Universidad de San Carlos de Guatemala
Facultad de Ingeniería
Escuela de Ciencias y Sistemas
Inteligencia Artificial 1
Ing. Luis Espino
Aux. Max Florian
1er Semestre 2021

Practica 2

Campo Minado

Objetivos

General

- Aplicar conceptos de Q Learning vistos en el laboratorio.

Específicos

- Utilizar Robocode para visualizar el aprendizaje por refuerzo aplicado a un robot.
- Entrenar un robot con el algoritmo Q Learning.
- Comprender la importancia que tiene el aprendizaje por refuerzo en las distintas situaciones de la vida real.

Descripción

El juego consiste en un robot de uso militar encargado de recuperar un material muy raro denominado por el mismo ejercito como “Elemento 115” que ha ingresado al planeta tierra. Se tiene la información de que parte del elemento se encuentra en un campo al norte en Alemania donde anteriormente se desarrollaron guerras internas por lo que debe evitar caer en las minas que aún se encuentran activas. Para simular dicha situación se hará uso del juego de tanques de batalla “Robocode” y el algoritmo de aprendizaje por refuerzo “Q Learning” para que pueda lograr los objetivos del juego cumpliendo cada una de sus reglas

Información del Juego

Elementos Principales

- Tablero y posición inicial
- Minas
- Robot detector de minas
- Elemento 115 (Meta)

Tablero y posición inicial

- La posición inicial del robot detector de minas siempre es la misma.
- La posición del elemento 115 siempre es la misma.
- La dimensión del tablero o campo de batalla es de 1000 x 800.
- Debido a que la cantidad de pixeles en el tablero es significativamente grande y los estados del agente (robot detector) serían demasiados, se dividirá el tablero en secciones de 100 x 100 pixeles resultando en un tablero de 10x8 secciones.

									(9,7)
								(8,6)	
(0,0)									

En el siguiente ejemplo de tablero los elementos del juego están representados de la siguiente manera:

- Verde: Robot detector de minas
- Rojo: Elemento 15 (Meta)
- Gris: Minas

Minas

- Actualmente no existe el elemento “mina” en Robocode por lo que las minas estarán representadas por coordenadas en el tablero.
- Si el robot cae en una mina ahí se quedará por el resto del juego.
- Para saber la posición de las minas y cuantas minas habrán se deberá leer un archivo de texto llamado “minas.txt” que tendrá el siguiente formato:
    - (mina1X,mina1Y)
      (mina2X,mina2Y)
      (mina3X,mina3,Y)
      …..

    - Ej:
      (1,2),
      (7,3)
      (5,5)

Elemento 115 (Meta)

- El elemento 115 estará representado por un robot enemigo en la siguiente coordenada: (8,6).

Robot detector de minas

Este será el agente al que se tendrá que entrenar con el algoritmo de aprendizaje por refuerzo Q Learning.

Acciones

- Arriba: El robot se desplaza 1 posición del tablero hacia arriba.
- Abajo: El robot se desplaza 1 posición del tablero hacia abajo.
- Izquierda: El robot se desplaza 1 posición del tablero hacia la izquierda.
- Derecha: El robot se desplaza 1 posición del tablero hacia la derecha.
- Disparar o Escarbar elemento: El robot usara su disparo para “desenterrar” el elemento 115 de su ubicación en el campo.

Estados

Los estados del robot (agente) están conformados por su posición en el campo de batalla.

$$\boldsymbol{S1 = (0, 0)}$$

$$S2 = (1, 3)$$

$$S1 = (4, 2)$$

Si se considera que se debe agregar más información al estado para mejorar el aprendizaje del agente, la puede agregar.

Recompensa

La función de recompensa es la que le indicara al agente que tan favorable fue haber realizado una acción especifica en un estado determinado. Esta depende de los siguientes factores:

- $Rwin$ si el robot ha logrado encontrar el elemento 115 se le recompensara positivamente.
- $Rtiempo$ por cada turno/ronda que pase sin encontrar el elemento, el robot será penalizado con una recompensa negativa (esto ayudará a que sea más veloz en su búsqueda).
- $R\text{mina}$ por cada turno que el robot pase ocupado con una mina recibirá una recompensa negativa.
- $Rdistancia$ el agente recibirá una penalización que será directamente proporcional a la distancia que existe entre él y el elemento 115 (esto le enseñará a acercarse al enemigo).
- $Rmuro$ si el robot choca con un muro, recibirá una recompensa negativa.

La recompensa total está dada por la siguiente expresión:

$$R = Rwin + Rtiempo + R\text{mina} + Rdistancia + Rmuro$$

Los factores que siempre se aplicaran a la recompensa son los de tiempo y distancia. Los demás solamente cuando aplica. Por ejemplo, si el robot no se topa con una mina no se aplicará el factor $R\text{mina}$.

Si se considera que se debe agregar más factores a la función de recompensa para mejorar el aprendizaje del agente, los puede agregar.

QTable

Corresponde a la estructura de datos donde se almacenan los valores Q que se han calculado en el proceso de aprendizaje del algoritmo. Para no perder las experiencias obtenidas durante un intento (episodio), al final del juego los valores de la tabla se almacenarán en un archivo de texto con el formato “<carnet>. tabla”. De este modo cuando el agente comience un nuevo juego no comenzara desde cero.

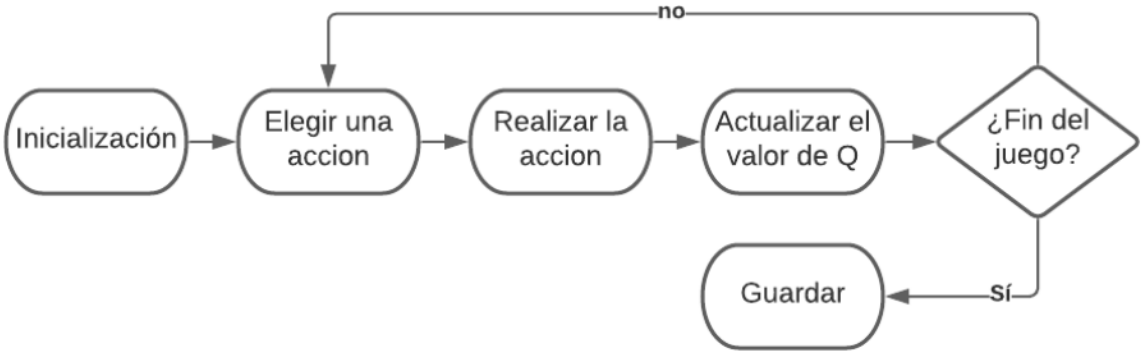
Hiper parámetros

Son valores que afectan el proceso de aprendizaje del algoritmo. Estos valores se leerán desde un archivo que tendrá el siguiente formato “<carnet>.par”:

- Alpha: tasa de aprendizaje
- Gamma: tasa de descuento
- Turnos Máximos: Turnos máximos que pueden transcurrir antes de terminar el juego.
- Modo: Indicará si se entrena o se juega.

En el mismo archivo se almacenará el valor de la probabilidad de exploración ϵ. De igual forma las constantes que el estudiante considere necesarias para controlar el comportamiento de la probabilidad.

Flujo



Inicialización

En esta parte se realizan acciones que solo se ejecutaran una vez, entre las cuales podemos mencionar:

- Crear la Qtable (inicializarla o recuperar los valores).
- Escanear la posición del elemento 115 (útil para la recompensa negativa de la distancia).
- Inicializar hiper parámetros.
- Inicializar variables.

Elegir acción

Dependiendo de si se está explorando el conjunto de estados o se elegirá en base al conocimiento obtenido de las experiencias previas.

Realizar la acción

Se realiza la acción elegida. La acción puede llevar al agente a un nuevo estado o dejarlo en el mismo estado.

Actualizar el valor Q

Se calcula la recompensa obtenida tras realizar la acción y se modifica el valor Q en la tabla.

Fin del juego

Se evalúa si se termina el juego. Esto solamente puede ocurrir si se lograron los objetivos del juego o si se llegó a la cantidad de turnos máximos.

Guardar

Se almacenan los valores Q de la tabla, y se almacenan los hiper parámetros.

Consideraciones

- La práctica se realizará en parejas. De no contar con una lo podrá realizar de forma individual pero no cambiará la exigencia de la práctica.
- Durante la calificación les estaré haciendo preguntas sobre como realizaron la práctica para verificar que esta sea de su autoría.

Entregables

- Archivo .java que contiene el código fuente del robot con el formato “<carnet1_carnet2>.java”.

Fecha de entrega: 28 de marzo de 2021