Evolucion de obligatorio IA

V1

* El algoritmo funciona y guarda los valores en la qtable considerando todas las dimensiones de la observación (velocidad angular, posición del carro, velocidad del carro, angulo del palo)
* Se comienza a partir de la versión 2 con el análisis de entrenamiento

V2

* Se utilizan valores
  + Learning rate 0.1
  + Discount 0.85
  + Cantidad de episodios 10000000000 (diez mil millones)
  + Discretización de los espacios de observación en 50 buckets (100 exceden la memoria utilizable por la ejecución)
* Se suspende el entrenamiento a los 80 millones de episodios
* Utilizando la ultima q table generada no se obtiene un reward promedio en el desempeño de una simulación de 100 episodios superior al 9
* Se intentará reducir el tamaño de la discretización de las observaciones

V3

* se disminuye la discretización a 20
* se decide realizar iteraciones de entrenamiento menos extensas que en la versión 2 para que los pasos sean mas cortos y significativos (de 1 a 10 millones de episodios)

V4

Se corren 10 millones de episodios, pero no se observa un entrenamiento adecuado del modelo ya que en promedio los rewards continúan siendo bajos.

Se incluye un código para generar métricas de rewards máximas obtenidas, mínimas y promedio en base a los episodios generados y se guardan las qtables sincronizadamente de forma que observando el grafico pueda se posible utilizar la qtable asociada al mejor desempeño.

Grafica

Gráfico, Gráfico de líneas, Histograma

Descripción generada automáticamente

Se corre una iteración de 100 episodios con la qtable guardada al inicio, a los 100 mil y a los 200 mil episodios y con la ultima qtable generada. En todos los casos se observa que el reward promedio es 10.

Impresión de pantalla luego de correr una simulación con 100 episodios. La impresión corresponde al ultimo episodio

Texto

Descripción generada automáticamente

Para la versión 4 se intentará aumentar el learning rate y disminuir a 0.5 el valor utilizado para el Discount (representando el valor asignado a la importancia de los episodios futuros frente a los actuales).

Adicionalmente se verificará que la qtable esté inicializándose, calculándose y guardándose correctamente. Se observa que ambas acciones tienen la misma “consecuencia” para nuestro modelo dado que tienen el mismo valor asociado.