Universidad Católica San Pablo | Departamento de Ciencia de la Computación

25 AÑOS

CS362 – Robótica 2022-01 (CCOMP10-1)
Laboratorio RL-03
Implementación y pruebas de VizDoom

Alumna: Rossmery Loayza Soto
rossmery.loayza@ucsp.edu.pe

I. Revise la configuración del Modelo Deep Q-Learning utilizada y verifique los siguientes ítems:

A. ¿Cómo es el modelo de mundo utilizado: MDP, Modelo de premios y modelo Q-Learning?

EXPERIMENTO 1:
El modelo utilizado fue MDP (Markov Decision Process) y Q-learning es usado para aprender la política.

EXPERIMENTO 2:
El modelo utilizado fue MDP (Markov Decision Process) y Q-learning es usado para aprender la política.

La variación en ambos experimentos se ven en el optimizador utilizado, mientras que en el experimento 1 se usó SGD, el experimento 2 utilizó RMSprop como optimizador.

B. ¿Qué tipo de política es usada?

EXPERIMENTO 1:
La política es e-greedy.

EXPERIMENTO 2:
La política es e-greedy

C. ¿Cómo se calculan y aproximan los valores Q?

EXPERIMENTO 1:
La función que actualiza los valores Q se aproxima con una red neuronal convolucional, que está entrenado con Gradiente Estocástico Decente (SGD).

EXPERIMENTO 2:
La función que actualiza los valores Q se aproxima con una red neuronal convolucional, que está entrenado con RMSprop.

D. ¿Qué modelo de optimización se usa?

EXPERIMENTO 1:
Reinforcement learning

EXPERIMENTO 2:
Reinforcement learning

E. ¿Existe estrategia de repetición de experiencias?

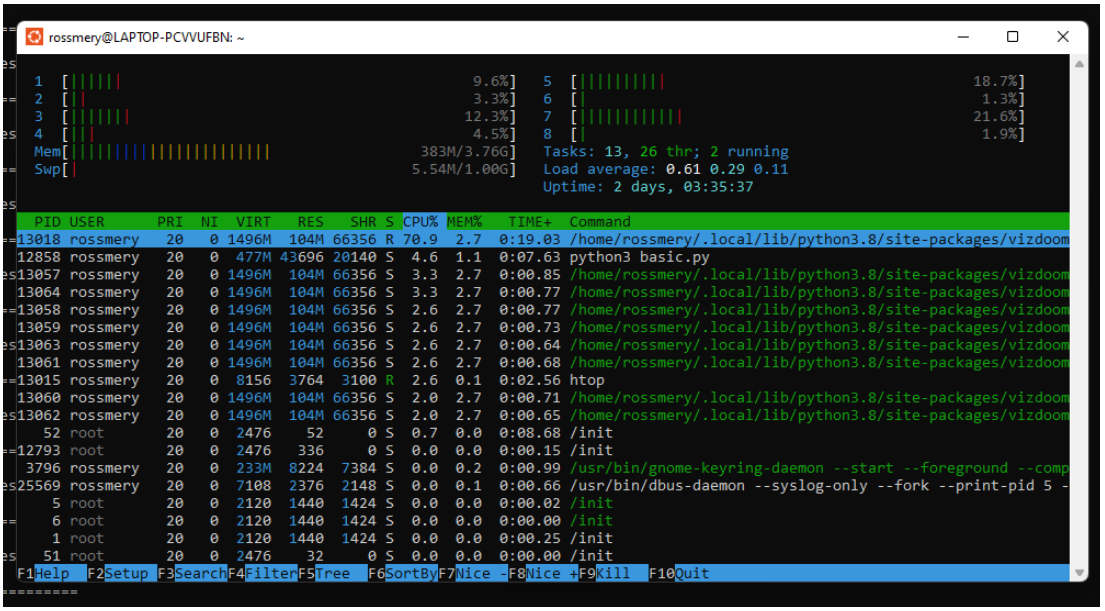
EXPERIMENTO 1:
Sí, se utiliza la repetición de experiencia, pero no la congelación de la red de destino (target network freezing, esta técnica consta de congelar los parámetros de la red de destino dentro de un cierto número de pasos esto hace que el algoritmo de Deep Q-learning se vuelva más estable).

EXPERIMENTO 2:
Sí, del mismo modo que el experimento 1, se utiliza la repetición de experiencia, pero no la congelación de la red de destino (target network freezing, esta técnica consta de congelar los parámetros de la red de destino dentro de un cierto número de pasos esto hace que el algoritmo de Deep Q-learning se vuelva más estable).

F. ¿Existen objetivos Q fijos?

II. Haga análisis de ejecución: tiempos de proceso, uso de memoria, uso de CPU/GPU para los dos experimentos

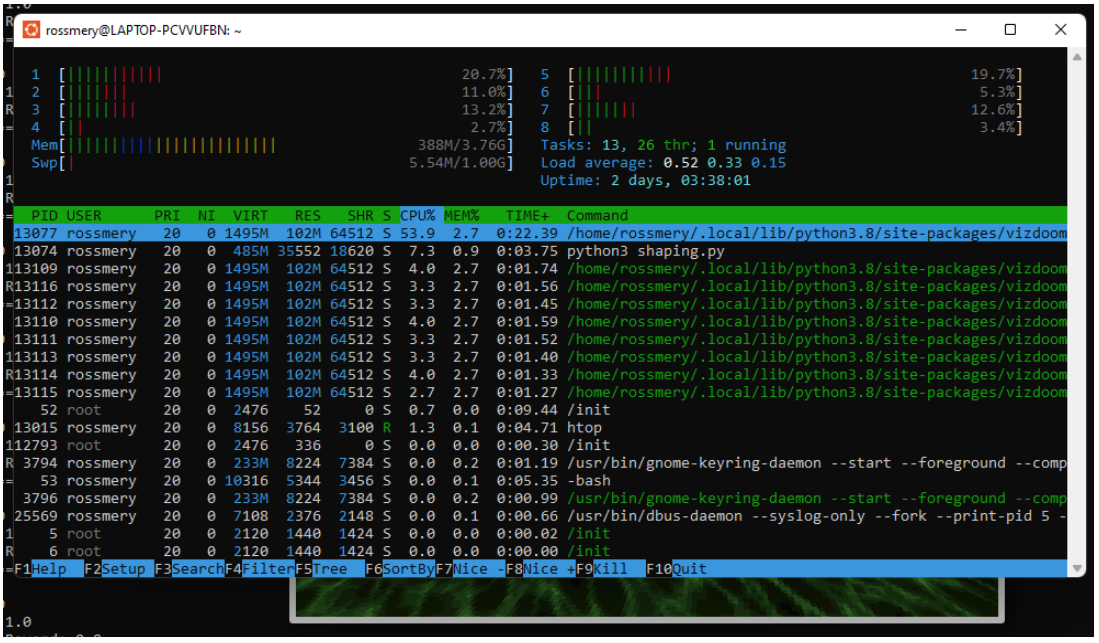
Experimento 1:



Como se puede verificar en la imagen:

- Se usan los 8 núcleos del procesador, pero no a su máximo rendimiento.
- La memoria utilizada es de 383 MB (2.7%).
- Uso de CPU es del 70.9%.

Experimento 2:



Del mismo modo que con el Experimento 1, se pueden hacer las siguientes observaciones:

- Se usan los 8 núcleos del procesador, en este caso en mayor porcentaje cada uno.
- La memoria en uso fue de 388 MB (2.7%).
- El uso de CPU en este experimento fue de 53.9%

III. Haga un análisis de convergencia y resultados obtenidos para ambos experimentos

A lo largo de los experimentos se pudo observar que ambos modelos muestran que los agentes mejoran con el tiempo, aunque en el segundo experimento no se pudo generar una política óptima que le permita al agente coleccionar todos los medikits. También en el paper es mencionado que uno de los factores que influyen en la velocidad del sistema de aprendizaje depende en gran medida de la cantidad de “frames” que el agente puede omitir durante el aprendizaje.