

ML Under Modern Optimization Lens - Reinforcement Learning -

Exercícios

Giovanni Amorim

Junho, 2023

A implementação do problema seguiu a lógica mostrada em sala, com os seguintes parâmetros:

- Espaço de estados: ["*high*", "*low*"]
- Espaço de ações: ["*search*", "*wait*", "*recharge*"]
- $\alpha = 0.8$
- $\beta = 0.7$
- $r_{search} = 3$
- $r_{wait} = 1$

Após a implementação, podemos avaliar as ações selecionadas para cada estado:

- "*high*" => "*search*"
- "*low*" => "*recharge*"

As ações selecionadas pela política fazem sentido uma vez que há uma penalidade alta em caso de falta de bateria pelo robô.

Realizando 1000 simulações do processo, podemos avaliar a distribuição das recompensas totais experimentadas pela política, como mostrado a seguir:

