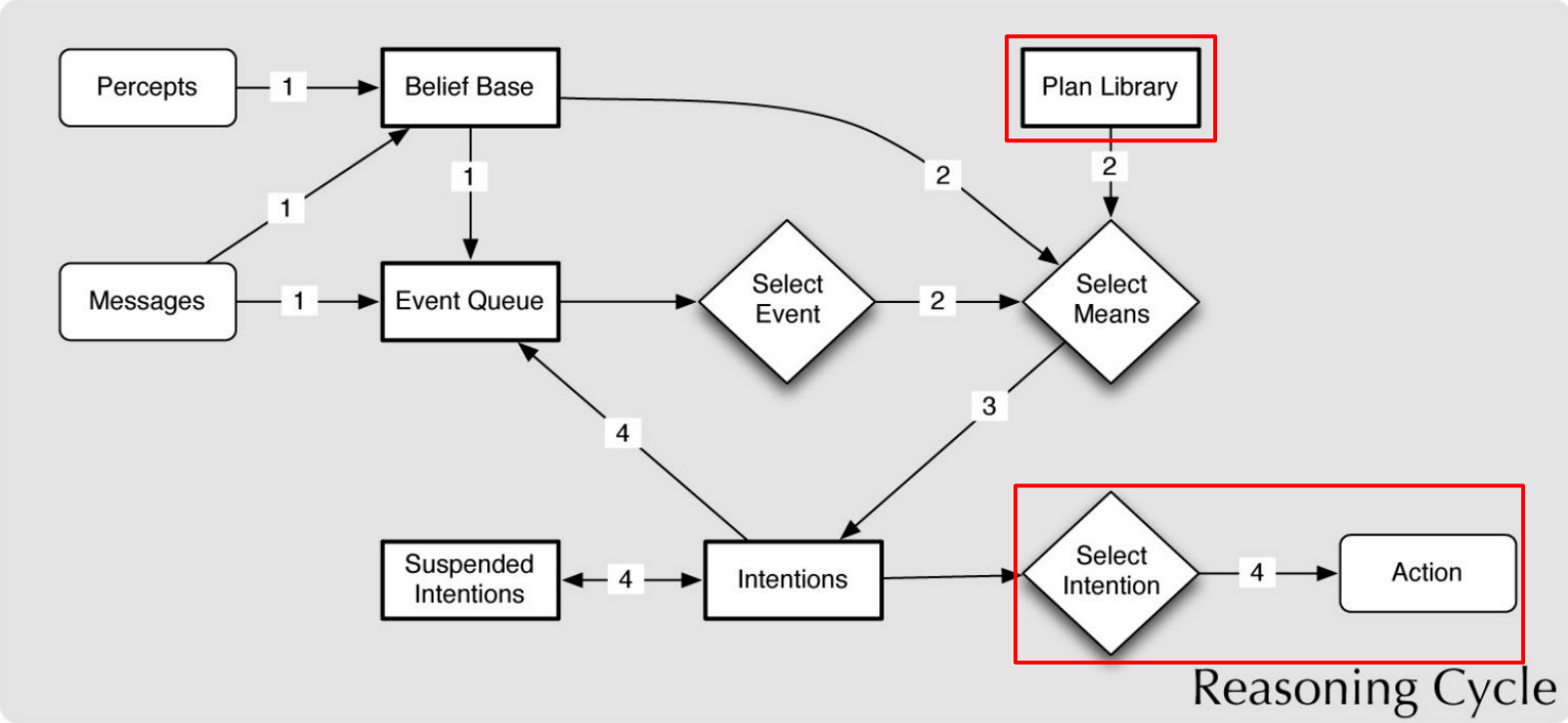
Aprendizagem em Sistemas Multiagentes (MAL)

Fundamentos e desafios

Motivação


Percepts
1
Belief Base
Plan Library
2
1
1
2
Messages
1
Event Queue
Select Event
2
Select Means
3
4
Suspended Intentions
4
Intentions
Select Intention
4
Action
Reasoning Cycle

Aviso

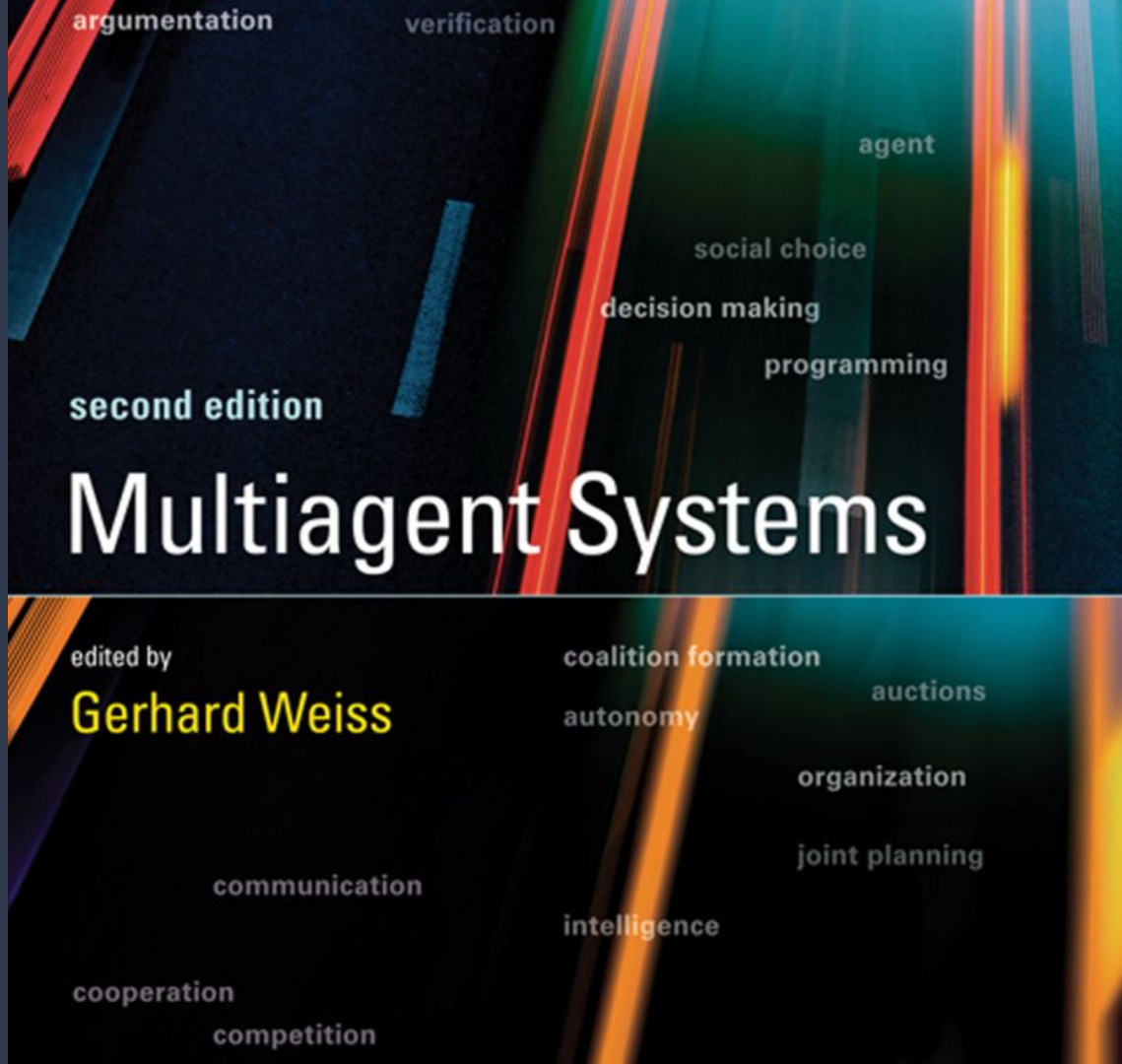
Aprendizagem por reforço (RL) não é a única abordagem para sistemas multiagentes.

Mas esta apresentação terá a “ótica” de RL.

Aviso

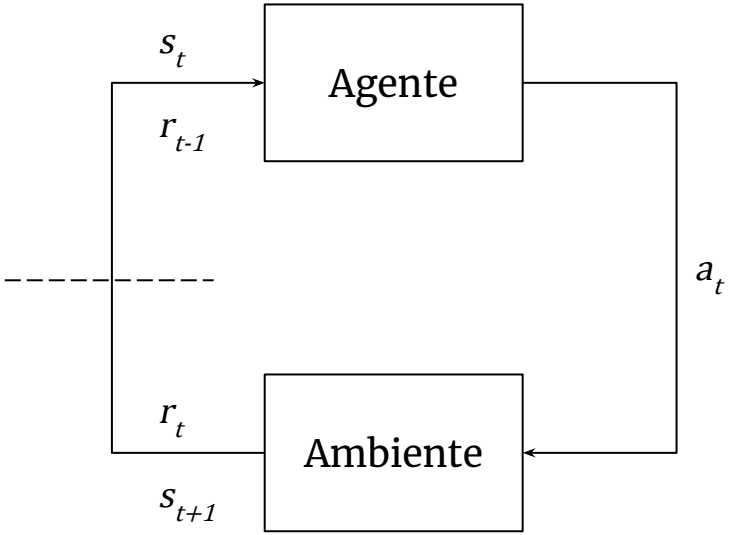
Esta apresentação tem como principal referência o livro “Multiagent Systems”, de Gerhard Weiss.

As demais referências serão mencionadas.



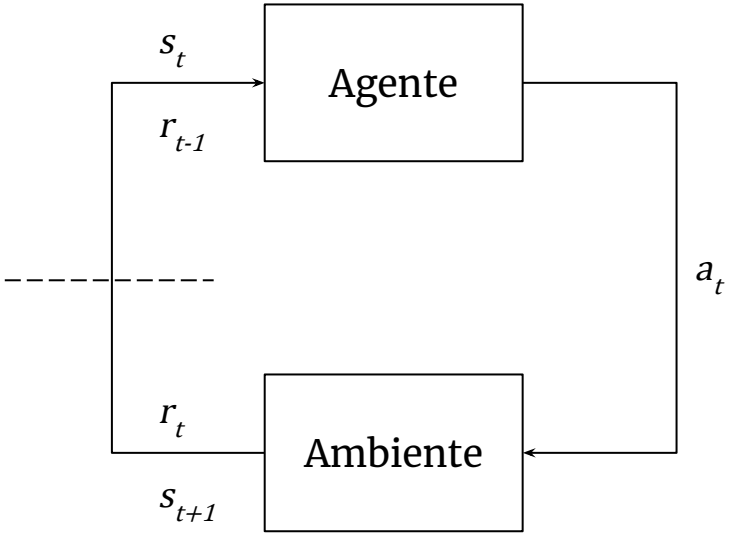
Introdução

RL Tradicional
ou
MAL com 1 Agente



*"The objective of a reinforcement learner is to discover a policy, i.e., **a mapping from states to actions**, so as to maximize the reinforcement signal it receives"*

RL Tradicional
ou
MAL com 1 Agente



Formulação como um processo de decisão de Markov (MDP):

$$\langle S, A, T, R \rangle,$$
$$T : S \times A \times S \to [0, 1],$$
$$\mathbb{P}[s_{t+1} \mid s_t, a_t] = T(s_t, a_t, s_{t+1}),$$
$$r_t = R(s_{t+1})$$

Objetivo:

$$\pi : S \to A$$

Aprendizagem Multiagente

*“The subfield of multiagent learning studies agent definitions, algorithms, interactions, and reward structures to create **adaptive agents** that can function in environments where their **actions shape and are shaped by the actions of other agents**.”*

adaptive agents ⇒ agentes que aprendem a lidar com as mudanças no ambiente e nas estratégias do outros agentes

actions shape and are shaped by… ⇒ interação bi-direcional com os outros agentes

Exemplo: Multinight Bar Problem

Toda semana, o agente precisa escolher uma entre n noites para ir a um bar, que tem capacidade c.

Toda ocupação r_i do bar na noite i vem de uma distribuição fixa, mas desconhecida.

A recompensa é máxima quando o bar está com ocupação intermediária (metade da capacidade).

⇒ O agente consegue estimar a expectativa de ocupação do bar pela média das tentativas anteriores, convergindo para uma política ótima.

Exemplo: Multinight Bar Problem

Agora, **Múltiplos** agentes precisam escolher uma entre n noites para ir a um bar, que tem capacidade c.

Dessa forma, r_i depende também da quantidade de agentes que foram ao bar na noite i.

⇒ O processo é não-estacionário e não existe garantia teórica de convergência.

Desafios

Desafios

A mudança de paradigma de MAL implica em 3 grandes desafios para algoritmos de aprendizagem:

- Não-estacionariedade
- Maldição da dimensionalidade
- Creditação das ações

Não-Estacionariedade

RL geralmente implica em trabalhar com um ambiente (T) (quase-)estacionário.

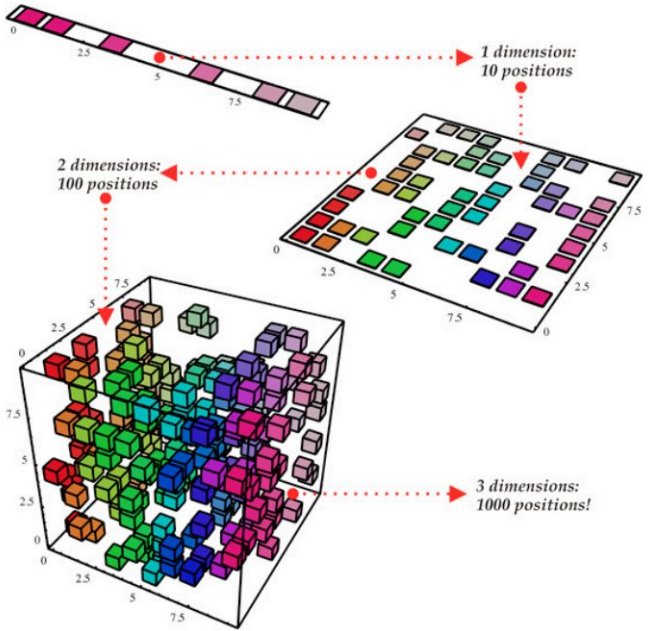
Em um MAS, é impossível para um agente estimar o estado seguinte sem total conhecimento do processo de decisão dos demais agentes.

Maldição da dimensionalidade

Encontrar uma política π é um problema de busca no espaço definido por S, A e R.

- $|S|$ aumenta com os estados dos agentes
- $|A|$ aumenta exponencialmente uma vez que $A = \Pi\ A^i$, onde A^i são as ações do i-ésimo agente

Dessa forma, a dimensionalidade do problema aumenta drasticamente em MAL.



Creditação das Ações

Em RL, creditação das ações já é um problema clássico uma vez que recompensas em problemas reais decorrem de atrasos e de sequências de ações.

Para MAL, uma nova camada surge que é a multiplicidade de ações em um mesmo instante.

Creditação das Ações

Projeto de Recompensas

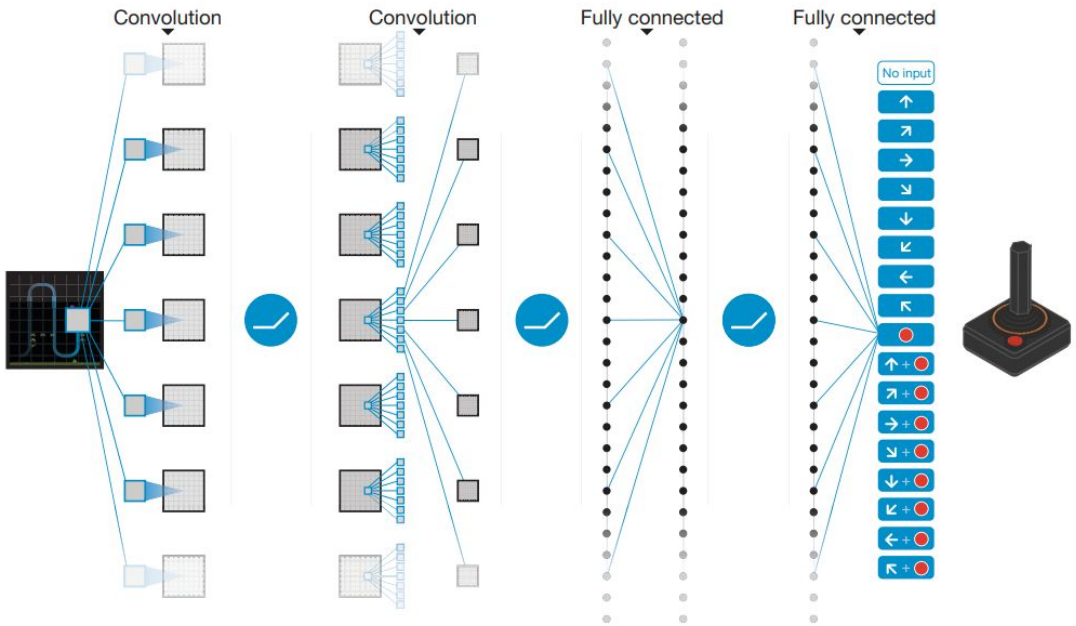
Recompensas afetam largamente as interações, os pontos de equilíbrio e a convergência das políticas.

Estado-da-Arte

Deep Reinforcement Learning, 2015

Até então, MAL+RL se limitavam a problemas simples, muitas vezes a jogos estáticos. [Tampuu et al., 2017]

Usando DNNs para resolver a maldição da dimensionalidade, esse trabalho “viabilizou” MAL+RL para problemas realistas.



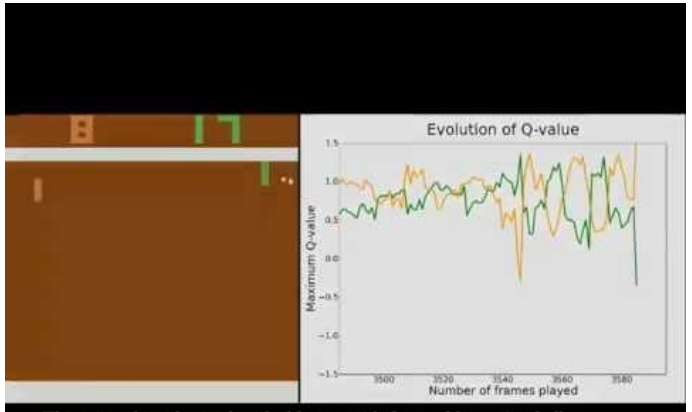
Mnih, V., Kavukcuoglu, K., Silver, D. et al. Human-level control through deep reinforcement learning. Nature 518, 529–533 (2015). https://doi.org/10.1038/nature14236

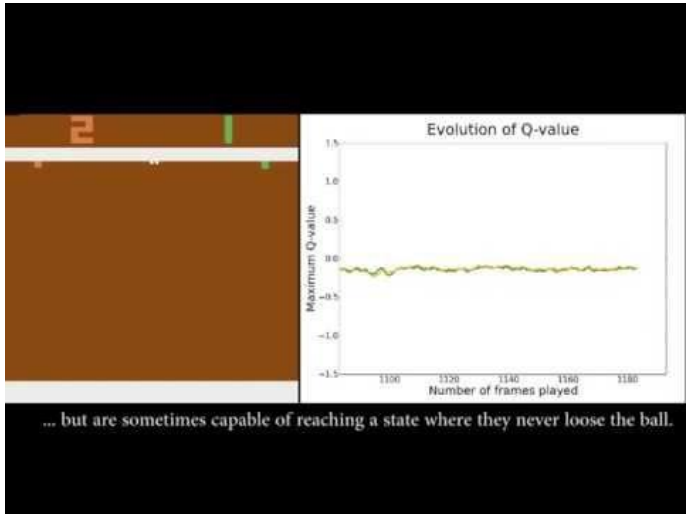
Cooperação e Competição, 2017

Através da recompensa, foram estimulados comportamentos de cooperação e competição no jogo *Pong*.

Além disso, em comparação com o cenário *single-player*, obteve-se um erro de generalização muito menor.

Tampuu A, Matiisen T, Kodelja D, Kuzovkin I, Korjus K, et al. (2017) Multiagent cooperation and competition with deep reinforcement learning. PLOS ONE 12(4): e0172395. https://doi.org/10.1371/journal.pone.0172395





Leniência, 2018

DRL utiliza as experiências passadas para tornar sua aprendizagem mais eficiente, mas a não-estacionariedade de MAL torna isso problemático.

Leniência é uma técnica para ignorar situações pouco frequentes que geraram recompensas pequenas.

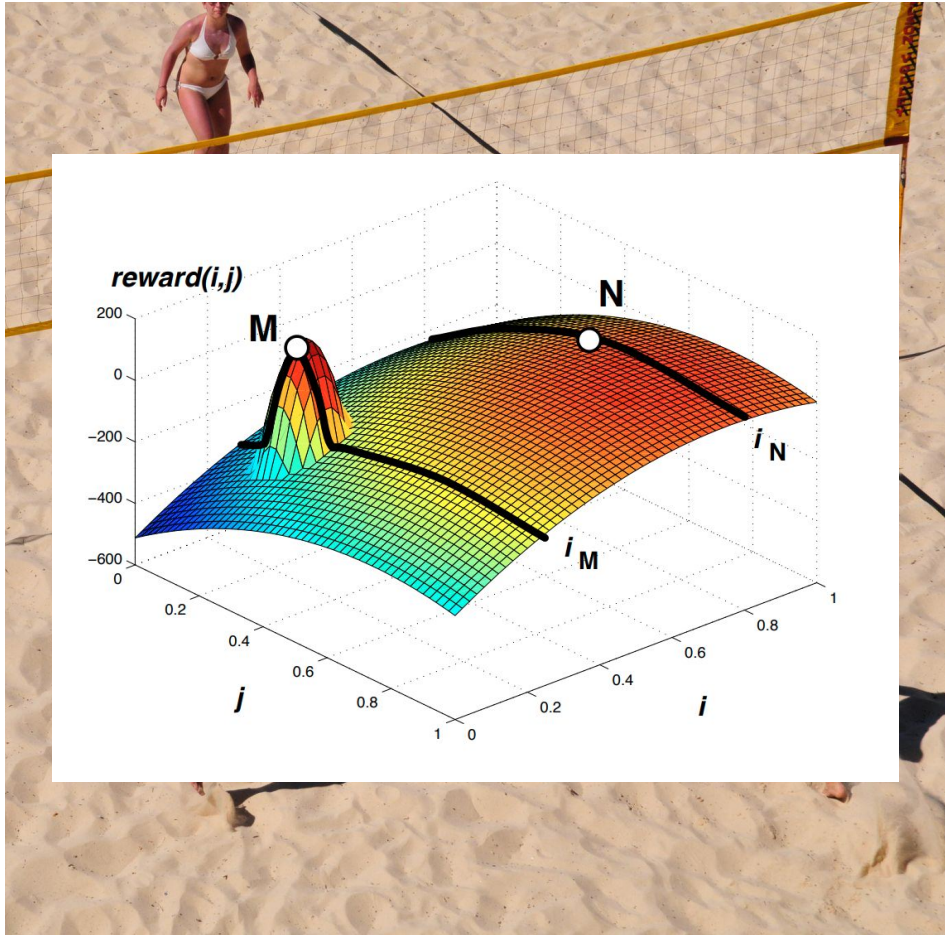
Gregory Palmer, Karl Tuyls, Daan Bloembergen, and Rahul Savani. 2018. Lenient Multi-Agent Deep Reinforcement Learning. In Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems (AAMAS '18). International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC, 443–451.

Leniência, 2018

DRL utiliza as experiências passadas para tornar sua aprendizagem mais eficiente, mas a não-estacionariedade de MAL torna isso problemático.

Leniência é uma técnica para ignorar situações pouco frequentes que geraram recompensas pequenas.

Gregory Palmer, Karl Tuyls, Daan Bloembergen, and Rahul Savani. 2018. Lenient Multi-Agent Deep Reinforcement Learning. In Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems (AAMAS '18). International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC, 443–451.



Conclusão

Conclusão

- MAL+RL é uma área “recente” com resultados promissores
- O sucesso depende de um trabalho grande no ajuste das recompensas do sistema, sendo quase sob-medida para a aplicação
- MAL+DRL é bastante custoso computacionalmente para problemas reais
- A comunicação entre agentes não parece ser muito explorada

Obrigado

Bruno M. Pacheco

mpacheco.bruno@gmail.com