UNIVERSIDAD

PANAMERICANA

Proyecto final
Q-Learning
(Aprendizaje reforzado)

Por:

Joaquín Alarcón	0189971
César Rodríguez	0197314

Profesor:	León Felipe Palafox
Materia:	Inteligencia Artificial
Ciclo:	Invierno 2019 (1198)

Introducción:

La inteligencia artificial es una herramienta muy importante implementada en medios digitales y software para toma de decisiones basadas en experiencias.

En este proyecto de inteligencia artificial se entrena a un agente por aprendizaje reforzado para que logre cumplir con una meta de la manera mas eficiente.

Objetivo:

Implementar herramientas y conocimientos aprendidos durante el semestre en la clase de inteligencia artificial, usando agentes y ambientes programados en Python.

Herramientas utilizadas:

Lenguaje de programación Python.

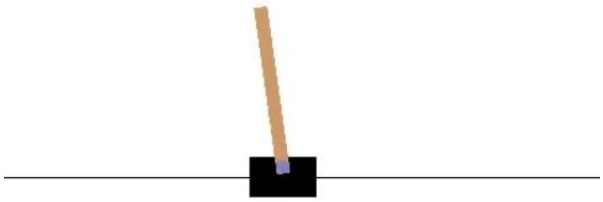
Collaborative (Google)

OpenAI Gym

Desarrollo:

El aprendizaje reforzado es una herramienta que se basa en el entorno y el agente, este último, toma decisiones de realizar determinadas acciones que generan nuevos estados en el entorno.

El problema de CartPole consiste en un mástil unido con una visagra a un carro que se mueve horizontalmente.



[Fig. 1 Imagen del entorno 'CartPole-v0']

Dicho mástil, o péndulo comienza en posición vertical, y el objetivo de dicho problema, es que se mantenga en dicha posición, para esto, se le otorga una recompensa cuando el péndulo se mantiene firme y un castigo cuando el mismo se desplaza mas de 15 grados de su posición vertical, que es cuando termina el entorno.

Nuestro agente consta de dos acciones posible, mover el carro hacia la izquierda o derecha

Para empezar, se configura el ambiente y el agente llamando a las herramientas de OpenAI '*import gym*' y '*gym.make('CartPole-v0')*' y se renderiza '*env.render()*' para poder visualizar a ambos (ambiente y agente).

Una vez que vemos y entendemos el entorno en el que se trabaja para dar una solución al problema, se pone a entrenar al agente, para que, basado en cada fracaso y éxito, obtenga experiencia y aprenda como debe manejarse ante una eventualidad aleatoria.

Para este paso se realiza un ciclo de 100,000 iteraciones para generar la tabla q, que es la fuente en la que se basa el agente para poder determinar qué acción realizar en un futuro, es decir, que el conocimiento que adquiere a partir de las recompensas otorgadas en cada acción, las almacena en dicha tabla y a partir de ella y de las opciones que el entorno le permita realizar, es que toma la decisión de actuar.

La formula para generar los valores dentro de la tabla q es la siguiente:

$$Q(state, action) \leftarrow (1-\alpha)Q(state, action)+\alpha\Big(reward+\gamma \max_a Q(next\ state, all\ actions)\Big)$$

Para esto, se establecen previamente 3 valores, alfa, beta y épsilon, que determinan la importancia que se les da a las partes de la fórmula, la experiencia o el valor futuro.

Después de unos minutos, en los que se cumplen las 100,000 iteraciones, y se llena la tabla q, entonces se realiza la prueba de que el agente haya aprendido de la manera esperada.

Finalmente se vuelve a renderizar para poder visualizar el entorno, en el que el agente realiza sus acciones a partir de lo que aprendió.

Conclusiones:

Se cumplieron con los objetivos estipulados, ya que se logró mantener el péndulo de manera vertical durante un tiempo mayor al que inicialmente permanecía.

Se realizaron tres pruebas, una prueba con 200 episodios donde logramos que el péndulo se mantuviera de manera vertical, después probamos con 500 y 1000 episodios, pero el resultado fue el mismo que en 200. Por lo tanto, el agente no necesita tantos episodios para aprender y lograr la meta.

Este proyecto nos ayudó a comprender como es que un agente aprende, entre más pruebas haga más aprenderá, dependiendo de la complejidad de la tarea es lo que se tardará en aprender, hay agente que se pueden tardar una cantidad de tiempo inmensa. Nuestro agente es sencillo por lo tanto no tarda tanto en aprender. Comprendimos la funcionalidad de una tabla q y como es que el agente la utiliza.