

Algoritmo SARSA

Fabício Barth

Inspér Instituto de Ensino e Pesquisa

Fevereiro de 2025

A regra para update da **Q-table** no algoritmo **Q-Learning** é:

$$Q(s, a) \leftarrow Q(s, a) + \alpha[r + \gamma \max_{A'} Q(s', A') - Q(s, a)] \quad (1)$$

A diferença entre o novo valor e a estimativa antiga é utilizada para atualizar a estimativa antiga. O algoritmo **Q-Learning** considera como o novo valor o valor máximo das possibilidades no novo estado s' :

$$\max_{A'} Q(s', A') \quad (2)$$

SARSA: on-policy

- No entanto, a ação realmente executada pelo agente pode não ser a ação que tem o valor máximo em s' devido a função de escolha da ação ser baseada no valor de ϵ e ter características aleatórias.
- Por isso que o algoritmo **Q-Learning** é chamado de **off-policy**.

SARSA: on-policy

O algoritmo SARSA é chamado de **on-policy** porque ele atualiza a **Q-table** da seguinte forma:

$$Q(s, a) \leftarrow Q(s, a) + \alpha[r + \gamma Q(s', a') - Q(s, a)] \quad (3)$$

o algoritmo SARSA atualiza $Q(s, a)$ considerando a real ação a' executada pelo agente em s' .

SARSA: on-policy

function *Sarsa*(*env*, α , γ , ϵ , ϵ_{min} , ϵ_{dec} , *episódios*)

inicializar os valores de $Q(s, a)$ arbitrariamente

for todos os episódios **do**

inicializar s a partir de *env*

$a \leftarrow \text{escolha}(s, \epsilon)$

repeat

$s', r \leftarrow$ executar a ação a no *env*

$a' \leftarrow \text{escolha}(s', \epsilon)$

$Q(s, a) \leftarrow Q(s, a) + \alpha[r + \gamma Q(s', a') - Q(s, a)]$

$s \leftarrow s'$

$a \leftarrow a'$

until s ser um estado final

if $\epsilon > \epsilon_{min}$ **then** $\epsilon \leftarrow \epsilon \times \epsilon_{dec}$

end for

return Q

Atividade de implementação

Comparando Q-LEARNING e SARSA

O objetivo desta atividade é implementar o algoritmo SARSA e compará-lo com o Q-LEARNING.

Atividades

Siga o roteiro descrito em https://insper.github.io/rl/classes/05_x_sarsa/

► Link

- Richard S. Sutton and Andrew G. Barto. 2018. Reinforcement Learning: An Introduction. A Bradford Book, Cambridge, MA, USA.

Capítulo 6.4

- Rummery, G. A. & Niranjan, M. (1994), On-Line Q-Learning Using Connectionist Systems (TR 166), Technical report, Cambridge University Engineering Department, Cambridge, England.
- Aurélien Géron. Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow, 2nd Edition, 2019.