
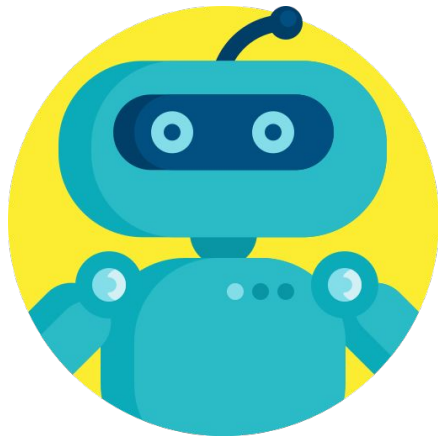

Aprendizaje por Refuerzo

- 
- Nombres:
 - Jhoan Sebastian Almeida Caicedo
 - Daisy Alejandra Ayala
 - Santiago Arredondo
- 
-
-

¿Qué es aprendizaje por refuerzo?

Es una rama del aprendizaje automático inspirada en la psicología conductista. El aprendizaje por refuerzo involucra uno o más agentes que tienen estados y acciones que pueden realizar en un ambiente dado con el fin de obtener recompensas.



¿En qué consiste el aprendizaje por refuerzo?

Es una área de la inteligencia artificial que está centrada en descubrir qué acciones se debe tomar para maximizar la señal de recompensa.



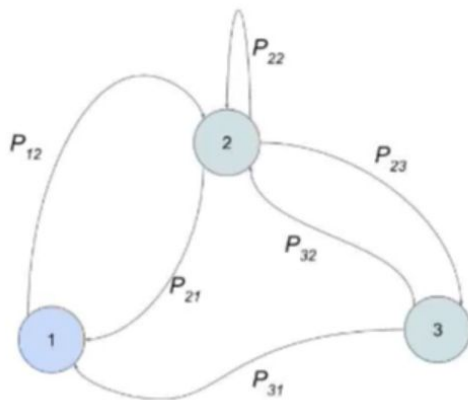
Elementos del Aprendizaje por refuerzo

- **Política:** es una regla utilizada por el agente para tomar decisiones.
- **Estados:** instancia o descripción completa del ambiente en el que se desenvuelve el agente.
- **Medio ambiente:** es el contexto donde se desenvuelve el agente, este tiene las reglas de juego también.
- **Acciones:** son distintas interacciones que tiene el agente con su entorno; diferentes entornos conducen distintas acciones basadas en el agente y estas tienden a ser finitas.
- **Recompensas:** son estímulos que recibe el agente, es un valor numérico que mide que tanto el agente se acerca o aleja de la solución ideal.

Fórmula matemática explicación

$$V^*(s) = \max_{a \in \mathcal{A}} \mathbb{E}_{s'} \{r(s, a) + \gamma V^*(s')\}$$

Procesos de Markov



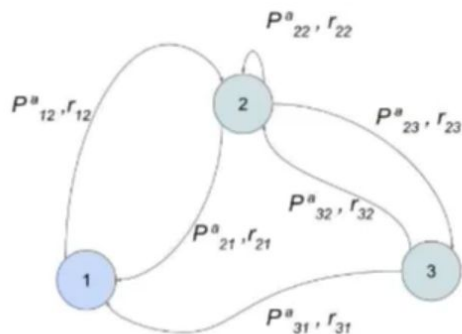
$$P_{ss'} = \begin{pmatrix} P_{11} & P_{12} & \cdots & P_{1n} \\ P_{21} & P_{22} & \cdots & P_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ P_{n1} & P_{n2} & \cdots & P_{nn} \end{pmatrix}$$

S_t = “Estado en tiempo t ”

$$\mathbb{P}[S_{t+1} | S_t, S_{t-1}, \dots, S_0] = \mathbb{P}[S_{t+1} | S_t]$$

Origen Algoritmo

Markov Decision Processes (MDPs)



$a_t \in \mathcal{A} = \{ \text{"Arriba"} , \text{" Abajo "}, \dots \}$

$r_t = \text{"Recompensa en tiempo } t\text{"}$

$$P^a_{ss'} = \begin{pmatrix} P^a_{11} & P^a_{12} & \dots & P^a_{1n} \\ P^a_{21} & P^a_{22} & \dots & P^a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ P^a_{n1} & P^a_{n2} & \dots & P^a_{nn} \end{pmatrix}$$

Encontrar la politica pi que maximice el valor para cada estado

$$V^*(s) = \max_{\pi} V^{\pi}(s)$$

Política de Control

Elegimos en cada momento una acción en función del estado

$$a_t = \pi(s_t)$$

Valor de un estado

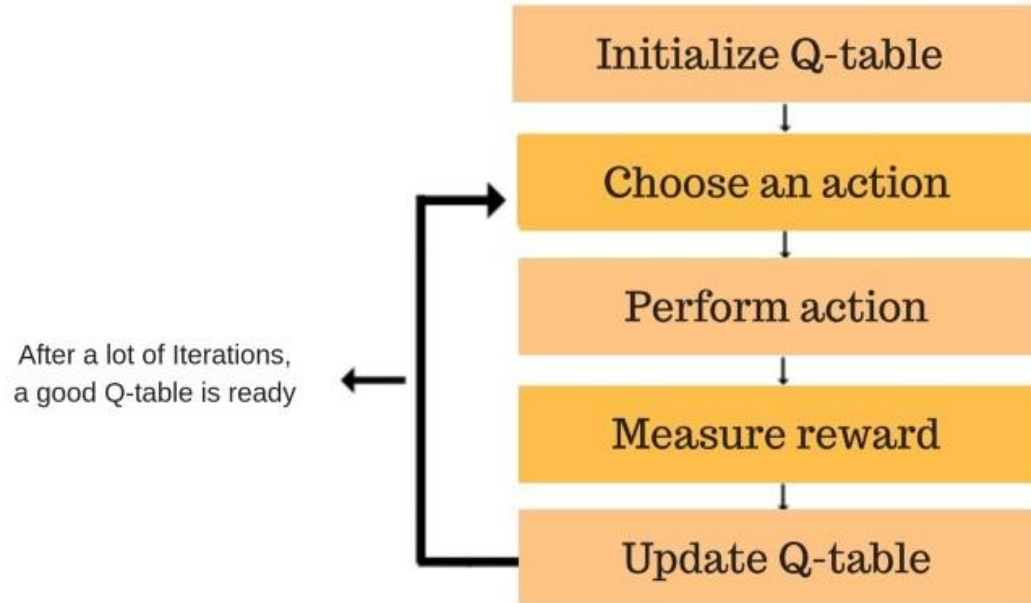
Recompensa a largo plazo si empezamos en s y seguimos la política π

$$V^{\pi}(s) = \mathbb{E} \left\{ \sum_{t=0}^{\infty} \gamma^t r_t | s_0 = s \right\}$$

Q-Learning

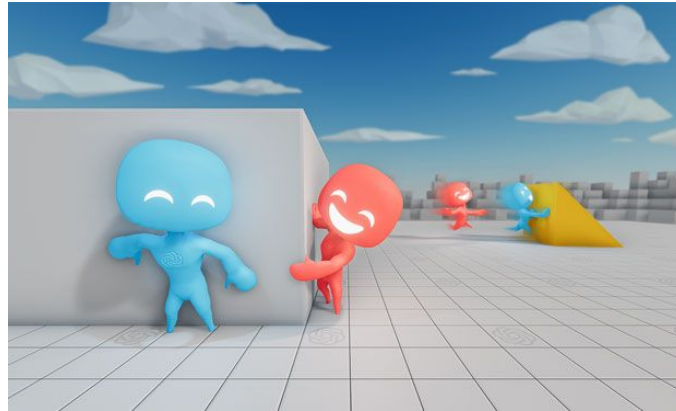
- Q-Learning es un **algoritmo** de aprendizaje de refuerzo basado en valores que se utiliza para encontrar la política óptima de selección de acciones mediante una función Q.
- Nuestro objetivo es maximizar la función de valor Q.
- La tabla **Q** nos ayuda a encontrar la mejor acción para cada estado.
- Ayuda a maximizar la recompensa esperada seleccionando la mejor de todas las acciones posibles.
- Q (estado, acción) devuelve la recompensa futura esperada de esa acción en ese estado.
- Esta función se puede estimar usando Q-Learning, que actualiza iterativamente Q (s, a) usando la **ecuación de Bellman**.
- Inicialmente exploramos el entorno y actualizamos la Q-Table. Cuando la Q-Table esté lista, el agente comenzará a explotar el entorno y comenzará a tomar mejores medidas.

Q-Learning



OPENAI y Unity

El objetivo de este sistema es demostrar usos prácticos del q-learning en juegos y simulaciones.



Vídeo juegos

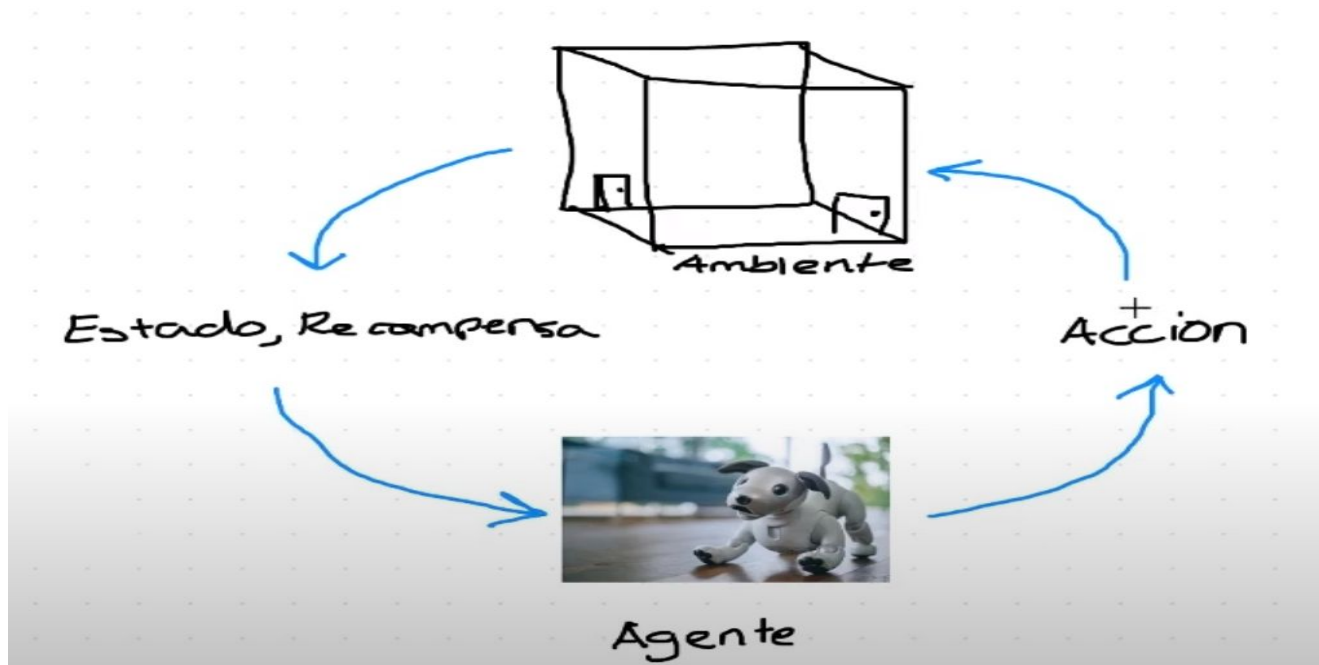
Open AI desarrolló un sistema llamado Dactyl

El objetivo de este sistema es demostrar que el entrenamiento por refuerzo en simulaciones puede lograr un gran impacto en la vida real.



Dexterity

Conclusión



Ejercicio Demostrativo

EjemploOPENAI

Bibliografía

[1] M. Silva, «medium,» [En línea]. Available: <https://medium.com/aprendizaje-por-refuerzo-introducci%C3%B3n-al-mundo-del/aprendizaje-por-refuerzo-introducci%C3%B3n-al-mundo-del-rl-1fcfbaa1c87>. [Último acceso: 15 03 2020].

[2] M. Silva, «medium,» [En línea]. Available: <https://medium.com/aprendizaje-por-refuerzo-introducci%C3%B3n-al-mundo-del/aprendizaje-por-refuerzo-procesos-de-decisi%C3%B3n-de-markov-parte-1-8a0aed1e6c59>. [Último acceso: 15 03 2020].

[3]<https://www.freecodecamp.org/news/an-introduction-to-q-learning-reinforcement-learning-14ac0b4493cc/>

[4] <https://www.youtube.com/watch?v=GtOsQCCVJDA>