6º Exercício

Aprendizagem por Reforço

A aprendizagem por reforço (ou *Reinforcement Learning – RL,* em inglês) é considerada uma forma mais genérica de aprendizado que os tradicionais aprendizado supervisionado ou não supervisionado. No aprendizado supervisionado têm-se a resposta desejada e busca-se capacidade de generalização. No aprendizado não supervisionado procura-se agrupar os dados em classes, refletindo a distribuição estatística dos dados. Já no aprendizado por reforço, um "agente" interage com o ambiente, recebendo reforço positivo caso a ação o aproxime do objetivo ou um reforço negativo caso se afaste deste. Uma sequência de Estado, Ação, Recompensa, Próximo Estado, ... é guiada pela recompensa imediata e pela expectativa da recompensa total final (Q-Learning). A programação dinâmica proposta por R. Bellman teve papel central na redução do esforço computacional em processos decisórios discretos. Problemas "pequenos" podem ser resolvidos com o uso de uma tabela para a matriz Q. Problemas "maiores" ou contínuos demandam o uso de modelos para aproximar Q.

1 – Aprendizagem por Reforço

As principais inspirações do Aprendizado por Reforço vêm da Biologia e da Psicologia, em que a exploração do ambiente por um agente rende a este recompensas e penalizações. O estado do agente é continuamente avaliado e as ações deste modificam o estado do agente (e.g., a posição x,y) no ambiente, Figura 1. Atualmente RL é uma abordagem interdisciplinar com ramificações em diversas áreas do conhecimento, Figura 2.

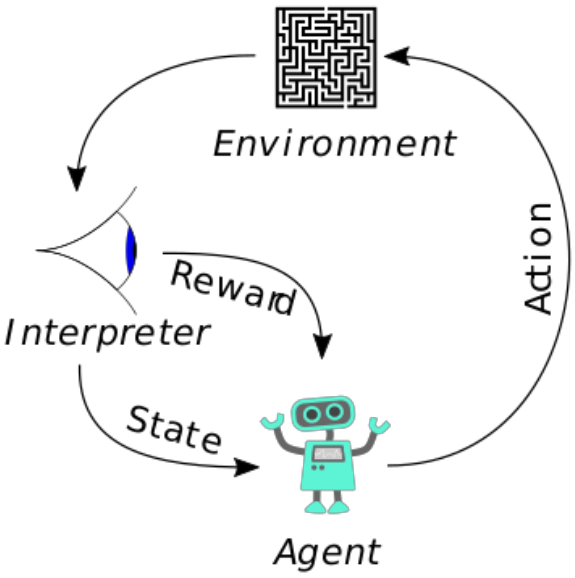

Figura 1 – Ilustração do contexto típico de RL.
(https://en.wikipedia.org/wiki/Reinforcement_learning)

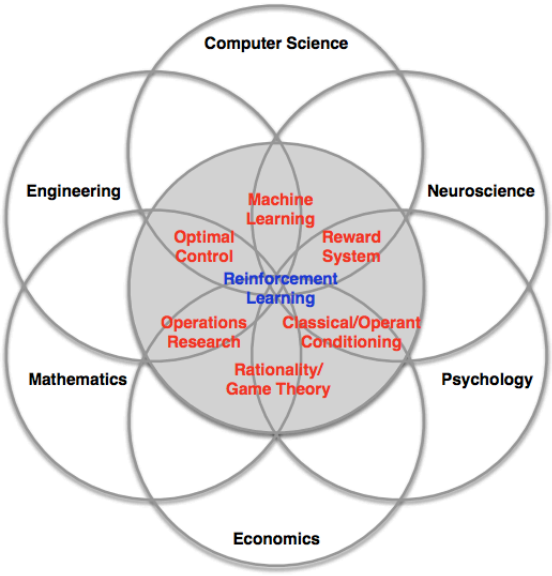

Figura 2 – Principais áreas de pesquisa em RL.
https://icml.cc/2016/tutorials/deep_rl_tutorial.pdf

2 – Problema do Labirinto

A busca da rota mais curta dentro de um labirinto entre ponto inicial e ponto objetivo (ou problema do "rato e do queijo", "pirata e tesouro" etc.) sem que haja, inicialmente, por parte do agente, nenhum conhecimento, seja do labirinto, seja da posição do ponto objetivo. Este é um problema típico que pode ser resolvido pelo RL.

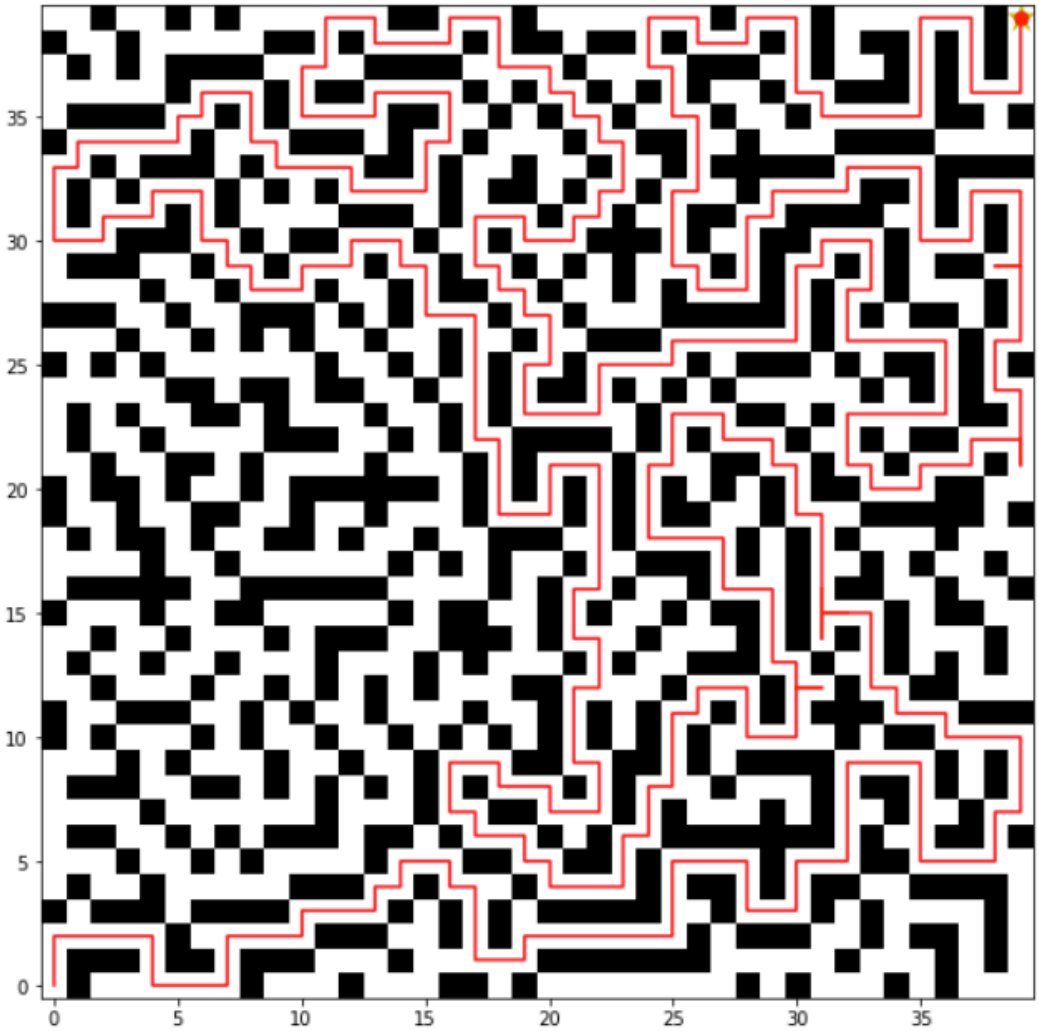

Figura 2 – Labirinto típico 40x40 com rota obtida.

3 – Repositório: jrieke/reinforcement-maze

Para a execução deste exercício sugere-se iniciar com os códigos disponibilizados por J. Rieke no GitHub (https://github.com/jrieke/reinforcement-maze.). O código Q-table.ipynb deverá ser adequado para verificar, na prática, características do algoritmo Q-Learning. Uma das características mais interessantes deste código é a geração automática de labirintos "interessantes". (Muitos programas de demonstração relegam ao usuário a criação manual dos labirintos).

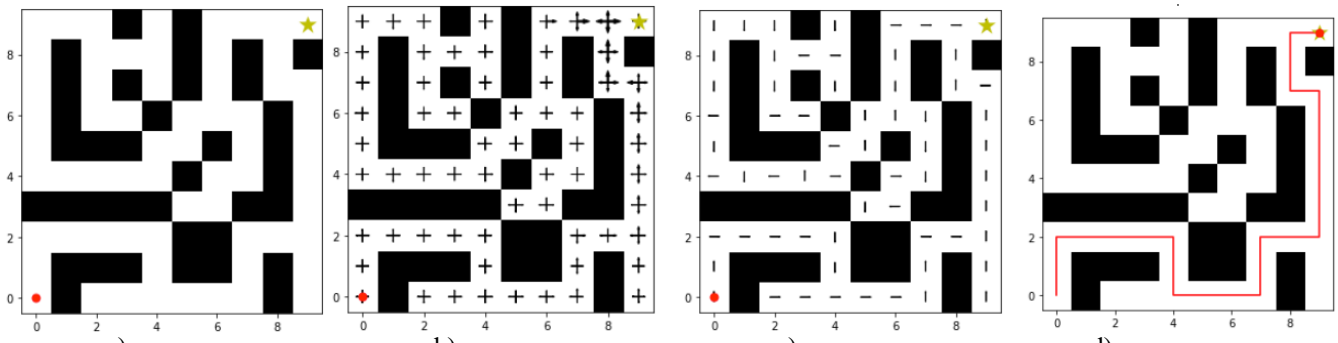

Figura 4 – Q-table.ipynb: Além da visualização em cores da Matriz Q, estas telas são úteis ao exercício.
a) Labirinto 10x10 com início e objetivo; b) Largura das setas correspondem ao valor predito em cada direção;
c) Apenas setas com o maior valor em cada estado. d) Rota entre início e objetivo.

4 – Procedimento: Experimentos com o Labirinto RL

Considerações Inicíciais:

Os melhores resultados são função dos parâmetros do RL. Podem acontecer as seguintes situações:

a) Não se encontrar nenhuma rota entre início o ponto objetivo.
b) Se encontrar *uma* rota válida entre início e objetivo.
c) Encontrar várias rotas válidas entre início e objetivo.

Uma vez que o Q-table esteja treinado é possível obter rotas de pontos iniciais distintos.

Experimentos:

Obter rotas válidas ligando início ao objetivo para:

- Tamanho do Labirinto: 20x10, 40x40
- Ponto Objetivo: aleatório x (mx-1, my-1)
- épsilon: 0 x 0.1 x 0.2 (probabilidade de ação aleatória em vez da melhor ação)
- gamma: 0.9 x 0.95 (desconto na recompensa)

Obs: Alterações necessárias no Código.
- epoch in range(200): Treinamento RL – 200 épocas não são suficientes para labirintos grandes.
- while not done: na construção de rota final. Inserir contador. Se não achou rota, laço infinito.
- O tamanho das setas deve ser ajustado de acordo com o tamanho do labirinto (Figs. 4.b) e 4.c))

5 – Relatório

- Apresente os parâmetros utilizados para cada labirinto testado
- Apresente rotas do ponto de início ao ponto objetivo.
- Apresente rotas alternativas caso encontradas.
- Apresente para cada RL, pelo menos uma rota iniciando de um ponto distinto do treinado.
- Comente o que acontece se épsilon = 0.
- Comente sobre os melhores valores de gamma para os labirintos 20x10 e 40x40.

_____
_____

BOM TRABALHO!!