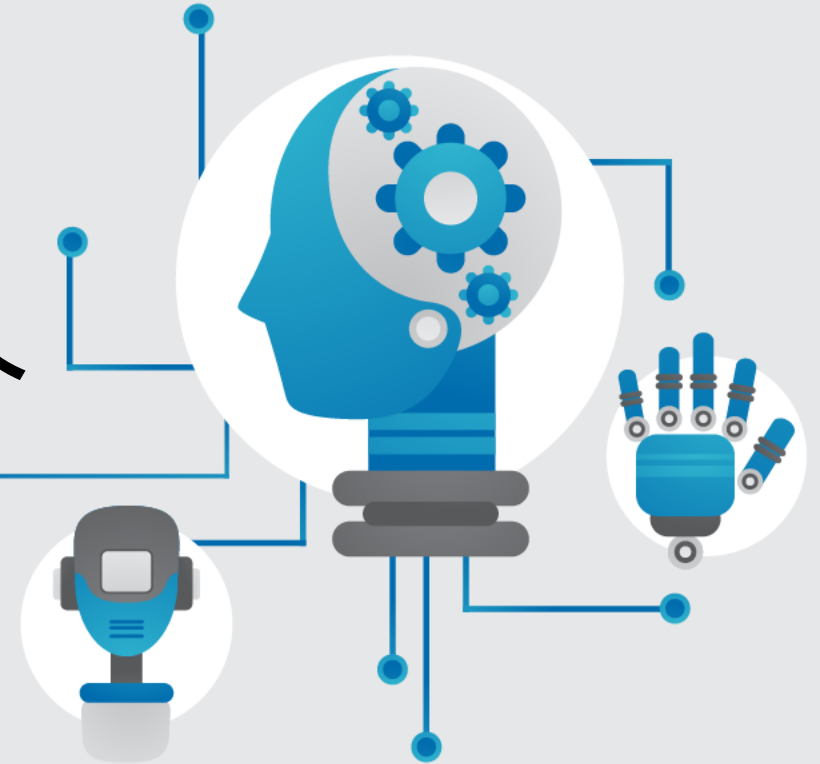


# 電腦圍棋Alpha Go簡介





# Google DeepMind

機器學習實務



## DeepMind

成立於2010年英國的人工智慧公司，2014年被Google收購，並於同年開始開發人工智慧圍棋軟體AlphaGo。

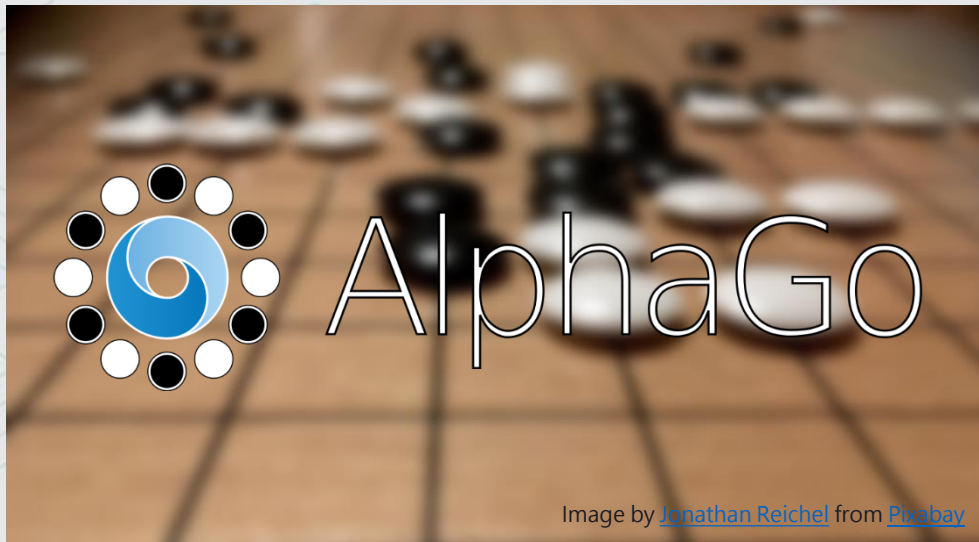


Image by [Jonathan Reichel](#) from [Pixabay](#)



# AlphaGo演算法

機器學習實務



蒙地卡羅樹搜尋 ( Monte Carlo based Tree Search, MCTS )

深層神經網路 ( Deep Neural Networks, DNN )

卷積神經網路 ( Convolutional Neural Networks, CNN )

強化學習 ( Reinforcement Learning, RL )

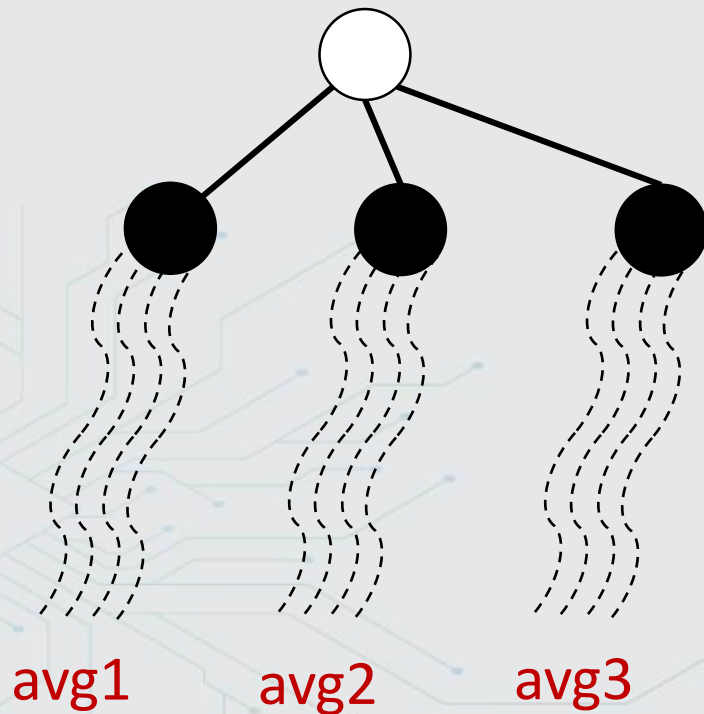


# 蒙地卡羅模擬搜尋 (原始想法)

機器學習實務



對於每一個可能的下一步，都進行大量的模擬，  
然後計算勝率，選擇最高勝率的著手落子。





# 多臂吃角子老虎機問題 ( Multi-armed bandit problem )

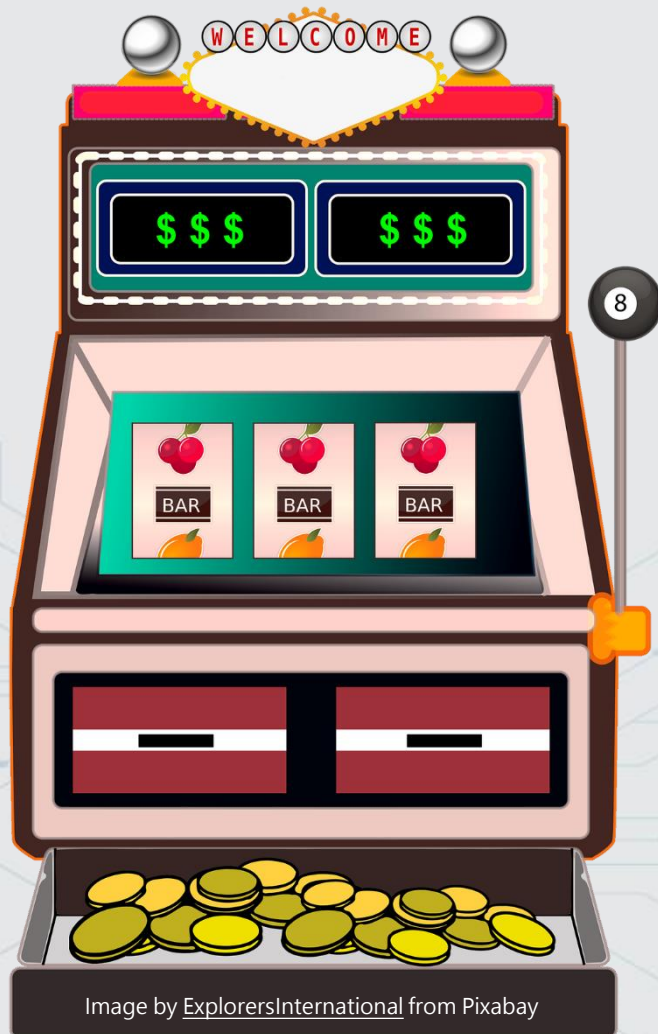


Image by [ExplorersInternational](#) from Pixabay

## [問題描述]

假設現在有 $K$ 台吃角子老虎遊戲機，每一台的報酬率不同。

假設你可以拉bar  $N$ 次，什麼是獲得最大報酬的策略？



# 解決策略



★ 主要策略：保持 **開發** 和 **探勘** 的平衡

Exploitation：開發目前最好的步

Exploration：探勘其它步，看是否可以獲得更好的報酬

★ 信賴上界 ( Upper Confidence Bound ; UCB )

$$UCB_i = \frac{\text{中獎次數}}{\text{遊玩次數}} + C \times \sqrt{\frac{\log(\text{總次數})}{\text{遊玩次數}}}$$

中獎次數越高，  
值越高

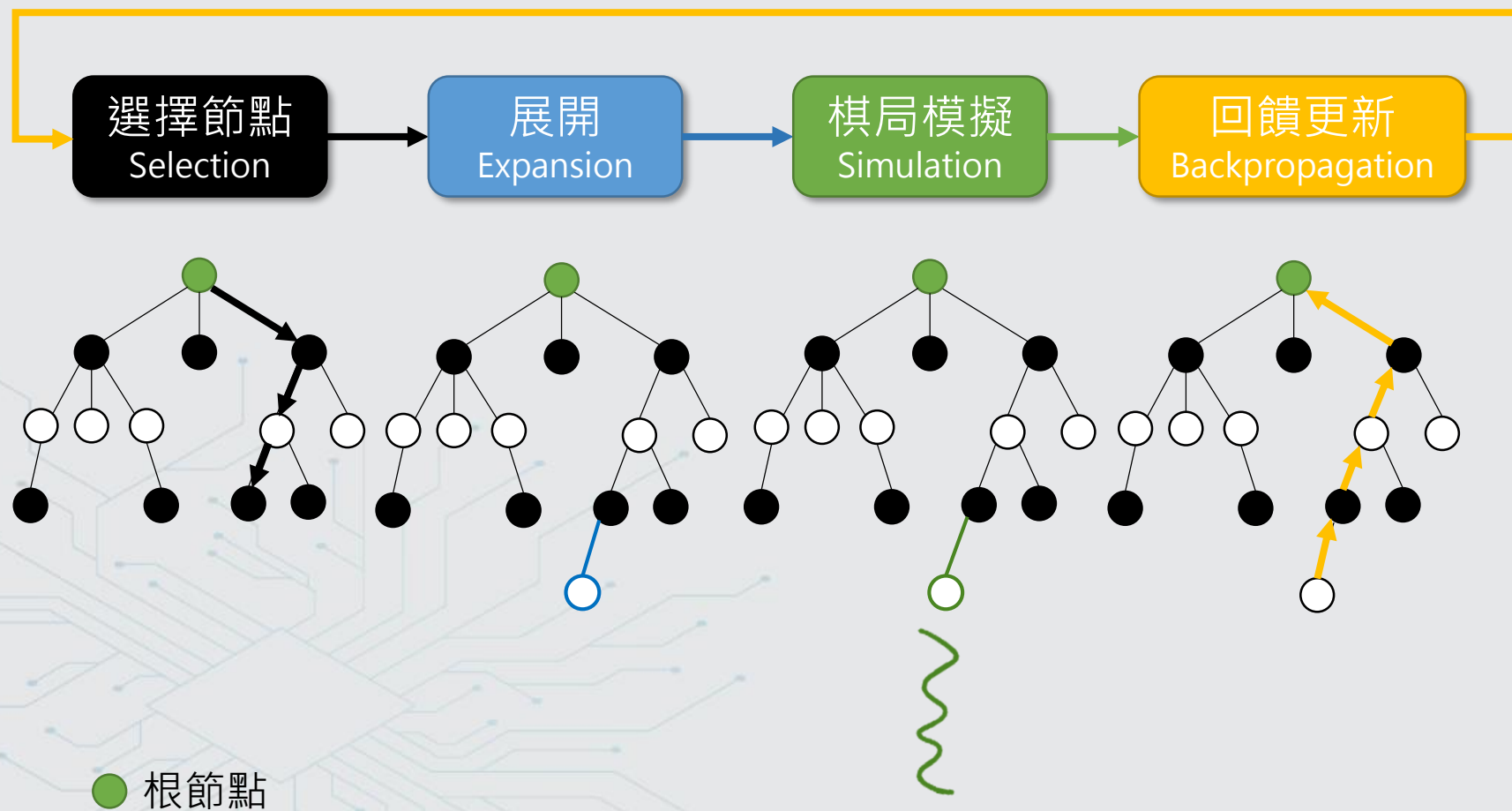
遊玩次數越高，  
值越低





# 蒙地卡羅樹搜尋演算法

機器學習實務





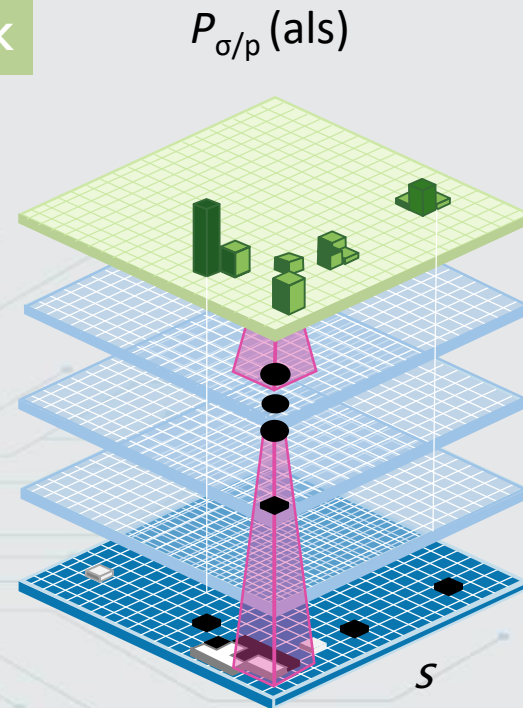
# 策略網路 ( policy network )

機器學習實務



為了降低遊戲樹的分支度，AlphaGo訓練一個策略網路（神經網路）來預測高手可能回應的著手。

Policy network





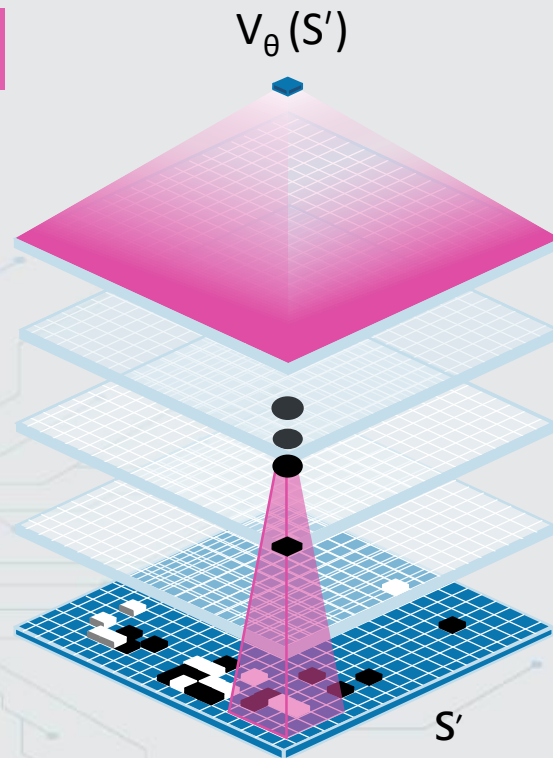


# 評估網路 ( value network )



為了讓自我模擬棋局能更像高手的對奕，AlphaGo訓練一個評估網路（神經網路）來引導棋局的模擬。

Value network





# 強化式學習 ( Reinforcement Learning )

機器學習實務



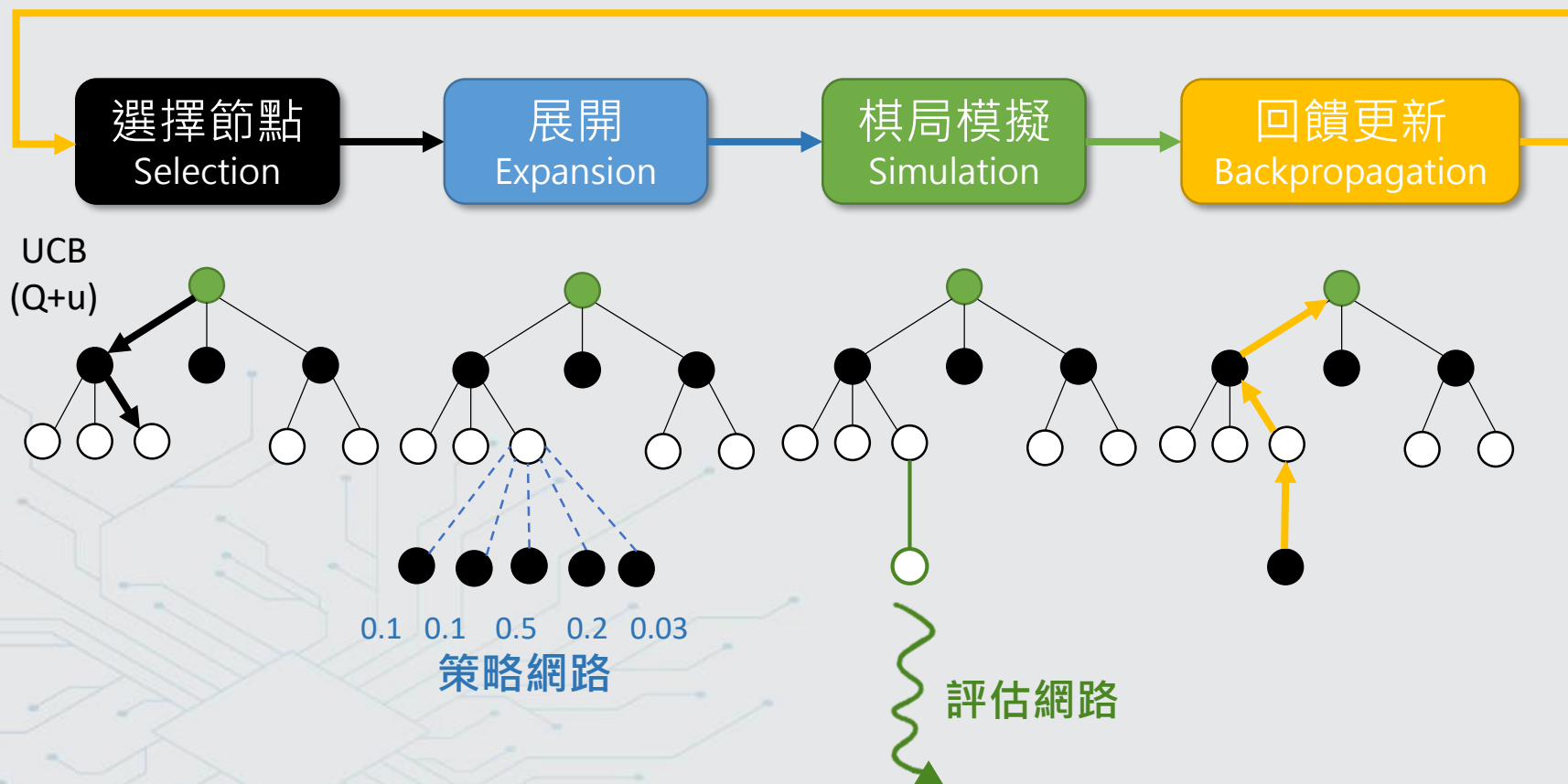
強化式學習的目的為代理人 ( Agent ) 對環境 ( Environment ) 進行一連串的動作，使得整個過程中可以獲得的累積獎勵 ( cumulative reward ) 最大化。





# AlphaGo運作圖解

機器學習實務





# AlphaGo



**AlphaGo is the first computer program to defeat a professional human Go player, the first to defeat a Go world champion, and is arguably the strongest Go player in history.**