MDPs: overview

Son un modelo matemático para la toma de decisiones bajo incertidumbre, este se desarrollo en los años 50-60 y están relacionados con las cadenas de Markov y permiten tomar decesiones en sistemas estocásticos.

Sus aplicaciones son:

- Robótica: decidir movimientos con actuadores defectuosos o obstáculos inesperados.
- Asignación de Recursos: decidir qué producir sin conocer la demanda exacta.
- Agricultura: decidir qué plantar sin conocer el clima futuro.

Definición de un MDP:

- **Estados (S):** conjunto de posibles estados del sistema.
- **Acciones (A):** conjunto de acciones disponibles en cada estado.
- **Transición (T(s, a, s')):** probabilidad de pasar de estado *sss* a estado ***s′s′***s′ al tomar acción *aa*a.
- **Recompensa (R(s, a, s')):** recompensa obtenida en una transición.
- **Factor de descuento (*γ\gamma*γ):** ponderación de recompensas futuras (valor entre 0 y 1).

Evaluación de Políticas

- **Política (*π\pi*π)**: función que asigna una acción a cada estado.
- **Valor de una política *Vπ(s)V_\pi(s)*Vπ (s)**: utilidad esperada de seguir la política desde *sss*.
- **Q-valor *Qπ(s,a)Q_\pi(s, a)*Qπ (s,a)**: utilidad esperada de tomar acción *aa*a en *sss* y seguir la política después.
- Se puede calcular mediante un algoritmo iterativo hasta la convergencia.

Algoritmos para Resolver MDPs

Evaluación de Políticas

- Se inicializa *V(s)=0V(s) = 0*V(s)=0.
- Se actualizan los valores usando:

*Vπ(s)=∑s′T(s,π(s),s′)[R(s,π(s),s′)+γVπ(s′)]V_{\pi}(s) = \sum_{s'} T(s, \pi(s), s')[R(s, \pi(s), s') + \gamma V_{\pi}(s')]*Vπ (s)=s′∑ T(s,π(s),s′)[R(s,π(s),s′)+γVπ (s′)]

- Se repite hasta convergencia.

Iteración de Valor (Value Iteration)

- Se busca la mejor política optimizando el valor esperado:

$V*(s)=max⁡a∑s'T(s,a,s')[R(s,a,s')+\gamma V*(s')]V^*(s) = \max_{a} \sum_{s'} T(s, a, s') [R(s, a, s') + \gamma V^*(s')]V∗(s)=amax s′∑ T(s,a,s′)[R(s,a,s′)+γV∗(s′)]$

- Se actualizan los valores de los estados iterativamente.

Iteración de Política

- Se alterna entre evaluar una política y mejorarla iterativamente.