CIENCIA DE LA COMPUTACIÓN

Laboratorio RL-03: Implementación y Pruebas de ViZDoom

Robótica

Nombre del Alumno:

Patrick Xavier Marquez Choque

Semestre: 10

Año: 2021

Grupo: CComp 10-1

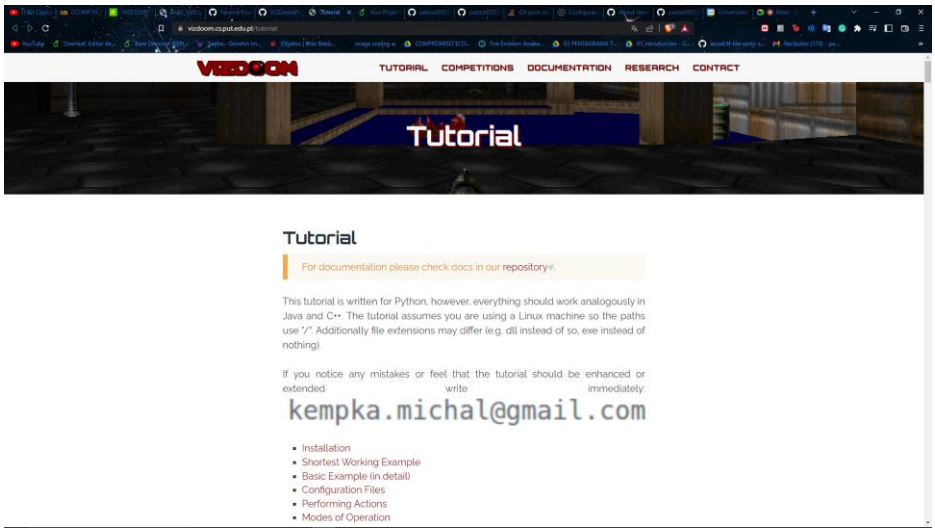
“El alumno declara haber realizado el presente documento de acuerdo a las normas de la Universidad Católica San Pablo”

_____

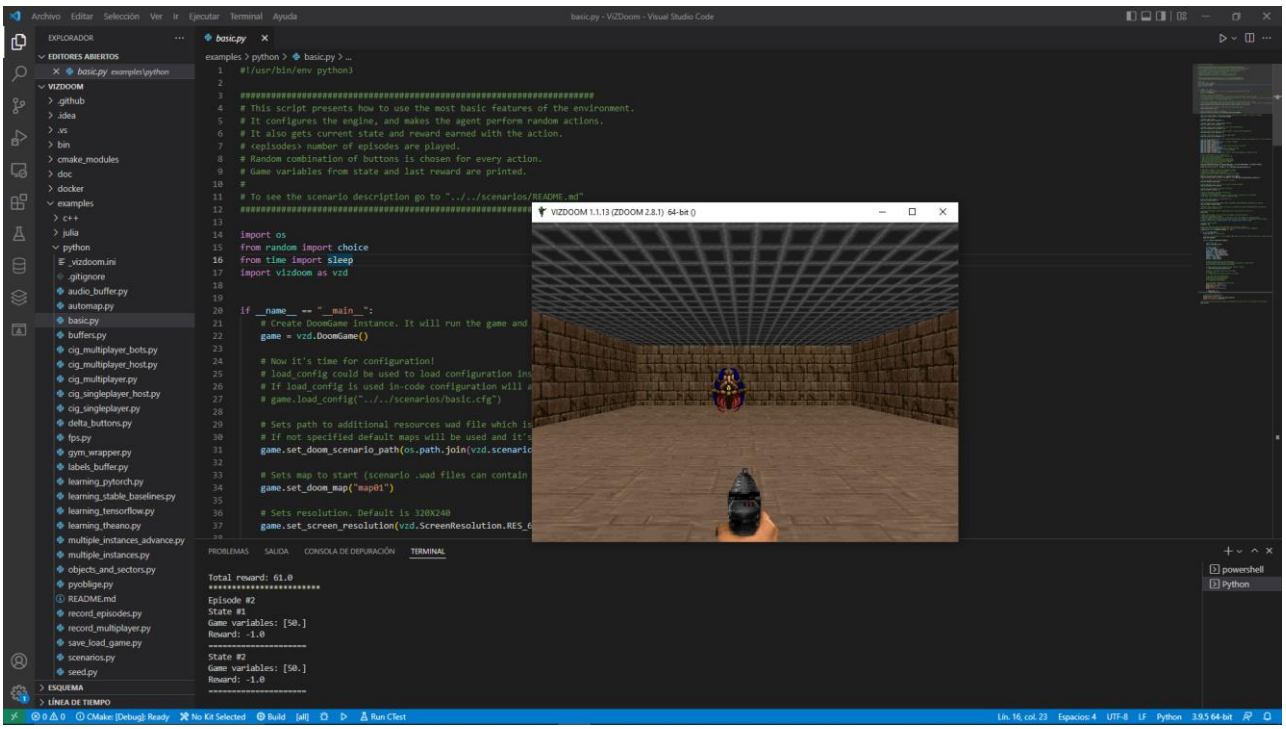
Ejercicios Laboratorio RL-03 Implementación y Pruebas de ViZDoom

1. Réplica del Artículo “ViZDoom: A Doom Based AI Research Platform for Visual Reinforcement Learning”

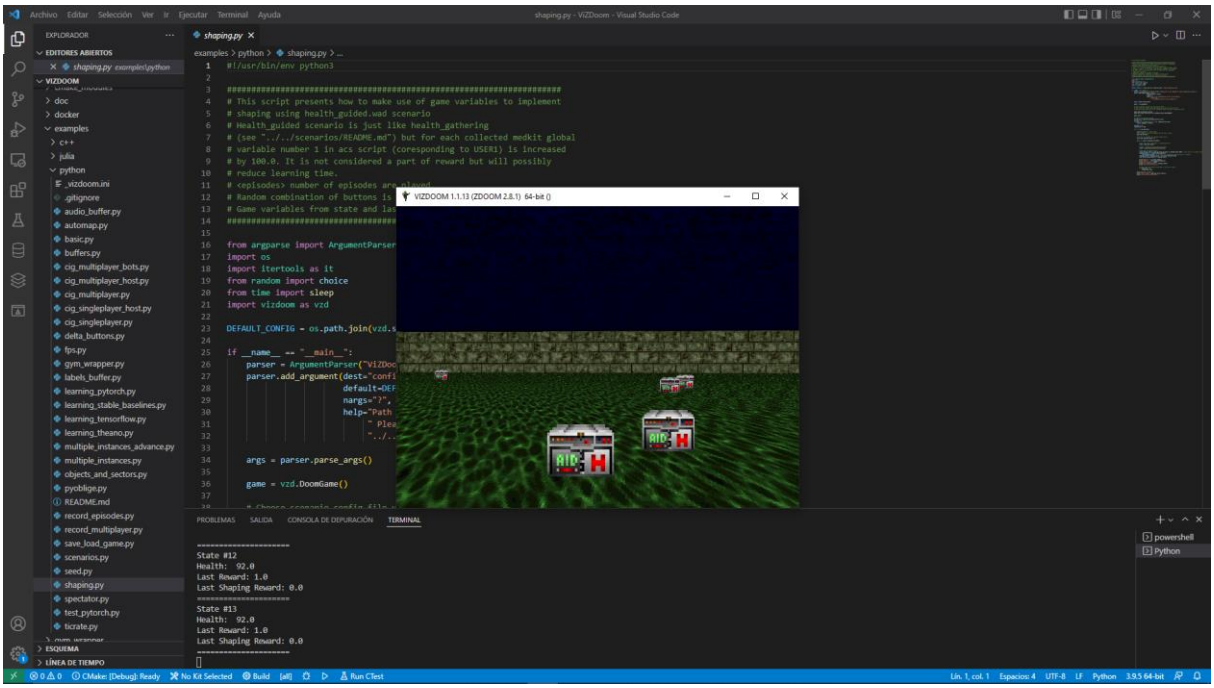
   La réplica de este proyecto se encuentra en el siguiente repositorio: https://github.com/patrick03524/Replica-Proyect-ViZDoom



2. Ejecución de los Experimentos
   a. Ejecutar el experimento básico con escenario del agente en un salón frente a un Cacodemon colocado en forma aleatoria. Verificar comportamiento del agente (Doom guy) y aprendizaje.



b. Ejecutar el experimento Medikit Collecting, que consiste en que el agente aparece aleatoriamente en un lago lleno de ácido sulfúrico que le va quitando la vida lenta pero constantemente. Verificar cuál es el comportamiento que se aprende.



3. Configuración del Modelo Deep Q-Learning

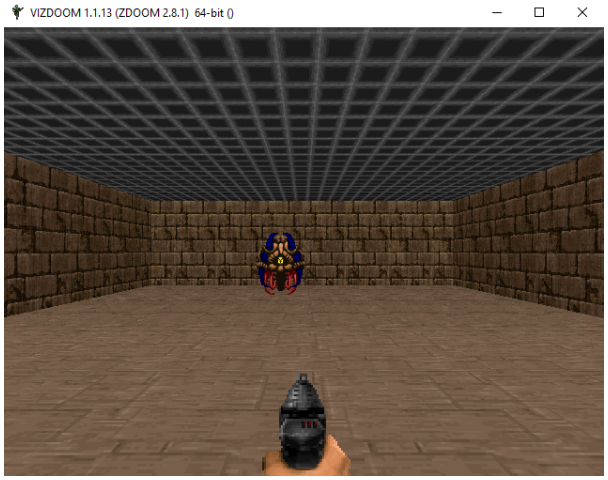
a. Cómo es el modelo de mundo utilizado: MDP, Modelo de premios y modelo Q-Learning

En primer lugar, debemos aclarar, ya que el juego base con el que se está trabajando es DOOM, el proceso de aprendizaje es un modelo Deep Q-Learning como otros juegos implementados de la consola Atari.

Ahora lo interesante es el Modelo de premios MDP, que necesita elegir caminos o acciones que lleguen a “un mejor escenario”, esta política se conoce como política por refuerzo greedy debido a que necesita de algún tipo de premio en caso de completar la tarea exitosamente (similar a la política estado-acción).

Haremos un análisis por experimento de la siguiente manera:

i. Experimento 1:
Dentro de este experimento habrá 3 acciones para nuestro agente “Doom Guy” que son las acciones de **movimiento hacia la derecha o izquierda y disparar**.

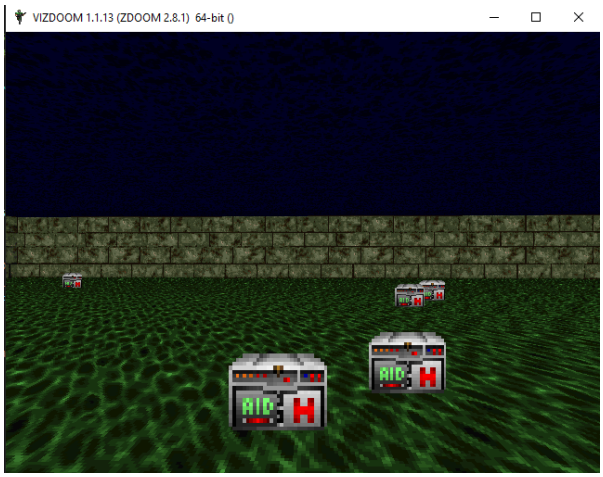


Cada vez que realiza la primera acción se le restará un punto al puntaje total y en la segunda acción se le restará 5 puntos si el disparo falló.

Como modificación se dará 1000 puntos al puntaje total. Esto se hace con el propósito que exista una mayor tendencia de que el modelo vaya siempre por la mejor opción con el fin de acumular el mayor valor posible y matar al monstruo en cada iteración.

ii. Experimento 2:

Dentro de este experimento existirán 4 acciones para nuestro agente “Doom Guy” que son las acciones de **movimiento hacia adelante, atrás, izquierda y derecha.**



Cada vez que se realiza una acción se estará restando la vida del agente hasta que, en caso de morir se le restará 1000 puntos, encuentre algún medikit y le recupere la vida.

Como modificación de este experimento no aumentaremos la recompensa por recoger un medikit pero aumentaremos el castigo

por tic de daño que reciba del veneno para que así se pueda observar el caso donde el agente decida desplazarse para encontrar la manera de no morir.

b. Qué tipo de política es usada.

   La política utilizada es E-Greddy con decaimiento lineal e.
   Esta política utiliza los premios como el valor del puntaje para que aumente el porcentaje de cada acción y que pueda deducir cual es la mejor acción en cada estado.

c. Cómo se calculan y aproximan los valores Q.

   Los valores Q se aproximan utilizando una red neuronal CNN que consta principalmente de 3 capas convolucionales, 1 capa de agrupación máxima, 1 capa totalmente conectada y por última 1 capa de salida. (Los parámetros y la cantidad de nodos en cada capa varían dependiendo del experimento en cuestión).

   La tabla de valores funciona de la siguiente manera:
   Primero es necesario la inicialización de una matriz de N estados x la cantidad de acciones en cada experimento. Todos los valores se inicializan con 0, por lo tanto, la primera acción de cada experimento en cuestión de movimiento será la misma en ambos experimentos y siguiendo tomando acciones aleatorias así obteniendo

d. Qué modelo de optimización se usa.

   Se utilizará un modelo de optimización con gradiente estocástica descendiente.

e. ¿Existe estrategia de repetición de experiencias?

   Si ya que esto es necesario para la experimentación.

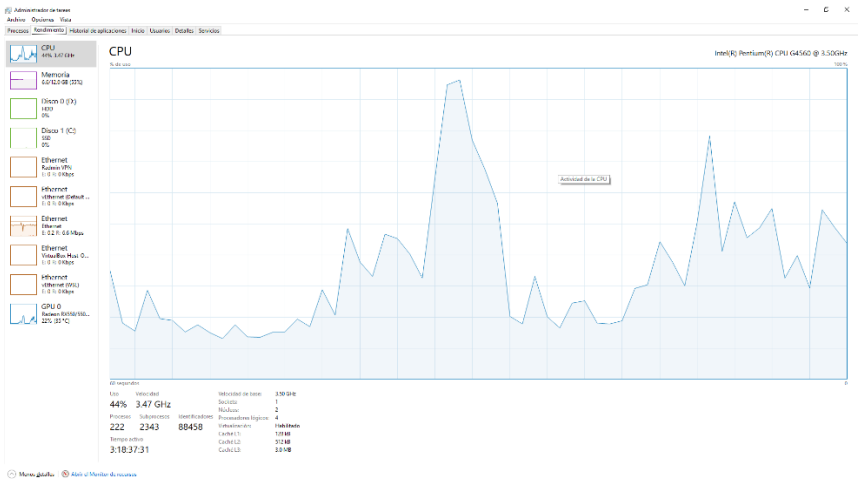
   Para ambos experimentos el agente siempre se iniciará en la misma posición que es en el medio del escenario.

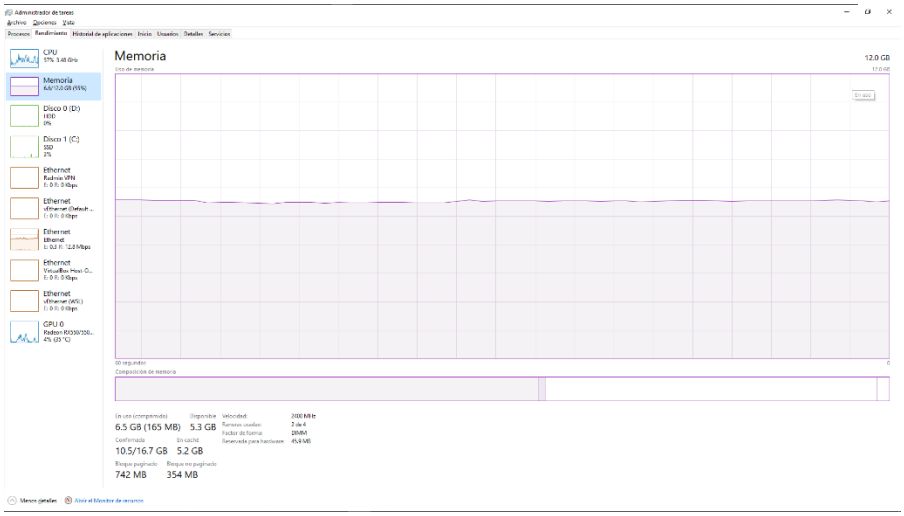
f. ¿Existen objetivos Q fijos?

   Si existe objetivos fijos ya que el objetivo de estos experimentos es ganar el juego.
   Para el primer experimento del Cacodemon se lográ matando al monstruo exitosamente.
   Para el segundo experimento del Medikit Collecting se logrará alcanzando una puntuación máxima de 2100 puntos.

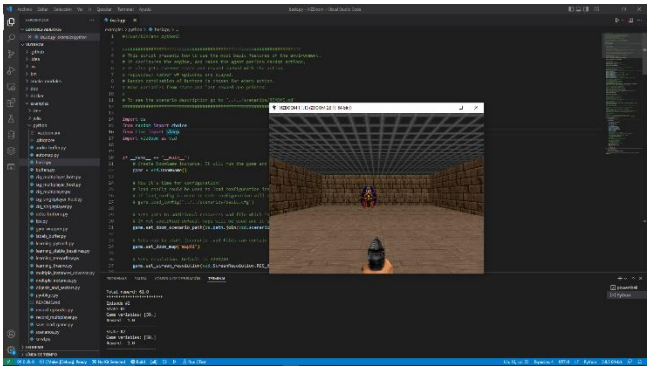
4. Análisis de los Experimentos

Sistema Operativo: Windows 10
Procesador: Intel(R) Pentium(R) CPU G4560 @ 3.50GHz 3.50 GHz
RAM: 12.0 GB
Tarjeta Gráfica: Radeon RX550/550 Series de 2.0GB





1. Experimento 1
   Se realizaron 10 ejecuciones de este experimento de las cuales se obtuvo como tiempo promedio aproximadamente 6.5s para alcanzar el objetivo del Cacodemon con un puntaje de 61 puntos en promedio.



2. Experimento 2

   Fue en este experimento que se encontraron serias dificultades con respecto al otro experimento ya que se obtenían valores demasiados extensos para los tiempo de encontrar los medikits y por lo tanto de llegar al puntaje total.