

ORIENTADOR: GUSTAVO CARLOS BUSCAGLIA

ALUNO: LUCAS PELIZZARI RAYMUNDO

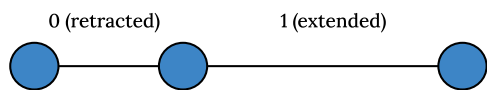
LUCASPELIZZAR@USP.BR

Objetivos

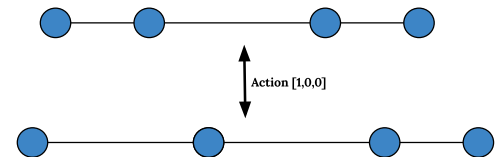
Micronadadores são agentes naturais ou artificiais que possuem capacidade de locomoção através de um fluido, e o fazem ao exercer forças sobre o fluido viscoso no qual estão submersos por meio de partes móveis, como flagelos ou pseudópodos. Em decorrência de suas dimensões reduzidas, os efeitos inerciais são desprezíveis, o que os obriga a exercer uma sequência de ações não-recíprocas para alcançar deslocamento não-nulo.

Os micronadadores artificiais mais simples consistem em esferas ligadas unidimensionalmente por meio de hastes rígidas, como o nadador de Najafi-Golestanian, representado nas figuras 1a e 1b, que também definem um estado e uma ação. As forças e velocidades em cada esfera do micronadador resultam das equações de interação corpo-fluido e das relativas à dinâmica do nadador. O estudo consiste em desenvolver modelos mecânicos de micronadadores utilizando elementos finitos e aproximações analíticas, e acoplá-los com algoritmos de *Reinforcement Learning* que treinam o agente por meio de redes neurais.

O estudo de micronadadores pode se mostrar relevante em medicina para, por exemplo, aplicação localizada de medicamentos, e em micro-robótica, já que os agentes reais exigirão um treinamento virtual prévio à transferência para os agentes físicos. Por fim, foram também analisados agentes nadadores bidimensionais, que têm a mesma estrutura dos unidimensionais de 3 esferas com a exceção de duas conexões na esfera central.



(a) O estado de um agente foi representado pela configuração de suas hastes. Nesse caso, como a 1ª haste está recolhida e a 2ª estendida, $s=[0 \ 1]$.



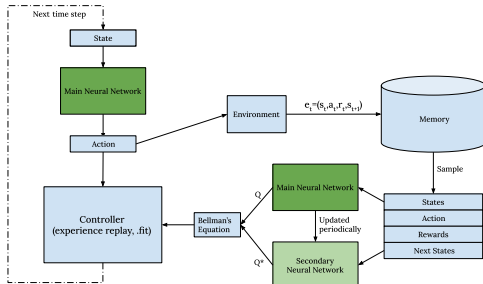
(b) A ação tomada por um agente foi representada pela alteração da configuração das hastes. Nesse caso, a ação $a=[1 \ 0 \ 0]$ significa alterar o comprimento da primeira haste.

Figura 1: Definição de estados e ações, necessárias para o funcionamento do algoritmo de aprendizado por reforço.

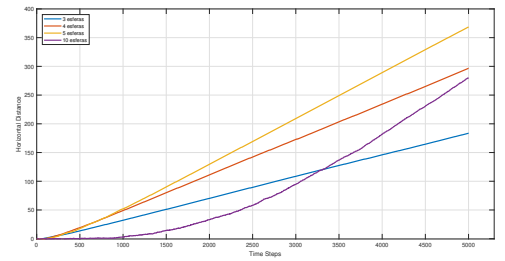
Métodos e Procedimentos

Para o caso unidimensional, a abordagem inicial, isto é, com o uso de *Q-tables* (aprendizado de reforço sem uso de redes neurais), se mostrou eficaz ao aproximar agentes simples, de 3 ou 4 esferas. No entanto, o aumento do número de esferas faz a *Q-table* crescer exponencialmente, o que reduziu a performance do agente. Por outro lado, a aplicação do *Deep Q-Learning* se mostrou eficaz tanto para agentes simples quanto para agentes complexos, de 3, 4, 5 e até 10 esferas.

Portanto, a implementação ideal é utilizando redes neurais. No entanto, isso exigiu algumas técnicas externas ao *Q-Learning*, como o *Experience Replay*, que previne *overfitting* no processo de treinamento da rede neural principal; e o uso de duas redes neurais para aumentar a estabilidade do aprendizado. O esquema geral do *Deep Q-Learning* está representado na figura 2a.



(a) Esquema geral do *Deep Q-Learning*.



(b) Comparação entre as fases de treinamento dos agentes de 3, 4, 5 e 10 esferas.

Figura 2: Esquema geral e resultados obtidos

Resultados

Verificou-se que todos os 4 casos estudados convergiram em um mesmo modo de locomoção (*gait*): o movimento começa com uma ação realizada na haste mais à esquerda, e depois na 2ª haste mais à esquerda, e assim por diante até chegar na haste mais à direita. Esse ciclo de ações gerou uma sequência de estados característica. De fato, esse ciclo de ações é similar ao encontrado em animais como lagartas e centopeias - as patas posteriores se locomovem, e essa locomoção é propagada até as patas anteriores, gerando um movimento.

Além disso, como foi evidenciado na figura 2b, quanto maior o número de esferas, maior a velocidade de locomoção do agente. No entanto, na fase de treinamento, o agente de 5 esferas obteve maior deslocamento porque é mais simples, e, portanto, aprende mais rápido (cerca de 200 passos de tempo). Enquanto isso, o agente de 10 esferas é mais complexo, e só aprende após 1000 passos de tempo.

Já o caso bidimensional usou as mesmas técnicas empregadas na modelagem do caso unidimensional, como *experience replay*, duas redes neurais para obter Q e Q^* , e o uso de funções de ativação adequadas. Além disso, foram construídas diversas redes neurais no formato *Multi-Layer Perceptron* para melhor entender o efeito da organização da rede neural no aprendizado. No entanto, não foi possível obter um agente satisfatório (consistente e reprodutível) que tenha aprendido a se locomover na direção de maior recompensa, seja seguindo uma trajetória complexa, seja simplesmente se locomovendo em direção ao quadrante que possuía maior recompensa.

Os dados de locomoção (velocidade e posição) de cada uma das esferas foram gerados de duas maneiras: utilizando um código de CFD em FEniCS, e utilizando a aproximação analítica de Oseen, que calcula o tensor de Oseen H , da equação $V_i = \sum_{n=1}^N H_{ij} F_j$. Essa equação, juntamente com a cinemática das hastes e equilíbrio de forças e torque, compuseram as equações do sistema. Por fim, foi imposto que as hastes se movem em uma senoide. Os resultados foram similares aos obtidos usando a abordagem de CFD, com a exceção de que o agente usando a aproximação de Oseen se mostrou mais sensível às condições iniciais (configuração inicial das hastes), e aprendeu mais lentamente.