Instituto Tecnológico de Aeronáutica - ITA
Inteligência Artificial para Robótica Móvel - CT213
Aluno: Matheus Felipe Ramos Borges

Relatório do Laboratório 12 - Deep Q-Learning

1 Breve Explicação em Alto Nível da Implementação

O objetivo deste laboratório é implementar um agente de *Deep Q-Learning (DQL)* para resolver o problema clássico de controle conhecido como *Mountain Car*, utilizando o ambiente *OpenAI Gym*. O algoritmo adotado, baseado em redes neurais profundas, estende o Q-Learning tradicional ao incorporar técnicas como *experience replay* e *fixed Q-targets*, promovendo maior estabilidade e eficiência no processo de aprendizado.

A função ação-valor $\hat{q}(s, a)$ é aproximada por uma rede neural densa, cuja arquitetura foi definida conforme a tabela a baixo.

Tabela 1: Arquitetura da rede neural utilizada para aproximar a função ação-valor $\hat{q}(s, a)$.

Camada	Neurônios	Função de Ativação
Densa (Hidden 1)	24	ReLU
Densa (Hidden 2)	24	ReLU
Densa (Saída)	*action_size*	Linear

2 Figuras Comprovando Funcionamento do Código

2.1 Sumário do Modelo

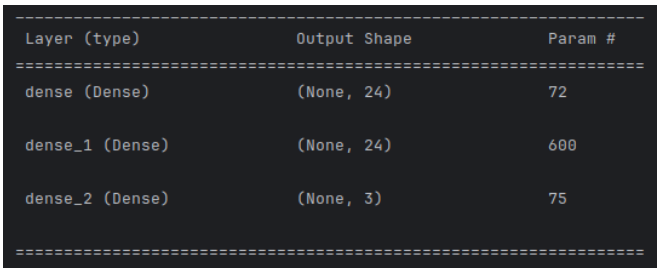

Figura 1: sumário do modelo.

2.2 Retorno ao Longo dos Episódios de Treinamento

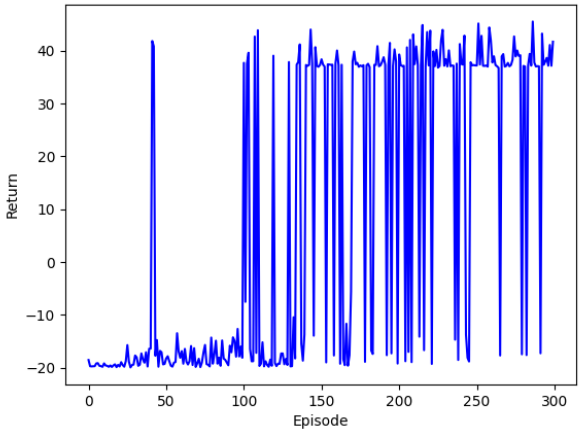

Figura 2: retorno ao Longo dos Episódios de Treinamento.

2.3 Política Aprendida pelo DQN

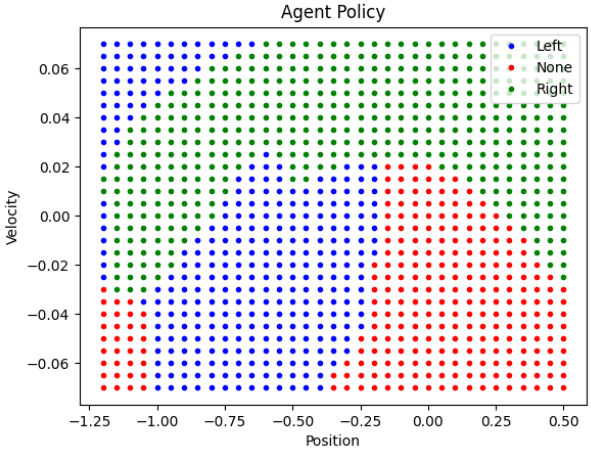

Figura 3: política Aprendida pelo DQN.

2.4 Retorno de 30 Episódios Usando a Rede Neural Treinada

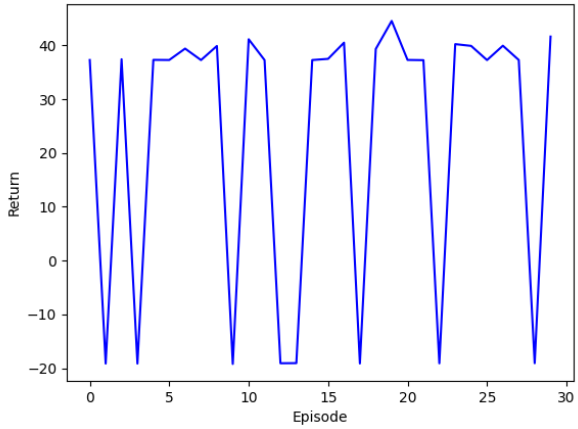

Figura 4: retorno de 30 Episódios Usando a Rede Neural Treinada.

3 Discussão dos Resultados

A implementação do agente utilizando o algoritmo de *Deep Q-Learning* foi bem-sucedida na resolução do problema do *Mountain Car*. A utilização de técnicas como *experience replay* e *fixed Q-targets* foi essencial para estabilizar o treinamento e permitir uma aproximação eficiente da função ação-valor $\hat{q}(s, a)$ por meio de redes neurais densas. A engenharia de recompensas também teve papel fundamental, fornecendo sinais mais informativos ao agente durante o processo de aprendizado e acelerando a convergência para uma política eficaz.

Os gráficos obtidos durante os testes demonstram uma clara evolução no desempenho do agente ao longo dos episódios, com evidências de que a política aprendida é capaz de generalizar para diferentes estados iniciais. A experiência proporcionou uma compreensão prática dos desafios inerentes ao aprendizado por reforço profundo, como o equilíbrio entre exploração e exploração, o impacto da escolha de hiperparâmetros e a sensibilidade ao design das recompensas.