Instituto Tecnológico de Aeronáutica - ITA
Inteligência Artificial para Robótica Móvel - CT-213
Aluno: Rafael Mello Celente

Relatório do Laboratório 12 - Deep Q-Learning

1. Breve Explicação em Alto Nível da Implementação

O código teve por objetivo implementar uma solução de Deep Q-Learning para a solução de um problema do OpenAI Gym, em que buscamos fazer um carrinho, inicialmente no ponto mais baixo de um vale, a aprender a melhor forma de chegar em um objetivo no topo do vale.

Como não sabemos a dinâmica do evento, a estratégia aplicada tem por objetivo estimar a função ação-valor do método da Q-Learning. Para estimar essa função, utiliza-se uma rede neural profunda, daonde sai o nome da solução Deep Q-Learning.

A implementação utilizou uma rede neural com 1 camada escondida de 24 *neurons*. Com uma discretização de 24 estados e 3 ações (as ações se dividiram em *push left, push right* e *no push*), as camadas de entrada e saída possuíam 24 e 3 *neurons*, respectivamente.

Para a atuação do agente, uma política ε-greedy na estimativa da função $\hat{q}$ foi implementada, de forma que um elemento de exploração seja introduzido. Entretanto, foi feito com que o valor de ε fosse diminuído exponencialmente com uma taxa de decaimento, diminuindo a cada episódio para facilitar o treinamento, uma vez que esperamos que, com o passar dos episódios, o algoritmo se torne cada vez mais próximo do ótimo, portanto não precisa de tanto fator de exploração.

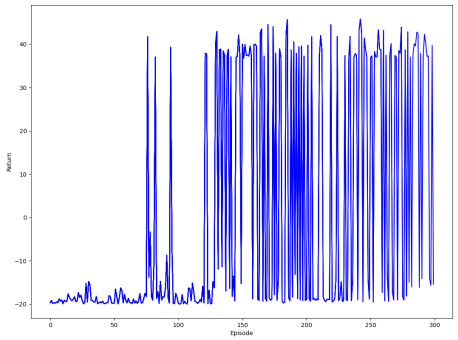
2. Figuras Comprovando Funcionamento do Código

2.1. Sumário do Modelo

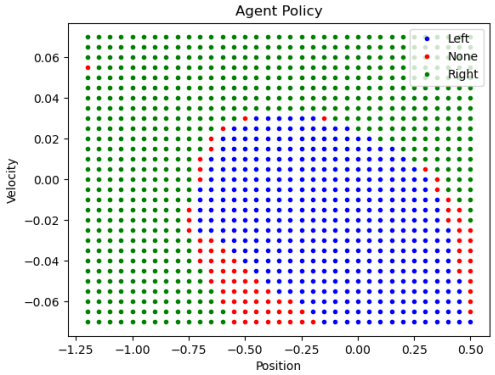
Layer (type)	Output Shape	Param #
dense (Dense)	(None, 24)	72
dense_1 (Dense)	(None, 24)	600
dense_2 (Dense)	(None, 3)	75

Total params: 747
Trainable params: 747
Non-trainable params: 0

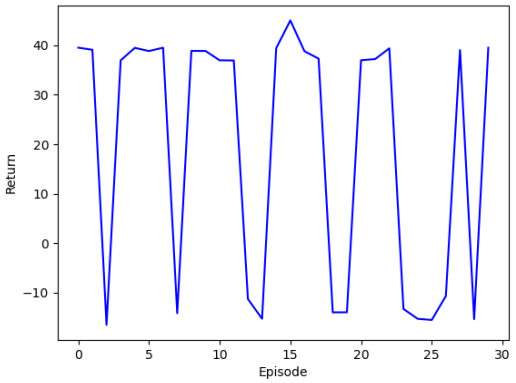
2.2. Retorno ao Longo dos Episódios de Treinamento



2.3. Política Aprendida pelo DQN



2.4. Retorno de 30 Episódios Usando a Rede Neural Treinada



3. Discussão dos Resultados

A partir dos resultados apresentados, podemos perceber que o código implementado não desempenhou tão satisfatoriamente a tarefa. O carrinho, utilizando o algoritmo de Deep Q-Learning, conseguiu alcançar o objetivo da bandeira em cerca de 63% dos casos de teste, com uma média de retorno de 19,83, o que é relativamente baixo para

um problema brinquedo. Entretanto, parte das tentativas do teste que falharam passaram por pouco do objetivo, o que pode indicar que o algoritmo deveria ter sido treinado por mais tempo e com uma taxa de decaimento menor.