Trabalho de Aprendizado por reforço

1st João Pedro Aguiar Formiga Matos *EMC-UFG*
Universidade Federal de Goiás
Goiânia, Brasil
Joao.formiga@discente.ufg.br

***Abstract*—A aplicação de aprendizado por reforço em problemas de tomada de decisão sequencial representa um desafio intrigante, que tem sido abordado com sucesso graças à capacidade das técnicas de aprendizado de máquina, em especial as redes neurais. Nesse contexto, este estudo tem como objetivo a implementação e avaliação de um algoritmo de aprendizado por reforço para auxiliar um agente a encontrar um tesouro em um grid nxn. O aprendizado por reforço capacita um agente a aprender a tomar ações em um ambiente interativo, visando maximizar uma recompensa cumulativa ao longo do tempo. As estratégias propostas serão avaliadas quanto à sua capacidade de aprender com a interação, adaptar-se a mudanças nas condições do ambiente e tomar decisões otimizadas. O estudo busca fornecer dados e desafios encontrados na realização do trabalho na materia de redes neurais profunda na universidade federal de goias.**

I. Metodologia

A. Introdução

A aplicação de técnicas de aprendizado por reforço em contextos que envolvem a tomada de decisões sequenciais tem sido objeto de extenso interesse, impulsionado pelas notáveis conquistas das redes neurais e do campo do aprendizado de máquina em geral. Essas abordagens proporcionam meios para capacitar agentes a tomar ações em ambientes interativos, visando otimizar a acumulação de recompensas ao longo do tempo. Dentro deste âmbito, o presente estudo direciona-se para a implementação e avaliação de um algoritmo de aprendizado por reforço, com o foco em empregá-lo para orientar um agente na busca de um tesouro em um grid nxn.

A proposta aborda a habilidade de um robô pirata operando em um ambiente modelado por uma grade, a utilizar estratégias de aprendizado por reforço para aprender eficientemente a navegar, minimizando a extensão da trajetória para atingir o tesouro. O ambiente simulado é caracterizado por obstáculos representados como buracos, que requerem esquiva por parte do agente, além do objetivo a ser alcançado, representado pelo tesouro no canto inferior direito da grade. A tarefa se torna desafiadora devido à limitação das informações observadas pelo agente, exigindo a tomada de decisões com base em observações parciais.

Neste estudo, além da implementação do algoritmo REINFORCE, será elaborada uma função de recompensa adequada para o ambiente do robô pirata. A avaliação do desempenho do algoritmo será conduzida através da análise gráfica do retorno médio dos episódios ao longo das épocas de treinamento, assim como da média de passos executados pelo robô em cada episódio durante o período de treinamento. O intuito é analisar a capacidade do algoritmo em aprender uma política de ações que guie o agente eficientemente ao tesouro.

Ademais, este estudo se propõe a investigar a adaptabilidade da estratégia de aprendizado diante das mudanças nas condições do ambiente e sua capacidade de aprender continuamente através da interação. A abordagem de implementação do algoritmo de aprendizado por reforço sem o recurso de bibliotecas específicas busca proporcionar um entendimento aprofundado do processo subjacente.

Dessa maneira, por meio deste trabalho, almejamos contribuir para a área de aprendizado por reforço e redes neurais, oferecendo uma avaliação compreensiva do desempenho do algoritmo implementado, suas limitações e os insights obtidos ao longo do estudo. Em última instância, este trabalho tem como finalidade enriquecer a compreensão das estratégias de aprendizado por reforço em cenários de tomada de decisões sequenciais e suas aplicações práticas.

B. Metodologia

O desenvolvimento da metodologia para a implementação e avaliação do algoritmo de aprendizado por reforço baseou-se na linguagem de programação Python, com o uso das bibliotecas PyTorch e OpenAI Gym. O trabalho foi conduzido em um ambiente computacional com recursos compatíveis para experimentação e análise.

Para a construção do ambiente de simulação, o OpenAI Gym foi empregado, fornecendo um meio flexível e padronizado para a criação de ambientes de aprendizado por reforço. A grade do ambiente, onde o agente robô pirata navega, foi modelada conforme a descrição da proposta, considerando buracos e o tesouro como elementos de interesse. Entretanto, devido a algumas dificuldades iniciais, problemas como tempos de aprendizado prolongados e estagnação do modelo em determinadas situações foram identificados.

Para contornar esses problemas, foram introduzidas paredes no ambiente, a fim de delinear caminhos mais concretos para o agente explorar e aprender. Adicionalmente, os valores de recompensas foram ajustados de maneira estratégica para incentivar o agente a evitar os buracos e buscar o tesouro de forma mais eficiente. Esses ajustes foram realizados iterativamente, considerando observações e insights adquiridos durante a fase de experimentação.

Ao longo do processo, foi observado que o algoritmo de aprendizado por reforço conseguiu adquirir uma habilidade robusta para desviar dos obstáculos e localizar o tesouro quando este estava posicionado no canto inferior direito da grade. Os problemas de estagnação foram mitigados com a introdução das paredes no ambiente, permitindo ao agente explorar de maneira mais eficiente e evitar situações que resultavam em retornos inadequados.

Em resumo, a metodologia adotada envolveu a implementação do algoritmo REINFORCE utilizando a biblioteca PyTorch e o ambiente de simulação do OpenAI Gym. A construção do ambiente de grade foi aprimorada com a introdução de paredes e ajustes nas recompensas, solucionando problemas de estagnação e contribuindo para a robustez do aprendizado do agente. As etapas metodológicas contemplaram a experimentação iterativa, a análise dos resultados e a adaptação estratégica para melhorar o desempenho do algoritmo na tarefa proposta.

C. dados e graficos de treinamento

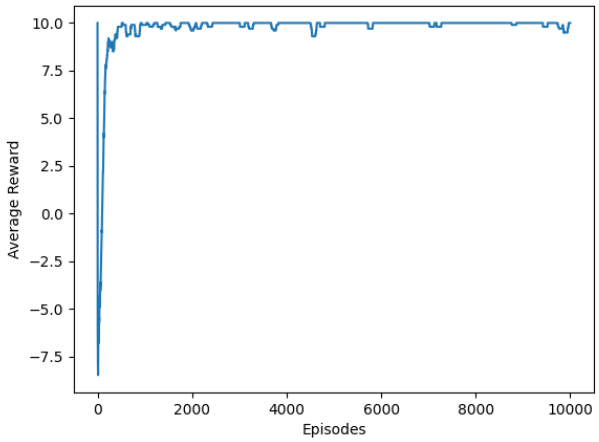

Fig. 1. recompensa media por episodio

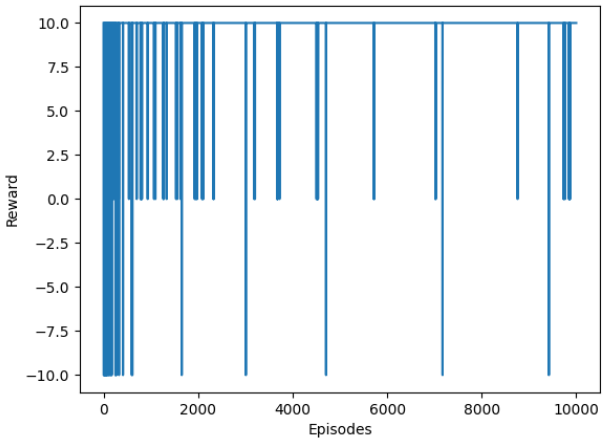

Fig. 2. recompensa por época

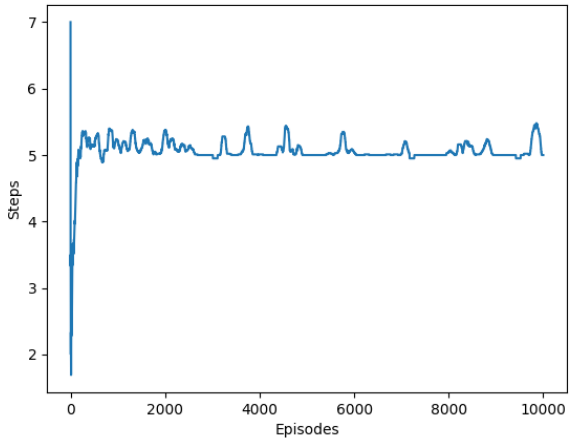

Fig. 3. passos medios por epoca

II. Conclusão

No decorrer deste estudo, exploramos a aplicação do aprendizado por reforço na resolução de um desafio de tomada de decisão sequencial, no qual um agente robô pirata busca um tesouro em um ambiente representado por uma grade. A implementação e avaliação do algoritmo REINFORCE utilizando a biblioteca PyTorch e o ambiente OpenAI Gym permitiram observar resultados promissores e insights relevantes.

Os resultados alcançados evidenciam a capacidade do modelo de aprender a desviar do obstáculo de forma precisa e encontrar o tesouro com eficiência quando este é posicionado no canto inferior direito da grade. O agente demonstrou um aprendizado robusto, evidenciado pelo fato de que, em condições controladas, conseguiu executar a tarefa de maneira confiável e consistente.

Entretanto, quando introduzimos a variabilidade na localização do obstáculo, notamos um desafio mais complexo para o agente. Apesar dos esforços em aumentar o número

de épocas de treinamento, o modelo não conseguiu aprender de forma eficaz a contornar o obstáculo em situações aleatórias. Esse resultado destaca a sensibilidade do algoritmo a mudanças significativas no ambiente e a necessidade de estratégias mais avançadas para lidar com essas variações imprevisíveis.

No panorama geral, este estudo reforça a robustez do modelo em relação ao obstáculo quando este é fixo, mostrando que o agente consegue aprender a navegar de forma eficaz. Por outro lado, também salienta os desafios inerentes à adaptação do aprendizado por reforço a mudanças não previsíveis no ambiente. Esses achados abrem caminho para futuras pesquisas que abordem estratégias mais avançadas de aprendizado para lidar com obstáculos variáveis.

Em resumo, este estudo contribui para o entendimento das capacidades e limitações do aprendizado por reforço em cenários de tomada de decisão sequencial, fornecendo insights valiosos sobre a adaptação do modelo a variações do ambiente. A busca por soluções que permitam ao agente lidar eficazmente com obstáculos variáveis permanece como um tópico interessante e relevante para investigações futuras.