



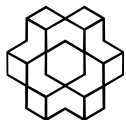
UFRJ
UNIVERSIDADE FEDERAL
DO RIO DE JANEIRO



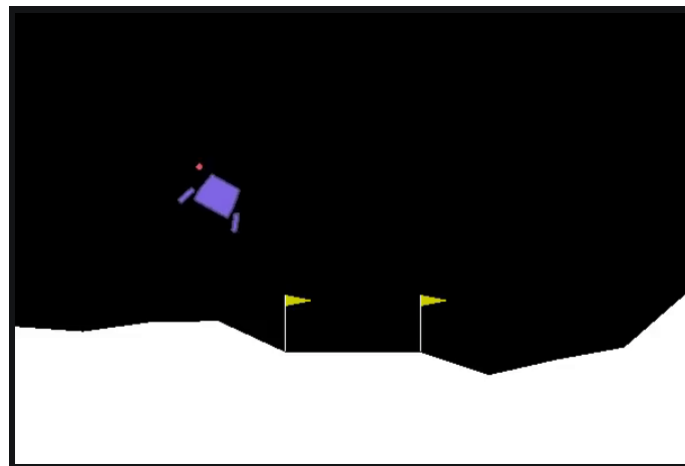
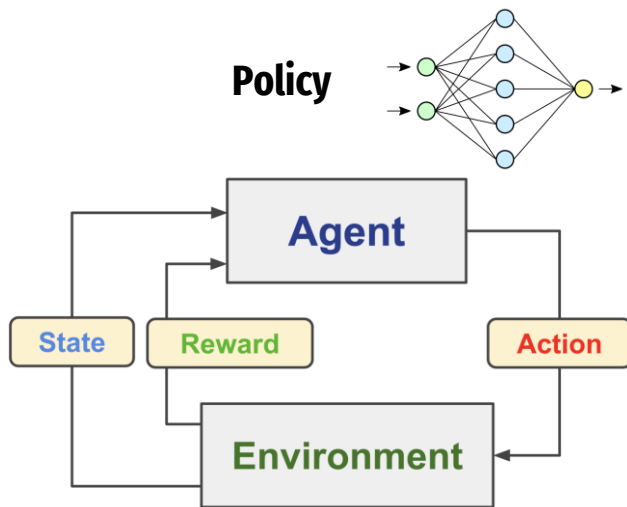
Aprendizado por reforço utilizando Algoritmos Genéticos

CPE 723 – Otimização Natural
Antonioni Barros Campos

Problema



Gym: Interface para aprendizado por reforço e uma **coleção diversa de ambientes de referência.**



Lunar Lander

Problema

| | |
|-------------------|----------------------------------------|
| Action Space | Discrete(4) |
| Observation Shape | (8,) |
| Observation High | [1.5 1.5 5. 5. 3.14 5. 1. 1.] |
| Observation Low | [-1.5 -1.5 -5. -5. -3.14 -5. -0. -0.] |

Estado

- Posição em x, y
- Velocidades em x, y
- Ângulo e velocidade angular
- Pernas esquerda e direita em contato com o solo

Recompensa (Aptidão)

A recompensa por se mover do topo da tela para a plataforma de pouso e parar é de cerca de 100 a 140 pontos.

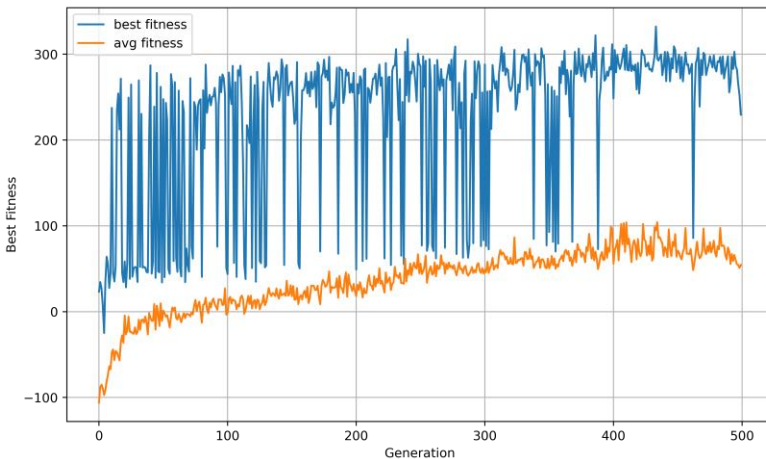
Estado Inicial

O módulo de pouso começa no centro superior da janela de visualização com uma força inicial aleatória aplicada ao seu centro de massa.

Análise

Estratégia de Evolução (ES)

- Representação Contínua
- Seleção dos pais – aleatória e uniforme
- Recombinação – Discreta e Intermediária
- Mutação – Não correlacionada com n tamanhos de passos
- Seleção dos sobreviventes – Determinística a partir dos filhos



Programação Evolucionária (EP)

Comentários

- Objetivo é utilizar o MBF e AES para comparar o ES com o EP.
- Rede Neural utilizada com uma camada escondida de 16 unidades. Nas próximas execuções, serão utilizadas 64 ou 128 unidades.
- Utilizar mais de um episódio para o cálculo da aptidão é muito custoso computacionalmente. Então, está sendo utilizado um único episódio.
- Dúvida entre realizar o processo utilizando um único estado inicial para os episódios ou gerar aleatoriamente.