Hola a todos,

Para esta actividad, has de comentar al menos una vez en este hilo, respondiendo a las cuestiones planteadas por el ponente Borja González con relación a su seminario.

La fecha límite para participar en esta actividad es el **17 de mayo de 2021.**

Gracias

**Actividades**

1.  ¿Cuales son las tres principales dificultades del DRL en conducción autónoma?

2.  Enumera técnicas para cubrir dichas debilidades y explica cómo funciona cada una

3.  Tenemos que entrenar a un coche que debe aprender a cruzar una rotonda con cuatro salidas por la salida número dos. ¿Cómo definirías la función de recompensas? Asume que el coche solo tiene las acciones acelerar hacia delante, frenar o "coger la salida", esta última acción sólo puede tomarla si el coche está contiguo a una salida de la rotonda.

4.  Define el problema anterior como un POMDP

5.  Open AI gym tiene muchos entornos de terceros muy útiles como <https://github.com/eleurent/highway-env> . Sigue el tutorial en colab de Model-Based RL para el entorno de parking <https://github.com/eleurent/highway-env/tree/master/scripts> ¿Que ventajas y limitaciones observas en esta implementación de Model-based RL?

Solución

1. ¿Cuáles son las tres principales dificultades del DRL en conducción autónoma?

* **Diseño de la función de recompensas.**
* **Proveer garantías de seguridad**.
* **Transferencia del entorno de simulación al real**.

1. Enumera técnicas para cubrir dichas debilidades y explica cómo funciona cada una

* Behavioural Cloning:

Se le enseña al coche una seria de trayectorias que los humanos han realizado y le aplico aprendizaje supervisado para que el sistema aprenda, el problema es que se necesitarían muchas trayectorias en muchos escenarios, adicional, estaríamos limitados a que la mejor conducción que se puede aspirar es a la realizada por el ser humano.

* Inverse Reinforcement learning

La idea de esta técnica es no imitar al humano, si no observar que esta logrando el humano, para hacerlo mejor o mínimamente igual.

Se obtiene al aplicar la intersección de la función de recompensa entregada por el algoritmo RL al algoritmo IRL para que busque la forma de optimizarla dependiendo del comportamiento de los expertos.

* Generative Adversarial networks

Se tiene una red neuronal, a la que se le ingresa una trayectoria generada por otra red o un experto y el discriminador debe acertar el mayor numero de veces cuando el imput es de un humano o de una maquina y por el otro lado está el generador tratando de engañar al discriminador.

* Constrained Reinforcement learning

Se basa en Markov Decision Process (MDP) aumentados con restricciones que reducen el espacio de políticas posibles.

En vez de recompensas, tenemos set de funciones de costes que el agente busca minimizar.

* Formal Methods in Reinforcement learning
  + Model Checking

Es un Framework donde se ingresan una seria de trayectorias o cadenas de Markov, el framework busca de forma exhaustiva si existe en el algoritmo algún comportamiento que va a causar que alguna regla necesaria no se seguirá.

Toma el estado actual y el busca todas las ramificaciones, es muy costoso e imposible de llevar a escenarios complejos.

* + Temporal Logic

Un tipo de lógica simbólica no tiene ambigüedad y su semántica esta muy bien definida, sintácticamente es muy fácil de componer, no se define función de recompensas si no que se define una formula lógica compleja que agrupa todo lo que se desea tener en cuenta y luego aplicar Model Cheking para validarla.

3.  Tenemos que entrenar a un coche que debe aprender a cruzar una rotonda con cuatro salidas por la salida número dos. ¿Cómo definirías la función de recompensas? Asume que el coche solo tiene las acciones acelerar hacia delante, frenar o "coger la salida", esta última acción sólo puede tomarla si el coche está contiguo a una salida de la rotonda.

**4**

**1**

**Auto**

**2**

**3**

Función Entrenamiento()

Acelera\_adelante()

Frena()

Analizar\_Si\_esta\_Cerca\_a\_Salida()

Si Cerca\_salida

Coger\_salida()

Si Salida = 2

Forward = 1

Si no

Forward = -1

Fin si

Reset\_recorrido ()

Entrenamiento()

Si no

Entrenamiento()

finsi

Fin funcion

4.  Define el problema anterior como un POMDP

Un POMDP es una tupla M = < S, A , Z , O , R , P , y > donde :

S -> Frenado, en marcha

A -> Acelerar\_hacia\_delante , frenar , coger\_salida

P -> S x A x S -> [0,1]

R -> Funcion de recompensas [-1,1]

Z -> set de observaciones posibles

O -> S x A -> Z

5.  Open AI gym tiene muchos entornos de terceros muy útiles como <https://github.com/eleurent/highway-env> . Sigue el tutorial en colab de Model-Based RL para el entorno de parking <https://github.com/eleurent/highway-env/tree/master/scripts> ¿Que ventajas y limitaciones observas en esta implementación de Model-based RL?

Se inicia con una tarea de control, donde el agente conduce el automóvil y debe estacionarlo en un lugar determinado con el rumbo apropiado, al ser un model-free agrega un nivel significativo de complejidad.

En el primer intento, con acciones aleatorias, el vehículo no esta ni cerca de lograr estacionar en el lugar indicado, lo que indica que un model-free no es conveniente para solucionar esta situación.

A continuación, se realiza una prueba con model-based produciendo un batch de experiencias

Se realiza un ajuste al modelo con descenso del gradiente estocástico y se puede observar que el overfiting es totalmente nulo, adicional, es un modeo totalmente confiable debido a que las curvas de training y validation tienen el mismo comportamiento

Luego, para mejorar la certeza se utiliza un algoritmo de optimización basado en muestreo (CEM), este debe encontrar la secuencia de acciones con mejor rendimiento.

Al utilizar el model-based, reemplazaron el problema original de optimización con otro problema, conocido como el problema de model bias.

Como se menciono en numeral 1, se debe tener mucho cuidado en la formulación de recompensas, esto debido a que el modelo podría divergir y alucinar con recompensas importantes o altas y el agente intentara alcanzarlas tomando acciones que pueden resultar peligrosas.