CONTROLE DE MANIPULADORES FLEXÍVEIS ATRAVÉS DE APRENDIZADO POR REFORÇO EM AMBIENTE SIMULADO

Marcelo Albergaria Paulino Fernandes Ferreira[1]; Taniel Silva Franklin [2]; Oberdan Rocha Pinheiro [2]
[1] Bolsista; Pesquisa e Desenvolvimento e Inovação (PD&I); marcelo.ferreira@fbter.org.br;
[2] Centro Universitário SENAI CIMATEC; Salvador-BA; {taniel.franklin, oberdan.pinheiro}@fieb.org.br;

RESUMO

Os manipuladores flexíveis, conhecidos como *Soft Robots*, inovam na robótica ao adotar materiais flexíveis, afastando-se da rigidez dos robôs convencionais. Além da flexibilidade para adaptação a diferentes formas e ambientes, os robôs moles proporcionam maior segurança e eficiência energética devido a sua construção leve e menor risco de danos em caso de colisões com humanos ou objetos. Através do aprendizado por reforço, os robôs podem adaptar-se a novas situações e aprender com a interação direta com o ambiente, tornando-se cada vez mais eficientes e autônomos em suas tarefas. Este trabalho apresenta ambiente de simulação que utiliza um modelo dinâmico baseado em curvatura constante por partes (*Piecewise Constant Curvature*) para descrever o movimento do robô no espaço, com o objetivo de resolver um problema clássico de cinemática inversa de manipuladores. O potencial desse método de aprendizado por reforço é ilustrado através de simulações utilizando um manipulador de três segmentos treinado com algoritmo Deep Q-Network.

PALAVRAS-CHAVE: Aprendizado por reforço; Robótica; Manipuladores Flexíveis; Redes Neurais;

1. INTRODUÇÃO

Os robôs flexíveis, ou Soft Robots, representam uma inovação significativa no campo da robótica, afastando-se das características rígidas dos robôs tradicionais para adotar materiais flexíveis e maleáveis [1] . Inspirados na natureza e na biologia, esses robôs são construídos com elastômeros, borrachas, hidrogéis e outros materiais que oferecem flexibilidade e capacidades elásticas [2]. Esta abordagem permite que os soft robots se adaptem a ambientes não estruturados, realizem movimentos complexos, e até mesmo imitem funções biológicas. Ao superar as limitações dos robôs convencionais em termos de rigidez, complexidade estrutural, segurança e eficiência energética os manipuladores flexíveis encontram aplicações em diversas áreas, como bioengenharia, resgate em desastres, cirurgias minimamente invasivas, produção industrial, e exploração em ambientes desafiadores. A pesquisa contínua nesse campo demonstra o potencial desses manipuladores para a interação segura com objetos e seres humanos em ambientes e tarefas complexas.

O surgimento do Deep Reinforcement Learning (DRL) representa uma convergência estratégica entre Reinforcement Learning (RL) e Deep Learning (DL). O Reinforcement Learning é uma abordagem de aprendizado de máquina na qual um agente aprimora seu comportamento em um ambiente por meio de tentativa e erro, com o intuito de maximizar recompensas. Em RL, o agente desempenha o papel de tomar decisões para enfrentar desafios complexos, interagindo com o ambiente, que fornece observações ou estados a partir dos sensores disponíveis, associados com recompensas ou custos [3]. Com o advento de computadores de alta capacidade e a abundância de dados disponíveis, surgiu a oportunidade de treinar modelos capazes de generalizar a partir de entradas, como imagens, textos e voz. A interseção entre RL e DL representa um avanço significativo no campo da Inteligência Artificial, permitindo a abordagem de problemas complexos de uma maneira mais eficaz, permitindo uma melhor generalização do comportamento do agente em contextos variados.

Neste estudo, as Redes Neurais Profundas serão empregadas para modelar o mecanismo de tomada de decisão do agente. A capacidade dessas redes de processar combinações complexas dos estados e aprender as melhores ações para cada possível cenário apresenta-se como uma ferramenta fundamental para modelar ambientes de RL.

Neste projeto procura-se resolver o problema da cinemática inversa de um manipulador flexível composto por três segmentos. O modelo dinâmico do manipulador utiliza simplificação de curvatura constante por partes (PCC), é implementado em linguagem Python e utiliza o sistema operacional Robot Operation System (ROS) [4] para estabelecer uma comunicação entre o agente de aprendizado e o ambiente.

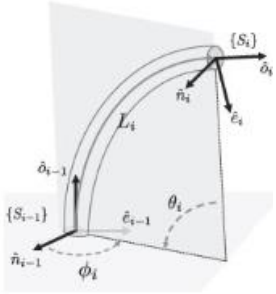
O modelo dinâmico PCC (Piecewise Constant Curvature) é uma abordagem para descrever o formato do robô no espaço, se destacando pela simplificação da representação da geometria do manipulador [5], aproximando cada seção como uma curva de curvatura constante, isso facilita significativamente os cálculos de cinemática e dinâmica e, o que facilita sua simulação numérica.

2. METODOLOGIA

O modelo dinâmico PCC é inicializado com base em um arquivo de configuração, que define as características de cada segmento do robô, o espaço de tarefas e os parâmetros de simulação. Supõe-se que o manipulador

é composto por três segmentos, o movimento é planar, os ângulos iniciais são fixos e a posição do alvo é aleatória. O manipulador de três segmentos possui três ângulos de articulação, consideramos que os ângulos de rotação dos arcos ϕ (Phi) são zero, e os incrementos e decrementos ocorrem nos ângulos de curvatura dos arcos θ (Theta). Essas premissas visam reduzir a dimensionalidade e o tempo de treinamento. A Figura 1, representa a cinemática de um segmento de curvatura constante, onde Li é o comprimento do segmento, theta o ângulo de curvatura e phi a orientação do plano que ocorre a curvatura.

Figura 1: Representação de Segmento de Curvatura Constante



Para o treinamento, foi utilizado o algoritmo Deep Q Networks (DQN) do Stable Baselines 3, uma biblioteca de aprendizado por reforço em Python [6]. O Stable Baselines 3 oferece implementações eficientes de algoritmos de aprendizado por reforço, incluindo o DQN, que combina redes neurais profundas com o Q-learning para aprender uma política ótima de ações em um ambiente complexo. A Rede Neural que representa o agente DQN tem uma arquitetura com três camadas ocultas com 128, 256 e 800 neurônios respectivamente e função de ativação Relu *(Rectified Linear Unit)*, sendo a entrada o número de estados e a saída a combinação de 3 possíveis ações: decremento, incremento e nada a fazer, para cada junta.

O processo de treinamento é iterativo, com o agente interagindo com o ambiente, escolhendo ações, recebendo recompensas e ajustando os parâmetros da rede neural para melhorar continuamente sua política de ação. O objetivo final é alcançar uma política ótima, maximizando a recompensa cumulativa ao longo do episódio no ambiente.

3. RESULTADOS E DISCUSSÃO

Foram conduzidas várias iterações de treinamento com o propósito de alcançar o alvo posicionado na cena. A Figura 2 ilustra a posição inicial do manipulador e o alvo posicionado na cena em cor vermelha. Já a Figura 3, representa o objetivo desejado após o treinamento, marcado pelo alvo em cor verde como indicativo de sucesso alcançado.

Figura 2: Posição Inicial do Manipulador

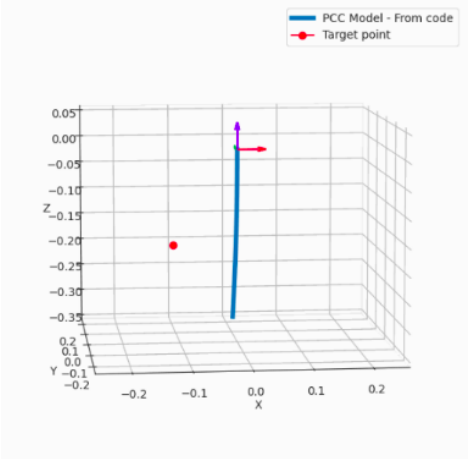
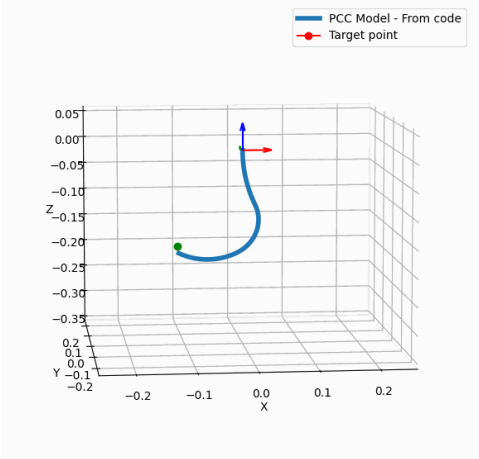


Figura 3: Posição Final do Manipulador



Neste projeto o ambiente foi configurado de forma simplificada, considerando 3 estados e 27 ações possíveis, sendo os incrementos e decrementos executados em três juntas do manipulador, com o objetivo simplificar a simulação, reduzir a dimensão do espaço de trabalho e o tempo de treinamento. O robô foi treinado em um ambiente computacional com os seguintes recursos de *hardware:* um processador core i7 de decima primeira geração, 8GB de memória e uma placa gráfica Nvidia Geforce MX450. O agente foi

submetido 95000 épocas, totalizando 37 horas de duração e alcançou uma taxa de sucesso em torno de 30%, conforme Figura 4.

Figura 4: Taxa de Sucesso do Agente

4. CONSIDERAÇÕES FINAIS

Esta pesquisa proporcionou uma análise sobre a eficácia do algoritmo DQN e do modelo dinâmico PCC na resolução do problema de cinemática inversa. Embora tenha alcançado uma taxa de sucesso em torno de 30%, é fundamental reconhecer as limitações e explorar abordagens mais avançadas para aprimorar o desempenho. Dentre as opções consideradas para estudos futuros e melhorias, destaca-se a adoção do algoritmo Ator-Crítico (A2C) no treinamento de um robô de dois segmentos. Com essa abordagem, será possível diminuir a dimensionalidade do ambiente devido a um menor número de juntas e implementar as vantagens do novo algoritmo, tais como a otimização direta de políticas e atualizações assíncronas para contribuir com a estabilidade no treinamento.

O ambiente de simulação permite o desenvolvimento das habilidades de controle e execução de movimentos do robô, além de oferecer uma plataforma propicia para experimentar ajustes no algoritmo de aprendizado por reforço. Trata-se de uma plataforma virtual com menor risco, controlada e de fácil instalação, sem depender de um robô real, economizando tempo e custos. Em algumas situações, esses aprendizados podem ser transferidos e testados em manipuladores reais após atingir um bom desempenho.

Agradecimentos

Esta pesquisa foi realizada em parceria entre o SENAI CIMATEC e a Shell Brasil. Os autores gostariam de agradecer à Shell Brasil Petróleo LTDA, à Empresa Brasileira de Pesquisa e Inovação Industrial (EMBRAPII) e à Agência Nacional do Petróleo, Gás Natural e Biocombustíveis (ANP) pelo apoio e investimentos em PD&I.

5. REFERÊNCIAS

[1] Liu, K.; Chen,W.; Yang, W.;Jiao, Z.; Yu, Y. **Review of the Research Progress in Soft Robots**. Appl. Sci. 2023, 13, 120. https://doi.org/10.3390/app13010120.

[2] Jin, G.; Sun, Y.; Geng, J.; Yuan, X.; Sun, L. **Bioinspired Soft Caterpillar Robot with Ultra-stretchable Bionic Sensors Based on Functional Liquid Metal**. Nano Energy 2021, 84, 105896.

[3]Data Science Academy. **Deep Learning Book**, 2022. Disponivel em: <https://www.deeplearningbook.com.br/?s=65 > Acesso em: 27. fevereiro. 2024.

[4] S. Macenski, T. Foote, B. Gerkey, C. Lalancette, W. Woodall, **Robot Operating System 2: Design, architecture, and uses in the wild, Science Robotics** vol. 7, May 2022.

[5] Webster RJ, Jones BA. **Design and Kinematic Modeling of Constant Curvature Continuum Robots: A Review**. The International Journal of Robotics Research. 2010.

[6] A. Raffin, A. Hill, A. Gleave, A. Kanervisto, M. Ernestus N. Dormann, **Stable-Baselines3: Reliable Reinforcement Learning Implementations**: Journal of Machine Learning Research vol. 22, 2021.