Trabalho Prático: Aprendizado por Reforço com Gymnasium

Disciplina de Inteligência Artificial

Professor(a): Ricardo de Andrade Lira Rabelo

Objetivo Geral

O objetivo deste trabalho é aplicar os conceitos de **Aprendizado por Reforço (Reinforcement Learning – RL)** na formulação, experimentação e análise de um problema modelado como um **Processo de Decisão de Markov (MDP)**. O aluno deverá compreender profundamente cada etapa do ciclo de aprendizado, desde a definição do problema até a avaliação dos resultados e da convergência do agente.

1 Tarefas Principais

O trabalho será desenvolvido em quatro etapas obrigatórias e interdependentes.

1.1 Etapa 1 – Formulação do Problema como MDP

O aluno deve formular o problema escolhido na forma de um **Processo de Decisão de Markov (MDP)**. A formulação deve conter e explicar claramente cada um dos elementos fundamentais:

- **Conjunto de estados (S):** descreve todas as possíveis situações em que o agente pode se encontrar.
- **Conjunto de ações (A):** define as decisões possíveis que o agente pode tomar em cada estado.
- **Função de transição ($P(s'|s,a)$):** especifica a probabilidade de transitar de um estado s para outro s' após a ação a.

- **Função de recompensa ($R(s, a)$):** determina o valor (positivo ou negativo) recebido após uma ação.
- **Fator de desconto (γ):** controla a importância das recompensas futuras.

O aluno deve justificar a escolha do problema e discutir como cada componente contribui para o comportamento esperado do agente.

1.2 Etapa 2 – Escolha ou Construção do Ambiente

Nesta etapa, o aluno deve escolher um ambiente já existente no **Gymnasium** ou desenvolver um ambiente próprio. É fundamental que a escolha seja **coerente com o MDP formulado** e que o funcionamento do ambiente seja descrito de forma detalhada. A descrição deve incluir:

- **Propósito do ambiente:** qual é o objetivo principal da tarefa?
- **Funcionamento:** como as ações afetam os estados e recompensas?
- **Objetivo do agente:** qual comportamento representa sucesso?

Caso o aluno opte por criar um ambiente, é importante descrever suas regras e limitações de maneira clara, explicando como ele pode ser utilizado para testar algoritmos de aprendizado por reforço.

1.3 Etapa 3 – Explicação do Processo de Treinamento

Nesta etapa, o aluno deve descrever e justificar o processo de **treinamento do agente**. A explicação deve abranger:

- **Algoritmo de aprendizado:** qual método foi utilizado (por exemplo, Q-Learning, SARSA, DQN, etc.) e por que foi escolhido.
- **Parâmetros principais:** valores de taxa de aprendizado, fator de desconto, taxa de exploração e número de episódios.
- **Critério de parada:** como o aluno determinou quando o treinamento seria encerrado (número de episódios, estabilização de recompensa, etc.).
- **Dinâmica do aprendizado:** como o agente atualiza seus valores, políticas ou redes neurais ao longo do tempo.

O aluno deve explicar o raciocínio por trás das escolhas feitas, relacionando o processo de treinamento com a teoria do aprendizado por reforço.

1.4 Etapa 4 – Análise e Visualização dos Resultados

Após a execução do experimento, o aluno deverá apresentar e discutir os resultados obtidos. A análise deve incluir, obrigatoriamente, representações visuais, tais como:

- Gráficos de recompensa média por episódio.
- Curvas de desempenho ao longo do treinamento.
- Comparações entre políticas ou diferentes configurações de parâmetros.

O relatório deve enfatizar o comportamento do agente durante o aprendizado, destacando se há progresso, estagnação ou regressão. Além disso, o aluno deve discutir possíveis causas para eventuais falhas ou comportamentos inesperados.

1.5 Etapa 5 – Análise de Convergência

Por fim, o aluno deverá analisar a **convergência ou divergência** do algoritmo utilizado. Essa análise deve responder às seguintes questões:

- O agente converge para uma política estável e eficaz?
- As recompensas se estabilizam ao longo dos episódios?
- Quais fatores podem ter dificultado a convergência (excesso de exploração, taxa de aprendizado inadequada, ambiente estocástico, etc.)?
- Em caso de divergência, quais ajustes poderiam ser feitos para melhorar o aprendizado?

A reflexão sobre a convergência é essencial para demonstrar o entendimento profundo do funcionamento dos algoritmos de aprendizado por reforço.

2 Entregáveis

O trabalho final deve conter:

- Um **relatório técnico** em PDF, contendo as quatro etapas descritas.
- Um **arquivo de código** (Python ou Jupyter Notebook) com a implementação utilizada, devendo ser submetido com instruções de execução.

O relatório deve ser redigido de forma clara, objetiva e com fundamentação teórica, abordando tanto o processo de modelagem quanto a análise dos resultados.

3 Recomendações

- Utilize ambientes simples para testes iniciais, como `FrozenLake-v1`, `CartPole-v1` ou um ambiente próprio de pequena escala.
- Ajuste parâmetros como taxa de aprendizado, fator de desconto e taxa de exploração para observar diferentes comportamentos.
- Registre as recompensas e métricas ao longo do tempo para facilitar a análise.
- Apresente os resultados com gráficos legíveis e interpretações coerentes.
- Lembre-se: entender as causas da convergência ou divergência é tão importante quanto alcançar bons resultados.