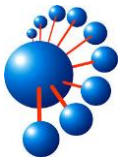


Tarea 2

Aprendizaje por Refuerzo (2023-2)

Integrantes: José Beltrán Alarcón

Dazhi Feng Zong

Pablo Zapata Schifferli

Profesor: Julio Godoy

Ayudante: Felipe Cerda

Fecha: 14 de diciembre, 2023

Arquitectura Red neuronal

El frame de las observaciones es transformado a escalas de grises y el tamaño y ancho es reducido a la mitad cada uno.

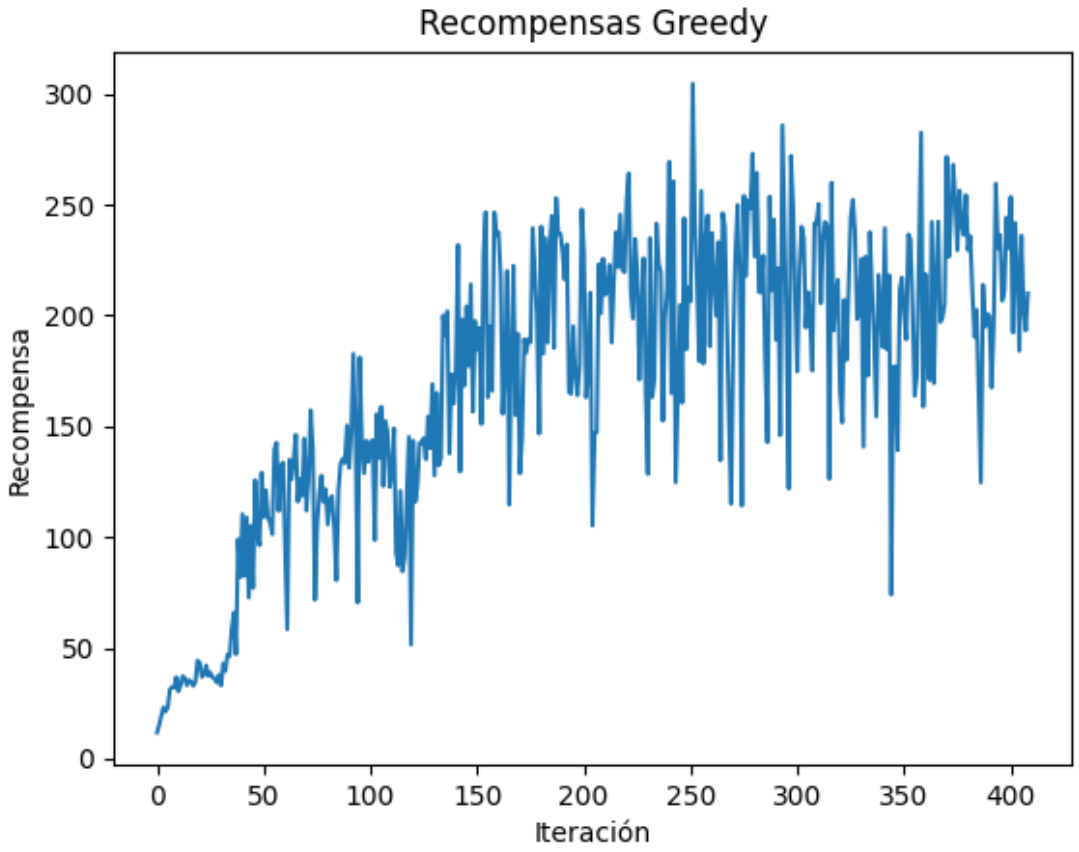
La red neuronal se define en la clase DQN, que hereda de nn.Module. A continuación, se explica la arquitectura capa por capa:

1. Capa convolucional 1:
    - Tipo: nn.Conv2d
    - Entrada: La observación del entorno (el estado).
    - Salida: 32 canales de características.
    - Kernel: Tamaño 8x8, con un paso (stride) de 4.
    - Función de activación: ReLU.
2. Capa convolucional 2:
    - Tipo: nn.Conv2d
    - Entrada: 32 canales de la capa anterior.
    - Salida: 64 canales de características.
    - Kernel: Tamaño 4x4, con un paso de 2.
    - Función de activación: ReLU.
3. Capa convolucional 3:
    - Tipo: nn.Conv2d
    - Entrada: 64 canales de la capa anterior.
    - Salida: 64 canales de características.
    - Kernel: Tamaño 3x3, con un paso de 1.
    - Función de activación: ReLU.
4. Capa totalmente conectada 1:
    - Tipo: nn.Linear
    - Entrada: Salida de la última capa convolucional aplanada.
    - Salida: 512 neuronas.
    - Función de activación: ReLU.
5. Capa totalmente conectada 2:
    - Tipo: nn.Linear
    - Entrada: 512 neuronas de la capa anterior.
    - Salida: 5 neuronas (acciones).
    - Función de activación: Ninguna.

- La red tiene dos instancias, la policy network y la target network. La target network se actualiza de manera pasiva.
- La función de pérdida utilizada es nn.SmoothL1Loss (Huber loss) y se optimiza utilizando el algoritmo AdamW. Se usa memory replay para la optimización.

Resultados

Cada iteración es el promedio de 10 juegos usando la política greedy, la cual ocurre cada 20 episodios de entrenamiento.



Las recompensas están definidas por lo siguiente:

1. Recompensas por defecto de Frogger, que son cuando gana puntaje.
2. Avanzar: 0.01
3. Retroceder: -0.01
4. Morir: -0.5

Al parecer frogger-v5 no da recompensas negativas y tiene 5 acciones en vez de las 3 mencionadas en la descripción de la tarea.

Hiperparámetros

BATCH_SIZE = 128

REPLAY_START = BATCH_SIZE

BUFFER_SIZE = 100000

TAU = 0.01

EPISODES = 10000

GAMMA = 0.90

EPS_START = 1

EPS_END = 0.05

EPS_DECAY = 0.99996

LEARNING_RATE = 0.00025