ĐẠI HỌC QUỐC GIA THÀNH PHỐ HỒ CHÍ MINH
TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN

Cơ sở trí tuệ nhân tạo

Tìm kiếm đối kháng

Nguyễn Ngọc Đức

2021

Nội dung

Mô hình Markov

Markov

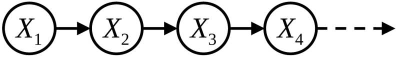
- Ta dùng từ ”Markov“ cho các quy trình mà hiện tại, tương lai và quá khứ độc lập với nhau
- Kết quả khi thực hiện một hành động chỉ phụ thuộc vào trạng thái hiện tại

Markov

- Ta dùng từ ”Markov“ cho các quy trình mà hiện tại, tương lai và quá khứ độc lập với nhau
- Kết quả khi thực hiện một hành động chỉ phụ thuộc vào trạng thái hiện tại
- Tương tự như bài toán tìm kiếm

Mô hình Markov

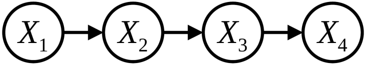
- Giá trị X trong một khoảng thời gian nhất định được gọi là trạng thái



$$P(X_1) \qquad P(X_t|X_{t-1})$$

- Tham số: Xác suất chuyển tiếp
- Giả định tính ổn định: Xác suất chuyển tiếp giống nhau ở mọi thời điểm

Hợp phân phối xác suất I



- Dựa trên quy tắc chuỗi, mỗi hợp phân phối xác suất X_1, X_2, X_3, X_4 có thể được viết thành:

$$\mathbb{P}(X_1, X_2, X_3, X_4) = \mathbb{P}(X_1)\mathbb{P}(X_2|X_1)\mathbb{P}(X_3|X_1, X_2)\mathbb{P}(X_4|X_1, X_2, X_3)$$

- Giả sử

$$X_3 \perp\!\!\!\perp X_1|X_2 \;\; \text{và} \;\; X_4 \perp\!\!\!\perp X_1, X_2|X_3$$

Hợp phân phối xác suất II

- Hợp phân phối xác suất:

$$\mathbb{P}(X_1, X_2, X_3, X_4) = \mathbb{P}(X_1)\mathbb{P}(X_2|X_1)\mathbb{P}(X_2|X_3)\mathbb{P}(X_4|X_3)$$

- Mọi hợp phân phối xác suất $X_1, X_2, \ldots, X_T$ có thể được viết dưới dạng:

$$\mathbb{P}(X_1, X_2, \ldots, X_T) = \mathbb{P}(X_1)\prod_{t=2}^{T}\mathbb{P}(X_t|X_1, X_2, \ldots, X_{t-1})$$

Hợp phân phối xác suất III

- Giả sử rằng với mọi t:

$$X_t \perp\!\!\!\perp X_1, \ldots, X_{t-2} | X_{t-1}$$

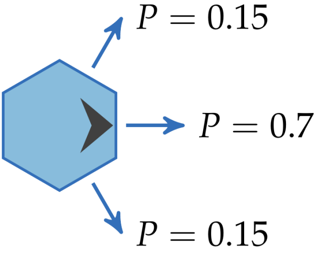
Ta có:

$$\begin{aligned}\mathbb{P}(X_1, X_2, \ldots, X_T) &= \mathbb{P}(X_1)\mathbb{P}(X_2|X_1)\mathbb{P}(X_3|X_2)\ldots\mathbb{P}(X_T|X_{T-1}) \\ &= \mathbb{P}(X_1)\prod_{t=2}^{T}\mathbb{P}(X_t|X_{t-1})\end{aligned}$$

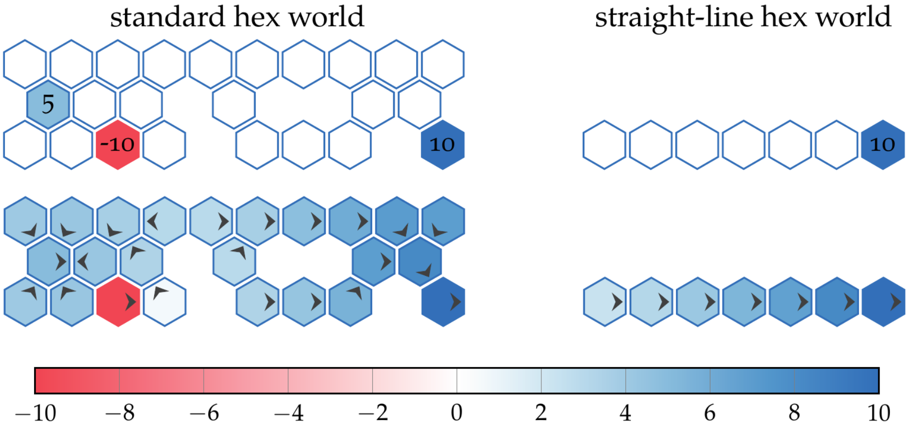
Tìm kiếm bất định

Ví dụ: Hex world I

- Mỗi ô là một trạng thái
- Di chuyển theo 6 hướng
- Nhiễu: Kết quả di chuyển ngẫu nhiên
- Tác tử nhận điểm thưởng với mỗi bước di chuyển

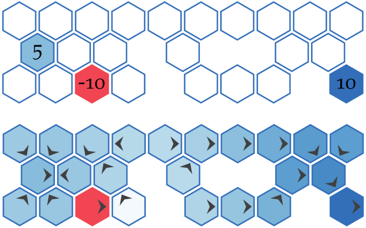


Ví dụ: Hex world II



Quy trình Markov

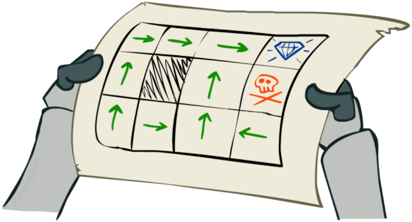
- Một quy trình Markov được định nghĩa dựa trên:
  1. Tập trạng thái $s \in S$
  2. Tập hành động $a \in A$
  3. Hàm chuyển dịch $T(s, a, s')$
  4. Hàm điểm thưởng $R(s, a, s')$
  5. Trạng thái bắt đầu
  6. Trạng thái kết thúc (có thể có hoặc không)



- MDP là một bài toán tìm kiếm bất định $\Rightarrow$ Expectimax

Chiến lược

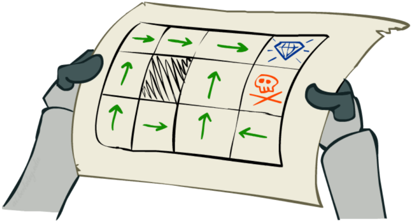
- Trong các bài toán tìm kiếm, chúng ta cần một kế hoạch tối ưu



- Với MDP:

Chiến lược

- Trong các bài toán tìm kiếm, chúng ta cần một kế hoạch tối ưu

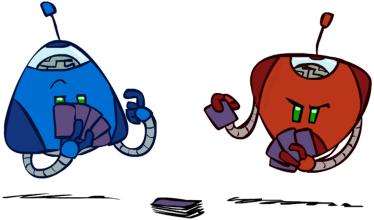


- Với MDP: Chúng ta cần một chiến lược tối ưu $\pi^* : S \rightarrow A$
  - Một chiến lược π đưa ra hành động tại một trạng thái cụ thể
  - Một chiến lược tối ưu sẽ tối đa hóa lợi ích kỳ vọng (expectimax)
  - Một chiến lược rõ ràng định nghĩa một tác tử

Trò chơi đối kháng

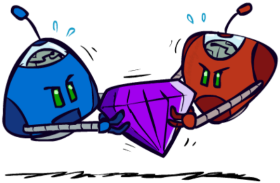
Trò chơi

- Trò chơi là một trong những nghiên cứu chính cho việc phát triển các chương trình thông minh
- **Đặc điểm:**
  1. **Môi trường đa nhân tố:** sự hiện diện của những người chơi với mục đích không đồng nhất
  2. **Bất định (uncertainty):** chiến lược di chuyển không rõ ràng



Các loại trò chơi

- Trò chơi có thể được phân loại dựa trên các yếu tố:
  1. Hành động
  2. Số lượng người chơi
  3. Đối kháng?
  4. Cụ thể (perfect information)?



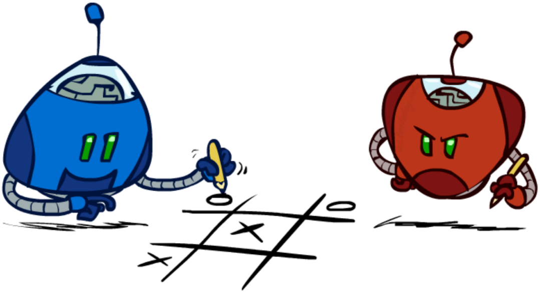
Trò chơi đối kháng 2 người I

Tập người chơi $Players = \{agent, opp\}$

1. s_0: trạng thái bắt đầu
2. $Action(s)$: Các hành động có thể thực hiện với trạng thái s
3. $Succ(s, a)$: trạng thái kết quả thực hiện hành động a ở trạng thái s
4. $IsEnd(s)$: kiểm tra trạng thái s có phải là trạng thái cuối (kết thúc trò chơi)
5. $Utility(s)$: điểm của agent ở trạng thái s
6. $Player(s) \in Players$: người chơi điều khiển trạng thái s

Trò chơi đối kháng 2 người II

- Có 2 đặc điểm chính:
  1. Tất cả mục tiêu đều nằm ở trạng thái cuối cùng
  2. Người chơi khác nhau điều khiển các trạng thái khác nhau



Cờ vua

- Players={đen, trắng}
- State(s): vị trí của các quân cờ
- Action(s): nước đi hợp lệ mà người chơi Player(s) có thể thực hiện
- IsEnd(s): kiểm tra trạng thái hiện tại là chiếu bí hay hòa
- Utility(s): $+\infty$ nếu đen thắng, $-\infty$ nếu trắng thắng, 0 nếu hòa

Thành tựu hiện nay I

- Cờ đam:
  - **Độ phức tạp:** $\approx 10^{18}$ nút
  - 1950: chương trình đầu tiên
  - 1991: Chinook đánh bại hoàn toàn nhà vô địch Marion Tinsley

Thành tựu hiện nay II

- Cờ vua:
  - **Độ phức tạp:** $b \approx 35, d \approx 100, 10^{154}$ nút
  - 1997: Deep Blue đánh bại Gary Kasparov trong ván đấu 6 trận.

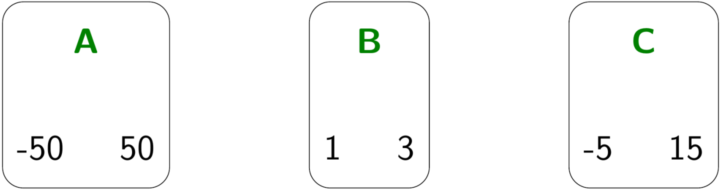
Thành tựu hiện nay III

- Cờ vây:
  - **Độ phức tạp:** $b \approx 361, d \approx 200, 10^{174}$ trạng thái bàn cờ
  - Không thể đoán trước được ngay cả ở giai đoạn thu quan
  - 2016: AlphaGo đánh bại kỳ thủ cửu đẳng Lee Sedol (4-1)

Cây trò chơi

Trò chơi 3 chiếc hộp

- Có 3 chiếc hộp, mỗi hộp chứa 2 con số (hình 1).
- Bạn chọn một chiếc hộp sau đó mình chọn một con số nằm trong hộp đó
- Nhiệm vụ của bạn là phải tối đa con số mà mình chọn

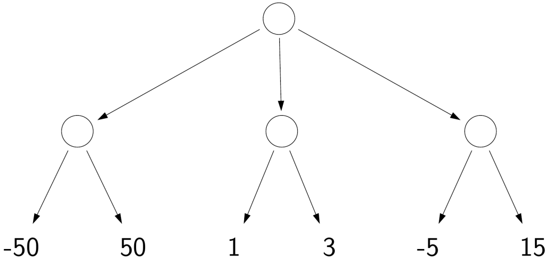


Hình 1: Ví dụ một trò chơi

Cây trò chơi

Cây trò chơi

- Mỗi nút là một điểm quyết định cho mỗi người chơi
- Mỗi đường đi tới nút lá là một kết quả của trò chơi



Expectimax

Chiến lược

- **Chiến lược xác định**: hành động người chơi p thực hiện ở trạng thái s

$$\pi(s) \in Action(s)$$

- **Chiến lược bất định**: xác suất người chơi p thực hiện hành động a ở trạng thái s

$$\pi(s, a) \in [0, 1]$$

Đánh giá trò chơi I

- **Lợi ích kỳ vọng** (giá trị của trò chơi)

$$\bigcirc \xrightarrow{\pi_{\text{agent}}} \bigcirc \xrightarrow{\pi_{\text{opp}}} \bigcirc \xrightarrow{\pi_{\text{agent}}} \bigcirc \ \dots$$

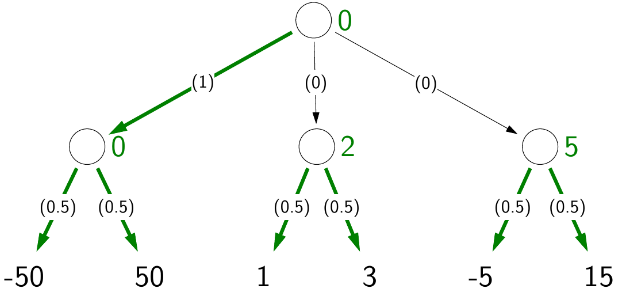
$$V_{eval}(s) = \begin{cases} Utility(s) & IsEnd(s) \\ \sum_{a \in Actions(s)} \pi_{agent}(s,a) V_{eval}(Succ(s,a)) & Player(s) = agent \\ \sum_{a \in Actions(s)} \pi_{opp}(s,a) V_{eval}(Succ(s,a)) & Player(s) = opp \end{cases}$$

1. Trò chơi kết thúc, lợi ích ở trạng thái cuối Utility(s)
2. Lượt của agent, dựa trên giá trị các successor trả về
3. Lượt của opp, tương tự agent

Đánh giá trò chơi II

$\pi_{\text{agent}}(s) = \text{A}$

$\pi_{\text{opp}}(s, a) = \frac{1}{2}$ for $a \in \text{Actions}(s)$

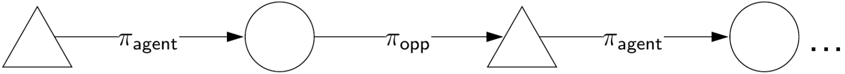


$$\boxed{V_{\text{eval}}(s_{\text{start}}) = 0}$$

Hình 2: Giá trị trò chơi

Expectimax I

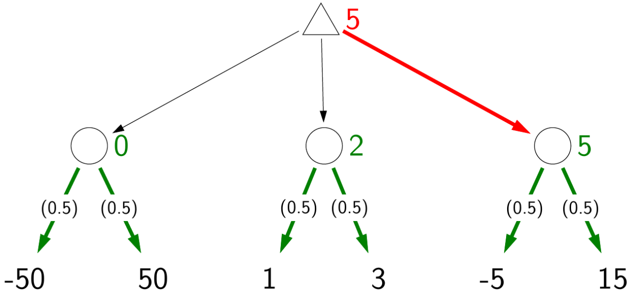
- **Giá trị expectimax** $V_{exptmax}(s)$ ở trạng thái s, là lợi ích tối đa của người chơi đạt được ở trạng thái s nếu biết trước chiến lược chơi của đối thủ



$$V_{exptmax}(s) = \begin{cases} Utility(s) & IsEnd(s) \\ \max_{a \in Actions(s)} V_{exptmax}(Succ(s,a)) & Player(s) = agent \\ \sum_{a \in Actions(s)} \pi_{opp}(s,a) V_{eval}(Succ(s,a)) & Player(s) = opp \end{cases}$$

Expectimax II

$$\pi_{\text{opp}}(s,a) = \frac{1}{2} \text{ for } a \in \text{Actions}(s)$$

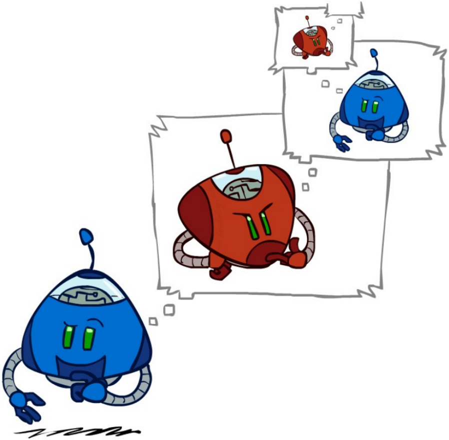


$$\boxed{V_{\text{exptmax}}(s_{\text{start}}) = 5}$$

Hình 3: Expectimax

Vấn đề

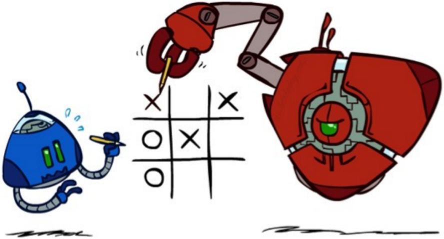
- Ta không biết rõ chiến lược của đối thủ
- Duy lý cá nhân
  - Mỗi người tham gia cuộc chơi sẽ cố gắng giành lợi ích tuyệt đối về bản thân
- Thuật toán Minimax



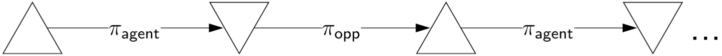
Minimax

Minimax I

- Nếu chúng ta có khả năng đọc suy nghĩ ⇒ **Expectimax**
- Tuy nhiên, trên thực tế chúng ta không biết chiến lược của đối thủ
- Giả định **trường hợp xấu nhất**: đối thủ làm mọi cách để giảm thiểu lợi ích



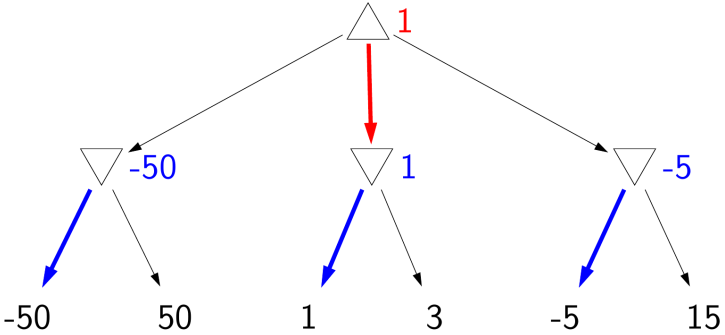
Minimax II



- Lợi ích của đối thủ được thay thế bằng lợi ích tối thiểu

$$V_{minmax}(s) = \begin{cases} Utility(s) & IsEnd(s) \\ \max_{a \in Actions(s)} V_{minmax}(Succ(s,a)) & Player(s) = agent \\ \min_{a \in Action(s)} V_{minmax}(Succ(s,a)) & Player(s) = opp \end{cases}$$

Minimax III



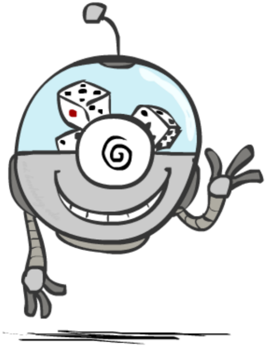
$$\boxed{V_{\text{minmax}}(s_{\text{start}}) = 1}$$

Hình 4: Minimax

Expectiminimax

Expectimax

- Nếu trò chơi có **yếu tố ngẫu nhiên**?
- Giả sử mỗi nút ngẫu nhiên ta đều biết được **phân phối** của các successor
- **Chiến lược bất định**



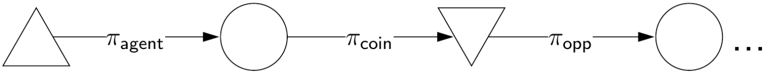
Trò chơi 3 chiếc hộp

- Có 3 chiếc hộp, mỗi hộp chứa 2 con số.
- Bạn chọn một chiếc hộp sau đó tung đồng xu; nếu ra mặt ngửa, thay vì chọn hộp của bạn, mình sẽ chọn hộp bên trái và một con số nằm trong hộp đó
- Nhiệm vụ của bạn là phải tối đa con số mà mình chọn

A		B		C	
-50	50	1	3	-5	15

Expectiminimax I

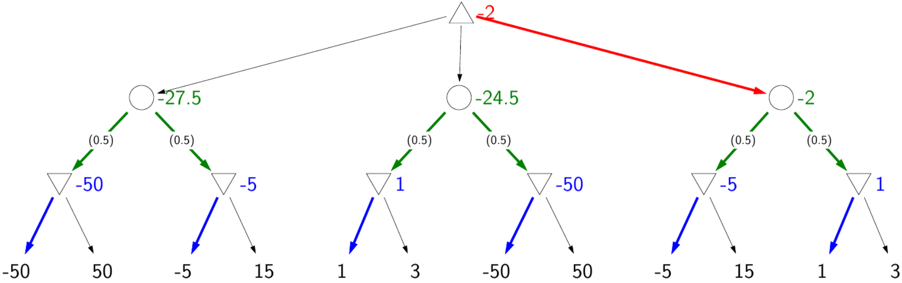
- Trò chơi trên có thể được mô hình bằng **expectiminimax**.



$$V_{exptminmax}(s) = \begin{cases} Utility(s) & IsEnd(s) \\ \max_{a \in Actions(s)} V_{exptminmax}(Succ(s,a)) & Player(s) = agen \\ \min_{a \in Actions(s)} V_{exptminmax}(Succ(s,a)) & Player(s) = opp \\ \sum_{a \in Actions(s)} \pi_{coin}(s,a) V_{exptminmax}(Succ(s,a)) & Player(s) = coin \end{cases}$$

Expectiminimax II

$\pi_{\text{coin}}(s, a) = \frac{1}{2}$ for $a \in \{0, 1\}$



$$\boxed{V_{\text{exptminmax}}(s_{\text{start}}) = -2}$$

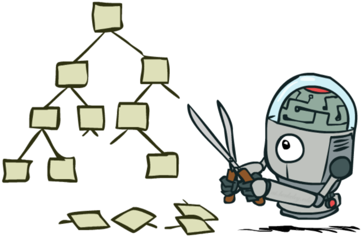
Hình 5: Expectiminimax

Tỉa nhánh Alpha-Beta

Tỉa nhánh Alpha-Beta I

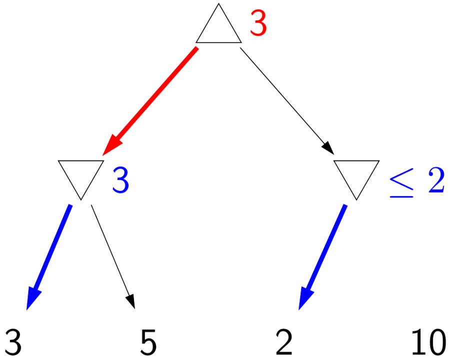
Ý tưởng chính:

- Giữ **ngưỡng trên - α** của nút **min** và **ngưỡng dưới - β** của nút **max**
- Tỉa nhánh một nút nếu khoảng giá trị của nút đó không giao với nút cha ($\alpha < \beta$)



Tỉa nhánh Alpha-Beta II

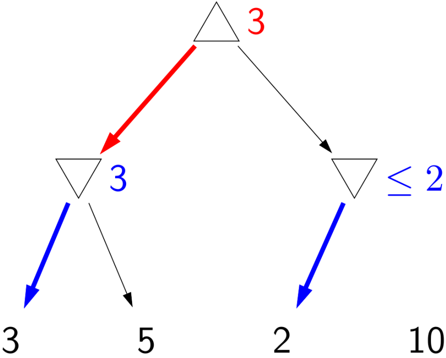
- Việc tỉa nhánh **không ảnh hưởng** đến giá trị cuối cùng của trò chơi
- Trong trường hợp tốt nhất độ phức tạp giảm xuống $O(b^{m/2})$
- Tìm kiếm toàn bộ cây trò chơi cờ vua???



Tỉa nhánh Alpha-Beta II

- Việc tỉa nhánh **không ảnh hưởng** đến giá trị cuối cùng của trò chơi
- Trong trường hợp tốt nhất độ phức tạp giảm xuống $O(b^{m/2})$
- Tìm kiếm toàn bộ cây trò chơi cờ vua???

vô vọng!!!



Hàm đánh giá

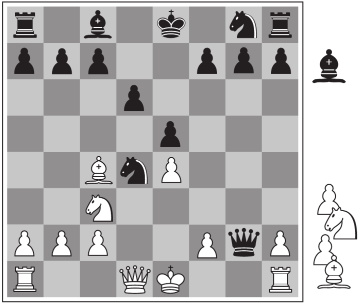
Hàm đánh giá

- **Xấp xỉ** lợi ích của trạng thái hiện tại
- Thứ tự trạng thái cuối cùng đúng với lợi ích của trò chơi
  - Các trạng thái chiến thắng cần được đánh giá tốt hơn hòa và hòa tốt hơn thua
- Việc tính toán không quá lâu!
- Với các trạng thái không phải là trạng thái cuối, việc đánh giá cần phải tương quan chặt chẽ với tỷ lệ chiến thắng

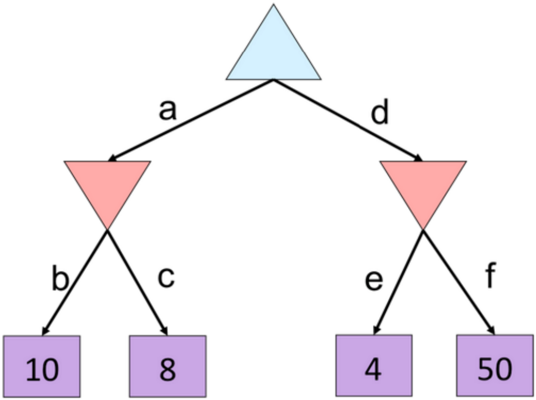
Ví dụ

$Eval(s) = material + mobility + king - safety + center - control$

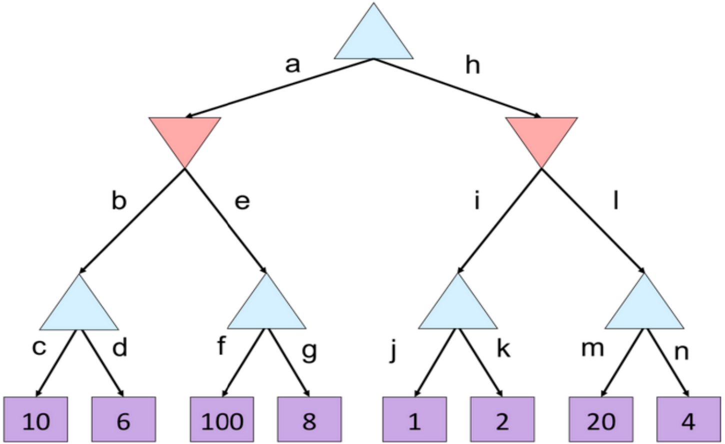
- $material = 10^{100}(K - K') + 9(Q - Q') + 5(R - R') + 3(B - B' + N - N') + 1(P - P')$
- ...



Bài tập I



Bài tập II



Tổng kết

Tổng kết

- **Cây trò chơi:** mô hình hóa trò chơi
- **Minimax:** tìm chiến lược chơi đối kháng
- **Hàm đánh giá:** tri thức cụ thể, xấp xỉ
- **Tỉa nhánh Alpha-Beta:** tri thức tổng quát, chính xác

Tài liệu tham khảo

[1] Michael Negnevitsky Russell, S. and Norvig, P. (2021).
"Artificial intelligence: a modern approach."

[2] Berkelev University
"CS188"