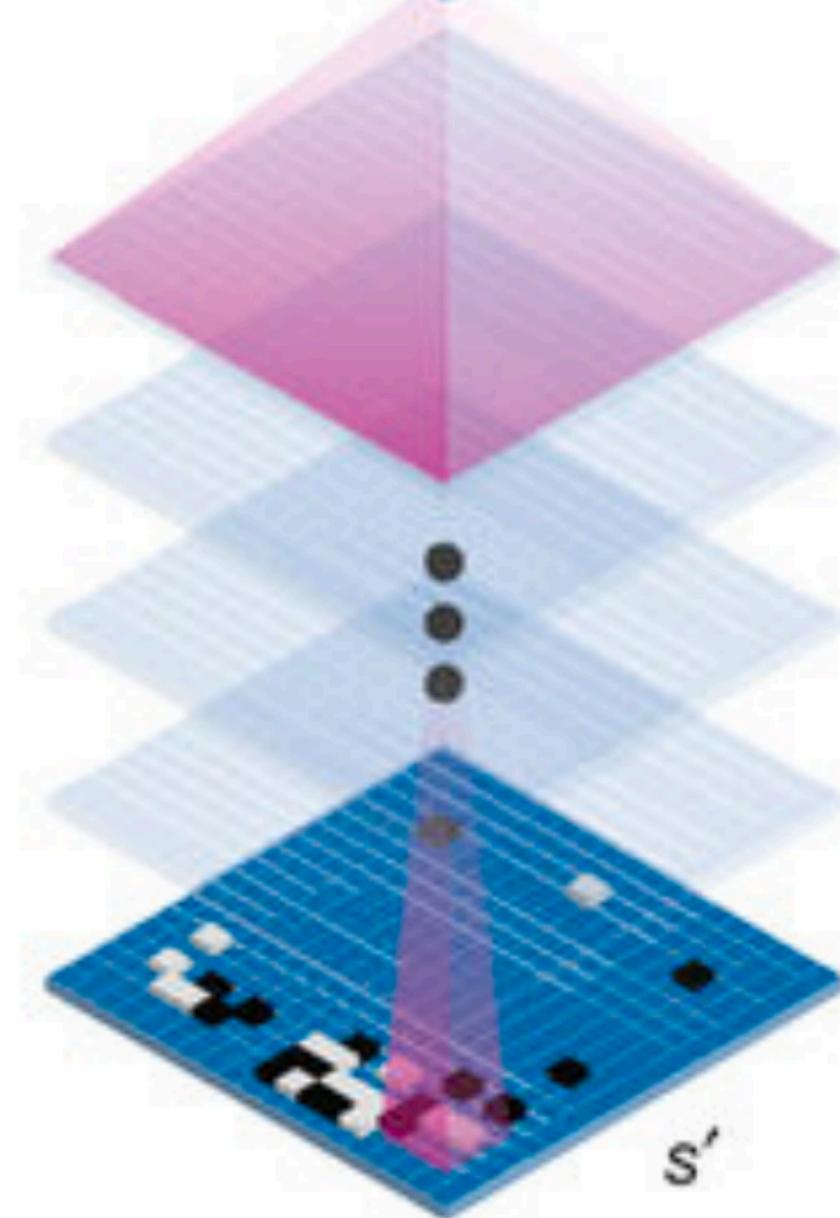
Policy network

$$p_{\sigma/\rho}(a|s)$$



Value network

$$v_{\theta}(s')$$



引用

- <https://u.cs.biu.ac.il/~krauss/advai2018/MCTS.pdf>
- [https://www.lri.fr/~sebag/Slides/InvitedTutorial CP12.pdf](https://www.lri.fr/~sebag/Slides/InvitedTutorial_CP12.pdf)
- <http://repository.essex.ac.uk/4117/1/MCTS-Survey.pdf>
- [https://discovery.ucl.ac.uk/id/eprint/10045895/1/agz unformatted nature.pdf](https://discovery.ucl.ac.uk/id/eprint/10045895/1/agz_unformatted_nature.pdf)
- <https://sci-hub.mksa.top/10.1126/science.aar6404>