IA

Tema 4: Markov

- **Definiciones del proceso de Decisión**
  - + **S**: conjunto de estados (s_0: estado inicial)
  - + **A(s)**: conjunto de acciones aplicables en s
  - + **Modelo de transición**: dado $P(s'|s,a)$, probabilidad de que s pase a s' dando una acción
  - + **Recompensa** ($R(s)$)
  - + **Propiedad de Markov**: El efecto de una acción sobre un estado solo depende de la acción y el estado al que se aplica
- **Definición de políticas**
  - + Una política recomienda en cada posible estado una acción a aplicar, esto es una función π, tal que $\pi(s) \in A(s)$
  - + La política óptima es aquella que maximiza la recompensa media en la secuencia de acciones posibles
- **Valoración de secuencia de estados**
  - + **Secuencia de estados**: $q_0\, q_1\, q_2 \ldots$
  - + **Valoración mediante recompensas con descuento**:

$$V([q_0 q_1 q_2 \ldots]) = R(q_0) + R(q_1)\cdot\gamma + R(q_2)\cdot\gamma^2 + \ldots$$

γ = factor de descuento (entre 0 - 1), con esto penalizamos las dimensiones del camino, si hay una R_{MAX} y $\gamma \in (0,1)$, la valoración no puede ser mayor que $R_{MAX}/(1-\gamma)$

  - + **Valoración mediante una política**

$$V^*(s) = R(s) + \gamma \cdot \sum_{s'} (P(s'|s, \pi(s)) \cdot V^*(s'))$$

$V^*(s)$ = Valor de la política en el estado s

$R(s)$ = Recompensa del estado s

γ = factor descuento

$\sum_{s'} (P(s'|s,\pi(s)) \cdot V^*(s'))$ = Sumatorio de la probabilidad de que se pase del estado s al s' (vecino) según su política para la valoración de dicho vecino

\# $V^*(s)$ devuelve un sistema de ecuaciones

* **Aproximación del cálculo de V^***

$$V^*_{i+1}(s) = R(s) + \gamma \cdot \sum_{s'} (P(s'|s, \pi(s)) \cdot V^*_i(s'))$$

\# Dándole un valor arbitrario V^*_0 a cada estado s

+ Políticas óptimas y valoración de estado

* Valoración de estado: mejor valoración de sus políticas $V(s) = \max_{\pi} V^{\pi}(s)$

* Política óptima (π^*): Desarrollar la anterior

$$\pi^*(s) = \underset{a \in A(s)}{\operatorname{argmax}} \sum_{s'} (P(s'|s,a) \cdot V(s'))$$

$V^{\pi^*} = V$

+ Ecuaciones de Bellman

Obtener $V(s)$ en función de los vecinos

$$V(s) = R(s) + \gamma \cdot \max_{a \in A(s)} \sum_{s'} (P(s'|s,a) \cdot V(s'))$$

Con esto, obtenemos la valoración de cada estado y de ahí la política óptima, de nuevo es un sistema de ecuaciones

Max de la valoración de tomar cada acción posible para ese estado

Se puede resolver de forma iterativa

La convergencia es rápida para γ pequeñas

* Cálculo de errores

1. Dada $\|V_{i+1} - V_i\| = \max_s |V_{i+1}(s) - V_i(s)|$

2. Si $\|V_{i+1} - V_i\| < \epsilon \cdot (1-\gamma)/\gamma \rightarrow \|V_{i+1} - V_i\| < \epsilon$

Criterio de parada

IA

Tema 4: Markov

Problema 1

Dado el conjunto $S = \{s_1, s_2, s_3\}$, con tres posibles acciones (a_1, a_2, a_3), siendo $A(s_1) = \{a_2, a_3\}$, $A(s_2) = \{a_1, a_3\}$ y $A(s_3) = \{a_1, a_2\}$.

$a_2 \rightarrow s_1 = (0; 0.3; 0.7)$ | $a_1 \rightarrow s_2 = (0.4; 0.1; 0.5)$ | $a_1 \rightarrow s_3 = (0.5; 0; 0.5)$

$a_3 \rightarrow s_1 = (0.1; 0.1; 0.8)^*$ | $a_3 \rightarrow s_2 = (0.6; 0.2; 0)^*$ | $a_2 \rightarrow s_3 = (0; 0.5; 0.5)^*$

$R(s_1) = -1$ | $R(s_2) = -0'04$ | $R(s_3) = 1$ | con descuento $(\gamma) = 0'9$

$\pi = [\pi(s_1) = a_3, \pi(s_2) = a_3, \pi(s_3) = a_2]$

a) Probabilidades de secuencia s_3, s_3, s_3, s_2, s_2 aplicando π:

$$P(\text{secuencia}) = P(s_3|s_3, a_2) \cdot P(s_3|s_3, a_2) \cdot P(s_2|s_3, a_2) \cdot P(s_2|s_2, a_3) = 0'5 \cdot 0'5 \cdot 0'5 \cdot 0'2 = 0'025$$

Valoración de la secuencia:

$$V(\text{secuencia}) = R(s_3) + \gamma R(s_3) + \gamma^2 (s_3) + \gamma^3 (s_2) + \gamma^4 (s_2) = 2'654596$$

b) Plantea V^*:

$$V^*(s_1) = R(s_1) + \gamma [(0.1) \cdot V^*(s_1) + (0.1) \cdot V^*(s_2) + (0.8) \cdot V^*(s_3)]$$
$$V^*(s_2) = R(s_2) + \gamma [(0.8) \cdot V^*(s_1) + (0.2) \cdot V^*(s_2)]$$
$$V^*(s_3) = R(s_3) + \gamma [(0.5) \cdot V^*(s_2) + (0.5) \cdot V^*(s_3)]$$

\# Recompensa del estado más el producto del sumatorio de la probabilidad según la política y el valor de ese estado

c) Plantear ecuaciones de Bellman:

\# Las ecuaciones de Bellman se plantean como el anterior tomando el máximo de los sumatorios de las acciones posibles

ej:

$$V(s_3) = R(s_3) + \gamma \cdot \max \begin{bmatrix} (0.5) \cdot V(s_1) + (0.5) \cdot V(s_3), \\ (0.5) \cdot V(s_2) + (0.5) \cdot V(s_3) \end{bmatrix}$$

\# La política óptima será la max para cada estado de las ecuaciones de Bellman

Plantear problema

Crear una entrada por cada estado posible y su recompensa, luego una entrada por cada acción que tenga cada estado, y la probabilidad de pasar a cada estado tras tomarla una acción

Estado
- $a_1 \rightarrow$ (s_1 0.1, s_2 0.2, s_3 0.7)
- a_2
- a_3

La política mantendrá que a_i se tenga

Hacer una iteración $k+1$

1º Calculamos los valores de V^{π}(Estado), con π_0

2º Calculamos las valoraciones para tomar la acción con máxima valoración, siendo V^{π}(Estado), los calculados en el primer punto

3º Comprobar que política tiene mayor valoración por estado