Instituto Tecnológico de Costa Rica
Escuela de Ingeniería Electrónica
Trabajo Final de Graduación
Proyecto: Método basado en aprendizaje reforzado para el control automático de una planta no lineal.
Estudiante: Oscar Andrés Rojas Fonseca
I Semestre 2024

Firma del asesor

Bitácora de trabajo

Fecha	Actividad	Anotaciones	Horas dedicadas
08/04/2024	**1**. Prueba de adaptación del código *CartPoleDQN.ipynb* para manejo de variables continuas.	*a*) Se buscaron opciones para la sustitución de la función *torch.gather*() utilizada en el código original.	4 horas
09/04/2024	**2**. Pruebas de implementación *CUDA* en Windows.	*a*) Reinstalación de paquetería *CUDA* 12.1 y librería *PyTorch* 2.2.2. *b*) La implementación *CUDA* fue exitosa, reduciendo importantemente los tiempos de entrenamiento.	4 horas
09/04/2024	**3**. Pruebas de entrenamiento del modelo *Pendulum DQN*.	*a*) Se implementó una primera versión de la adaptación del código original para el manejo de variables continuas. *b*) Entrenamiento de modelo controlador de hasta 600 episodios.	4 horas
10/04/2024	**4**. Reunión de seguimiento con el asesor del proyecto.	*a*) Revisión de avance en el código y errores de forma. *b*) Se acordó realizar entrenamientos con diferentes formatos de indicación del *target_angle*.	2 horas

11/04/2024	**5**. Corrección de potenciales errores en el código *PendulumDQN* señalados por asesor.	*a*) Replanteo de función de recompensa *calculate_reward*() para evitar salto. *b*) Adición de lógica para guardado de *checkpoints* al entrenamiento y corrección del guardado del modelo.	8 horas
12/04/2024	**6**. Continuación de corrección de errores potenciales en el código.	*a*) Replanteo de función *select_action*(); cambio de acción aleatoria en exploración a adición de ruido a la opción elegida.	4 horas
12/04/2024	**7**. Estudio de conceptos *MDP* [1] y *DQN* [2] .	*a*) Revisión de aplicación mediante *MDP* dada la mensión en una fuente en línea donde se utiliza [3]. *b*) Estudio de teoría *DQN* para mejor comprensión de la lógica de la función *optimize_model*() del código original [4] y su adaptación a *Pendulum*.	4 horas
12/04/2024	**8**. Pruebas de entrenamiento de modelos *CartPole* y *Pendulum*.	*a*) Se crearon los cuadernos *ctrlCartPoleDQN.ipynb* y *ctrlPendulumDQN.ipynb* para pruebas de carga de modelos. *b*) Entrenamiento del modelo *Pendulum_1000eps.pth*. *c*) Se descubrió un error grave en *select_action*(), corrección en proceso.	6 horas
		Total de horas de trabajo:	36 horas

Contenidos de actividades

Primeros entrenamientos con $PendulumDQN$

La primera versión de la adaptación del código a variable continua permitió los primeros entrenamientos del modelo controlador de $Pendulum$, donde la implementación del procesamiento con $CUDA$ permitió la disminución del tiempo de entrenamiento a aproximadamente a la mitad del tiempo de procesamiento con CPU, permitiendo los entrenamientos de hasta 600 episodios en 20 minutos y posteriores 1000 episodios en aproximadamente 40 minutos.

Los modelos anteriormente mencionados, a pesar de las comparaciones con el modelo exitoso entrenado de $CartPole$ a 600 episodios, no presenta mejoría en las pruebas realizadas, por lo que se procede a plantear una nueva forma de la adaptación al manejo de valores continuos en la función $optimize\_model()$.

Estudio de MDP y DQN

Se requirió una revisión de la teoría que sustenta al método DQN en [2] para comprender a mayor profundidad el proceso que realiza la función $optimize\_model()$, de manera que la base de la técnica de $Q-learning$ también fue estudiada.

Además, la presencia del proceso de decisión de Markov (MDP) en algunas de las revisiones de las opciones de implementación de variable continua al método DQN [3], requirió un análisis respectico del algoritmo, por lo que se estudió en [1], también a manera de contextualización al tema del $Q-learning$.

Referencias

[1] J. P. A. Moya, "EL5857 Lección 25: Aprendizaje Reforzado (1/4): MDP," 2021, [Vídeo de YouTube]. [Online]. Available: https://www.youtube.com/watch?v=FBaoss_Pb5Q

[2] ——, "EL5857 Lección 27: Aprendizaje Reforzado (3/4): DQN y Q-Learning," 2021, [Vídeo de YouTube]. [Online]. Available: https://www.youtube.com/watch?v=oXnNRSCe5T4

[3] S. Israilov, L. Fu, J. Sánchez-Rodríguez, F. Fusco, G. Allibert, C. Raufaste, and M. Argentina, "Reinforcement learning approach to control an inverted pendulum: A general framework for educational purposes," *PLoS ONE*, 2023.

[4] A. Paszke and M. Towers, "Reinforcement learning (dqn) tutorial," *PyTorch*.