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

Cơ sở trí tuệ nhân tạo

Quy trình Markov

Nguyễn Ngọc Đức

2024

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

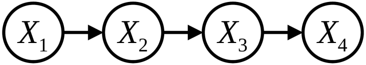
- Giá trị X trong một khoảng thời gian nhất định được gọi là trạng thái

$$X_1 \rightarrow X_2 \rightarrow X_3 \rightarrow X_4 \dashrightarrow$$

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

- 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

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

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:

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

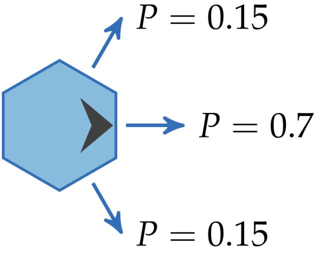
Ta có:

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

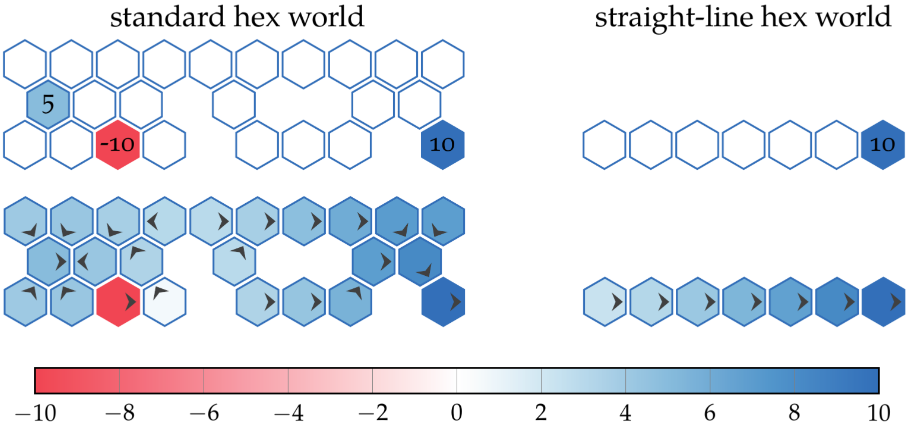
Tìm kiếm bất định

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

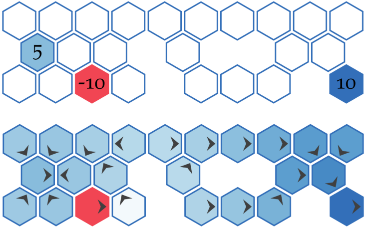


Ví dụ: Hex world II



Quy trình Markov

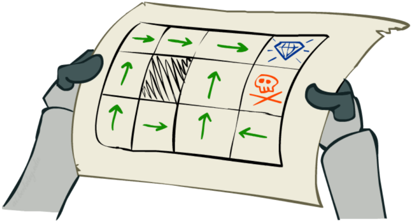
- Một quy trình Markov được định nghĩa dựa trên:
  1. Tập trạng thái $s \in S$
  2. Tập hành động $a \in A$
  3. Hàm chuyển dịch $T(s, a, s')$
  4. Hàm điểm thưởng $R(s, a, s')$
  5. Trạng thái bắt đầu
  6. Trạng thái kết thúc (có thể có hoặc không)



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

Chiến lược

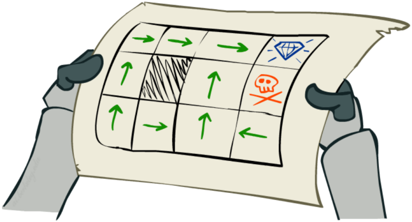
- Trong các bài toán tìm kiếm, chúng ta cần một kế hoạch tối ưu



- Với MDP:

Chiến lược

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

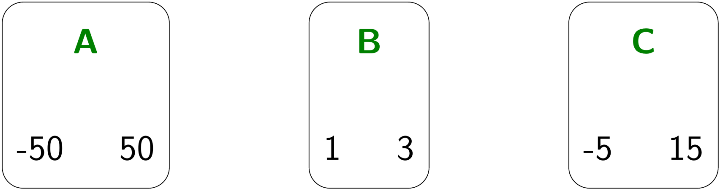


- Với MDP: Chúng ta cần một chiến lược tối ưu $\pi^* : S \to A$
  - Một chiến lược π đưa ra hành động tại một trạng thái cụ thể
  - Một chiến lược tối ưu sẽ tối đa hóa lợi ích kỳ vọng (expectimax)
  - Một chiến lược rõ ràng định nghĩa một tác tử

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

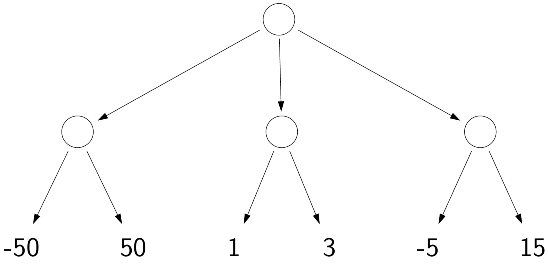


Hình 1: Ví dụ một trò chơi

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)

$$\bigcirc \xrightarrow{\pi_{\text{agent}}} \bigcirc \xrightarrow{\pi_{\text{opp}}} \bigcirc \xrightarrow{\pi_{\text{agent}}} \bigcirc \;\ldots$$

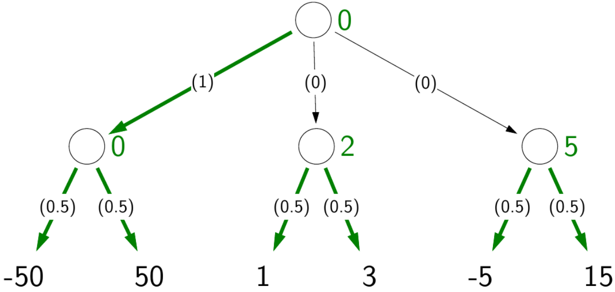
$$V_{eval}(s) = \begin{cases} Utility(s) & IsEnd(s) \\ \sum_{a \in Actions(s)} \pi_{agent}(s,a) V_{eval}(Succ(s,a)) & Player(s) = agent \\ \sum_{a \in Actions(s)} \pi_{opp}(s,a) V_{eval}(Succ(s,a)) & Player(s) = opp \end{cases}$$

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 $a \in \text{Actions}(s)$

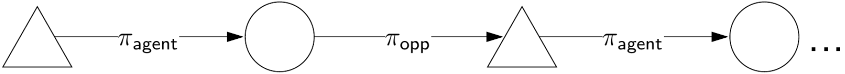


$\boxed{V_{\text{eval}}(s_{\text{start}}) = 0}$

Hình 2: Giá trị trò chơi

Expectimax I

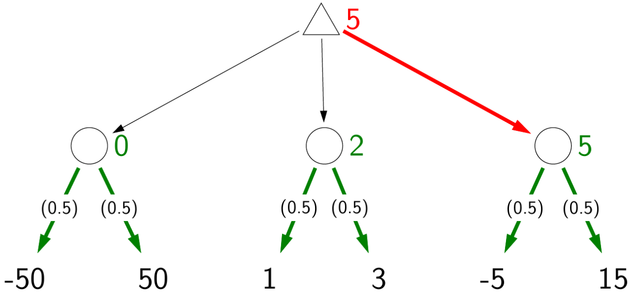
- **Giá trị expectimax** $V_{exptmax}(s)$ ở trạng thái s, là lợi ích tối đa của người chơi đạt được ở trạng thái s nếu biết trước chiến lược chơi của đối thủ



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

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

Tài liệu tham khảo

[1] Bùi Tiến Lên, Bộ môn Khoa học máy tính

Bài giảng môn Cơ sở trí tuệ nhân tạo

[2] Michael Negnevitsky

Artificial Intelligence: A Guide to Intelligent Systems (3rd Edition)

[3] Dan Klein and Pieter Abbeel

CS188: Introduction to Artificial Intelligence