

Proyecto de aprendizaje por refuerzo

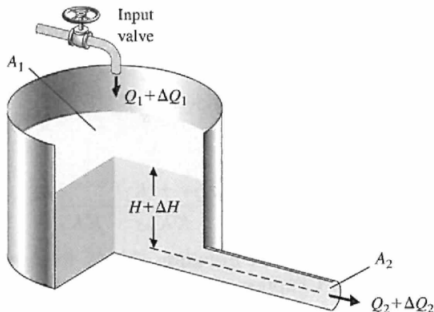
Emmanuel Peto Gutiérrez

IIMAS
UNAM

11 de junio de 2023

El sistema

Sistema de tanque de agua con una bomba de entrada y un agujero de salida.



El modelo matemático

La ecuación diferencial:

$$\frac{dH}{dt} = -\left(\frac{A_2}{A_1} \sqrt{2g}\right) \sqrt{H} + \frac{1}{\rho A_1} Q_1$$

La función de transferencia continua:

$$\frac{0.0013}{s + 0.0221}$$

La función de transferencia discreta ($t_s=0.2s$):

$$\frac{2.5409 \times 10^{-4}}{z - 0.9956}$$

La altura en un instante:

$$h^{(i)} = 0.9956 h^{(i-1)} + 2.5409 \times 10^{-4} u$$

Introducción

Implementación

Resultados

- ▶ **Estados:** $s \in [0, 2]$ (continuo)
- ▶ **Acciones:** $a \in [0, 50]$ (continuo)
- ▶ **Probabilidades:** $p(s', r|s, a) = 1$, donde $s' = 0.9956s + 2.5409 \times 10^{-4}a$.
- ▶ **Recompensa:** La recompensa es -1 por llegar a un estado no terminal y 200 por llegar a un estado terminal. Un estado terminal se alcanza cuando la altura actual del tanque se acerca a la altura objetivo con un margen de 0.01m.
- ▶ **Descuento:** $\gamma = 0.99$

Se utilizó el algoritmo DQN para entrenar al agente, usando ϵ -greedy para la selección de la acción. Al final se obtiene una red neuronal entrenada, la cual se utiliza para hacer control sobre el sistema y llegar a la altura objetivo. Los hiperparámetros usados en el algoritmo son

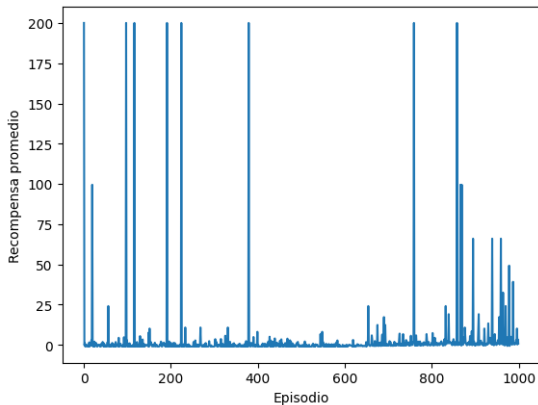
- ▶ $\alpha: 1 \times 10^{-4}$
- ▶ ϵ : valor decreciente de 1 a 0.01
- ▶ Tamaño de batch: 32
- ▶ Episodios: 1000
- ▶ Pasos (máximos) por episodio: 1000

Recompensa promedio

Proyecto de
aprendizaje por
refuerzo

Emmanuel Peto
Gutiérrez

En la siguiente gráfica se muestra la recompensa promedio de cada episodio de la fase de entrenamiento.

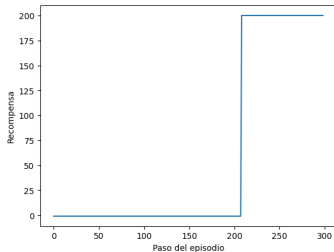
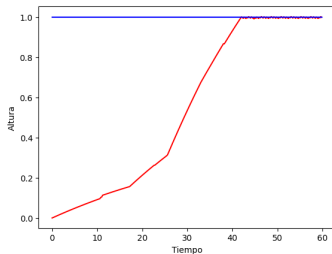


Introducción

Implementación

Resultados

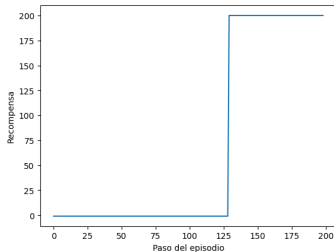
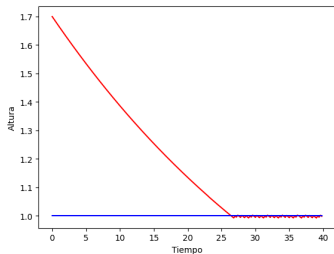
Fase de control cuando la altura inicial es 0.



Resultados

Control - parte 2

Fase de control cuando la altura inicial es 1.7.



Proyecto de
aprendizaje por
refuerzo

Emmanuel Peto
Gutiérrez

Introducción

Implementación

Resultados