

## Proyecto 3 - DoomRL



Reinforcement Learning es un tipo de aprendizaje automático que no requiere de labels para el entrenamiento, sino un entorno con el cual interactuar para aprender usando un premio o reward. Este premio dirige el aprendizaje, más específicamente el gradiente, en el caso de Policy Gradient.

El concepto de Policy Gradient es muy poderoso, nos permite crear un modelo diferenciable que se entrena a partir del reward, sin necesariamente proveer al

algoritmo información a priori sobre el entorno, más allá que una función de reward que define si las acciones del agente fueron útiles en el entorno. En este proyecto crearemos un modelo de Deep Reinforcement Learning usando como entorno el juego de Doom. La red será nuestro policy y aprenderá sola a jugar Doom a punta de prueba y error.

### Policy

Crearemos dos redes neuronales haciendo uso de PyTorch sus diversas capas. Diseñar cada red es parte del trabajo del estudiante, una debe ser el caso base y la otra debe tener ideas mas elaboradas con tal de mejorar el rendimiento. Pueden hacer uso de todo el conocimiento adquirido en el curso para hacerlo, es decir, pueden aplicar desde MLP hasta Transformers, CNNs, RNNs, y en general una combinación de todo eso. Lo importante es que el modelo de policy sea diferenciable. El modelo de policy gradient toma decisiones de cual acción tomar, dado el estado actual del agente, el reward, y las acciones anteriores.

### Pre-procesamiento

Es importante pre-procesar las imágenes antes de meterlas como input al modelo. Por ejemplo es posible resaltar aquellos pixels que realmente nos interesan, como enemigos u otros.

Se recomienda además hacer un resize a la imagen para evitar trabajar con imágenes de alta resolución, en caso de que lo vean necesario. También binarizar la imagen puede servir, o al menos pasarla a grises, aunque tambien puede dejarse a color si se cuenta con buen GPU. Adicionalmente, se recomienda hacer uso de imágenes que sean la diferencia entre la actual y la anterior, para poder capturar movimiento, o bien usar varios frames si se modelan secuencias.

### Entorno y Aprendizaje

Usaremos el entorno de VizDoom que se se puede bajar e instalar aquí:

<https://github.com/mwydmuch/ViZDoom>

Este entorno permite obtener los frames del juego para usarlos como entrada al modelo de policy gradient. Es importante aprender a usar ViZDoom y acceder el reward que definen para este juego en particular. Normalmente el reward cambia dependiendo de si se mata un enemigo, si se pierde un episodio, incluso por cada unidad de tiempo que pasa durante el juego. También pueden averiguar otros rewards que se puedan usar en el caso particular de Doom. El aprendizaje será usando Policy Gradient, calculando la gradiente del modelo usando el reward, definido por episodios. El aprendizaje será entonces una adaptación de supervised learning a un problema de reinforcement learning.

Es importante recordar que deben guardar los pesos de la red cada cierto tiempo para poder continuar entrenando ya que son procesos largos. Además, deben tener alguna forma de medir cuánto va mejorado el modelo con el paso del tiempo.

## Experimentos

Se espera que hagan 2 experimentos, y discutan sobre los resultados obtenidos:

1. Preprocesamiento versus sin preprocesamiento con la base: Probar cómo el preprocesamiento afecta el aprendizaje.
2. Comparar ambas arquitecturas.

## Documento en Latex / PDF

Deben crear en latex (y entregar los fuentes de latex) un documento donde describen los métodos usados así como los experimentos. El documento debe tener la estructura de un paper normal (abstract, introducción, trabajo relacionado, metodología, experimentos, resultados, conclusiones).

El documento debe seguir el template de la IEEE para publicaciones en ingeniería el cual pueden encontrar aquí:

<https://www.ieee.org/conferences/publishing/templates.html>

## Evaluación

Tarea	Puntaje Máximo
Diseño y programación del modelo de Policy Gradient	20
Episodios, entrenamiento	25
Paper	20
Experimento 1 (ejecución, discusión, gráficos)	15
Experimento 2 (innovación, ejecución, discusión, gráficos)	20