



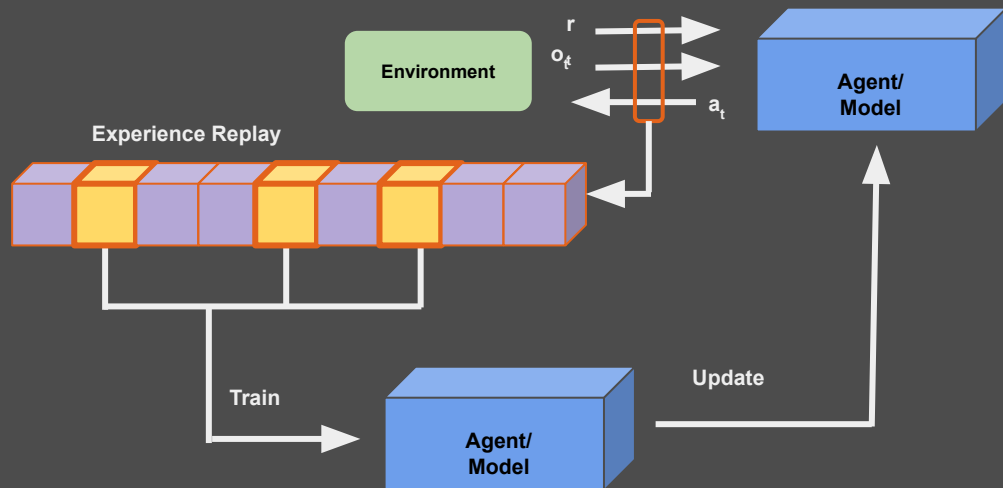
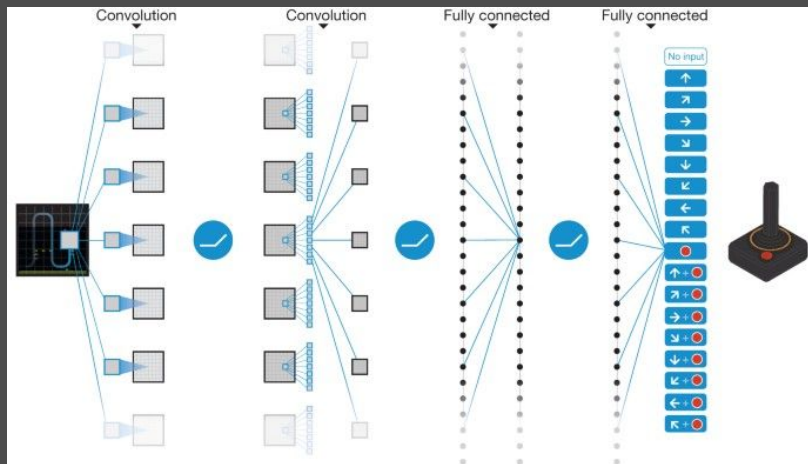
Aprendizado por Reforço

AULA - 7

Tópicos Avançados

De onde viemos...

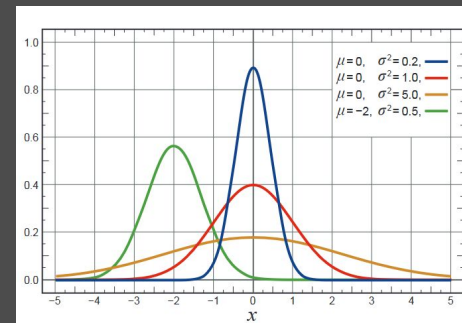
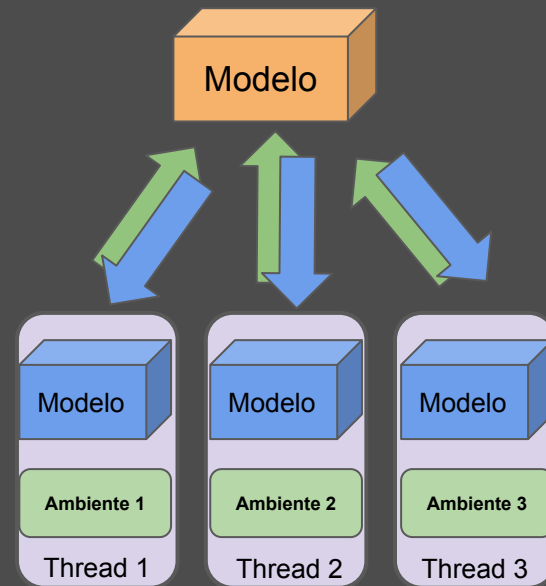
- Deep Q-Network



$$L(\theta) = \left(r + \gamma \max_{a'} Q(s', a'; \phi) - Q(s, a; \theta) \right)^2$$

De onde viemos...

- A3C + PPO**

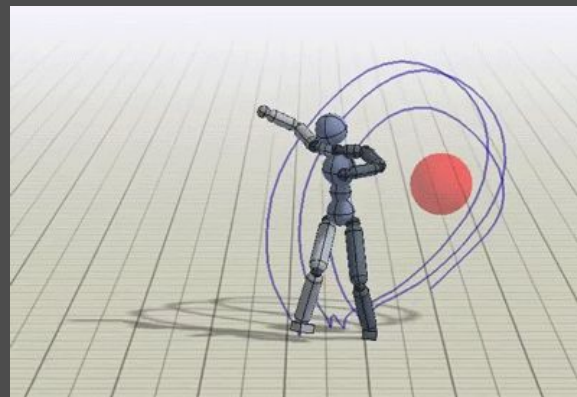
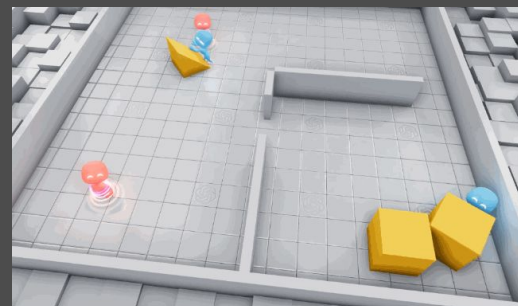
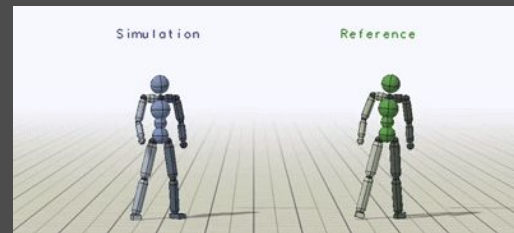


Method	Training Time	Mean	Median
DQN	8 days on GPU	121.9%	47.5%
Gorila	4 days, 100 machines	215.2%	71.3%
D-DQN	8 days on GPU	332.9%	110.9%
Dueling D-DQN	8 days on GPU	343.8%	117.1%
Prioritized DQN	8 days on GPU	463.6%	127.6%
A3C, FF	1 day on CPU	344.1%	68.2%
A3C, FF	4 days on CPU	496.8%	116.6%
A3C, LSTM	4 days on CPU	623.0%	112.6%

Table 1. Mean and median human-normalized scores on 57 Atari games using the human starts evaluation metric. Supplementary Table SS3 shows the raw scores for all games.

Hoje

- Como aprender comportamentos extremamente complexos?
- Posso aprender várias tarefas ao mesmo tempo?
- O que acontece se eu tiver vários agentes aprendendo?
- Posso imitar comportamento com generalização?
- E se eu quiser aplicar meu algoritmo no mundo real?





Inverse Reinforcement Learning

Como IRL funciona

- Aprender uma função de recompensa tendo uma política
- Funciona com trajetórias de experts

- Inicializar Política(θ) e estimativa de Recompensa(ϕ)
- Treinar Política(θ) usando Recompensas(ϕ)
- Calcular erro entre Política(θ) treinada e Expert
- Atualizar estimador de Recompensa(ϕ) com o erro

- Existem variações para o cálculo do erro



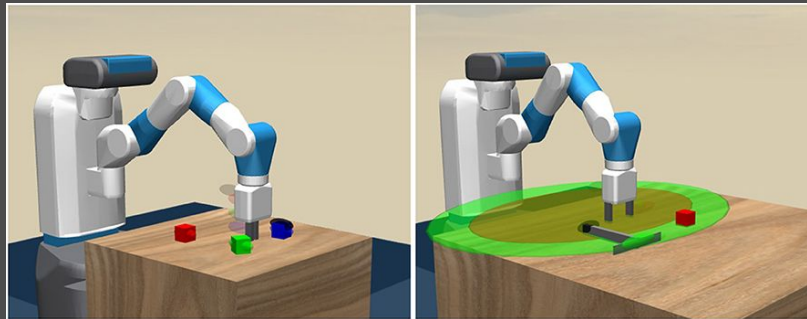
Meta-Learning




Transfer Learning

Transferência de Conhecimento

- Transferir conhecimento de uma tarefa para outra
- Tarefas similares = Soluções similares
- Pegando um modelo treinado em uma tarefa, ele pode ter facilidade de aprender tarefas similares
- Buscando generalização entre tarefas diferentes

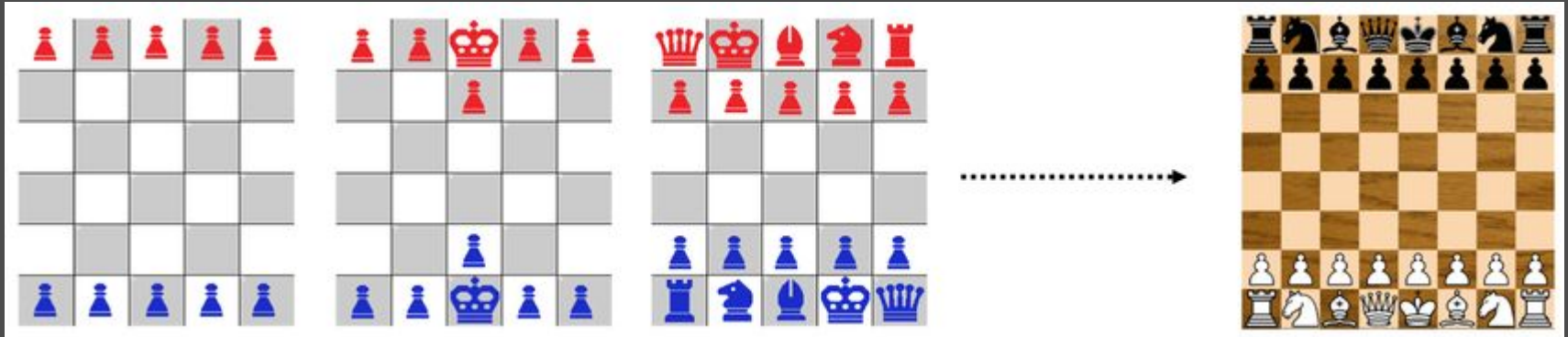




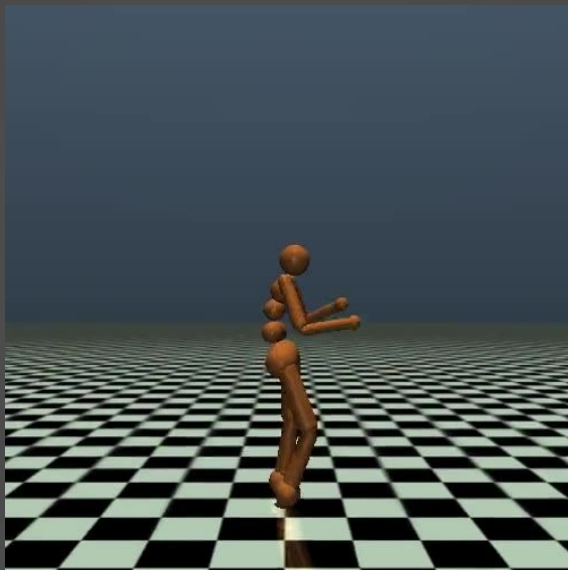
Curriculum Learning

Ideia Principal

- Quebrar tarefas complexas em subtarefas de dificuldade ascendente
- Transferência de conhecimento é comprovada







Ficar de Pé



Andar



Correr/Pular



Aumentar inimigos, adicionar unidades diferentes



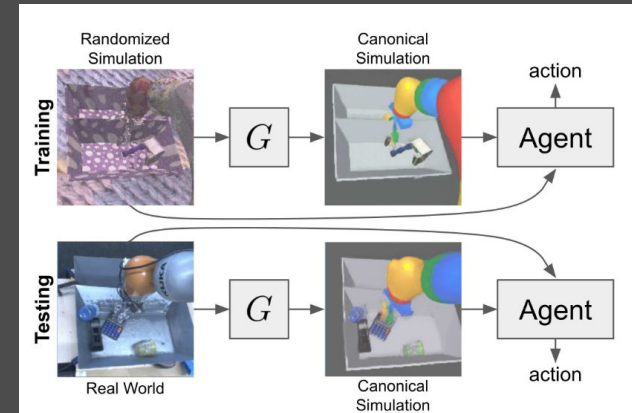
Sim-to-Real

Aplicando Soluções no Mundo Real

- Algoritmos treinados no mundo real são ok
 - Algoritmos treinados em simulação precisam de adaptação
 - Simulações não descrevem corretamente o mundo real
-
- Principal solução: **Generalização**
 - Presumindo: um algoritmo que consegue agir em ambientes diferentes, conseguirá agir no mundo real

Adicionando Ruído

- Domain Randomization (imagens)
- Modelo Generativo (imagem)
 - Faz imagens reais parecer vindas do simulador
- Ruído em medidas (dados do simulador)





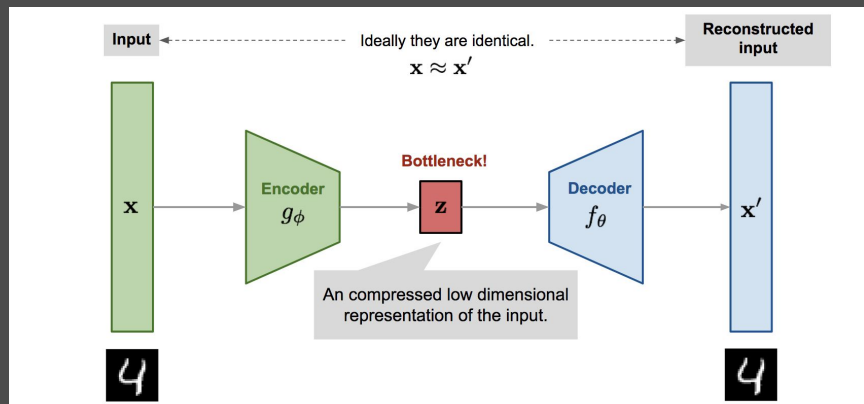
Model Based



Representation Learning

Aprendendo Representações

- Aprender uma representação diferente, simplificada, menor, ou comprimida, de algo
- Ex: Autoencoders



man	→	0.6	-0.2	0.8	0.9	-0.1	-0.9	-0.7
woman	→	0.7	0.3	0.9	-0.7	0.1	-0.5	-0.4
king	→	0.5	-0.4	0.7	0.8	0.9	-0.7	-0.6
queen	→	0.8	-0.1	0.8	-0.9	0.8	-0.5	-0.9

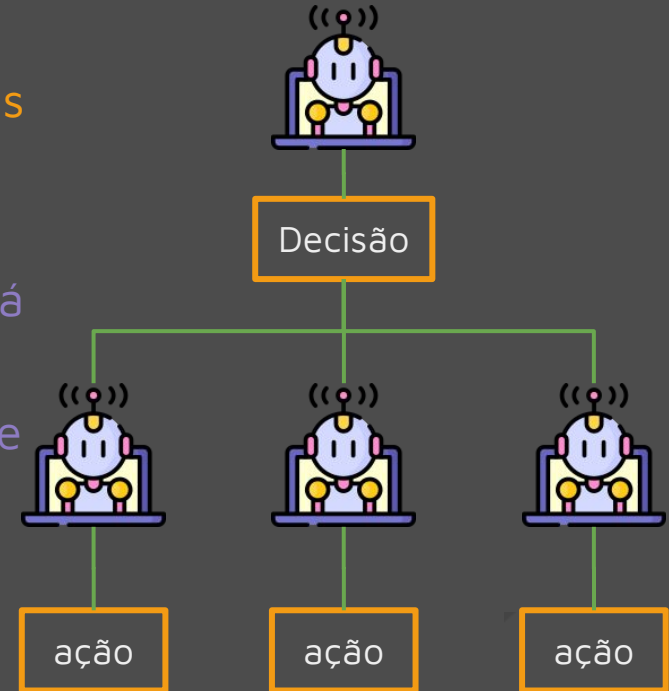
Word Word embedding D



Hierarchical Reinforcement Learning

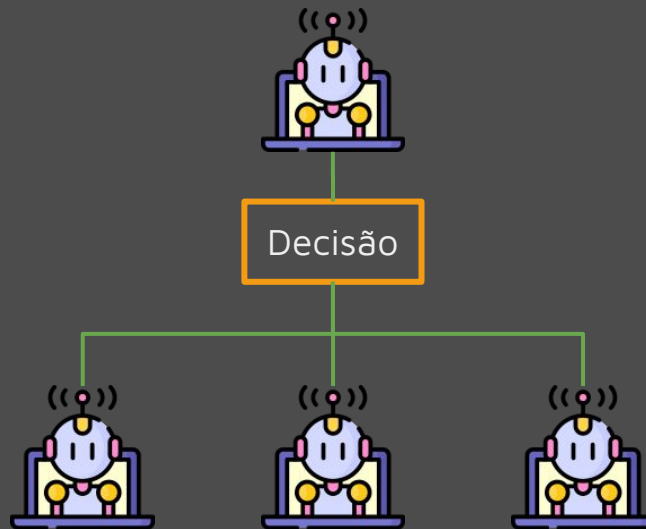
Hierárquico

- Decisões são feitas em vários níveis
- Níveis baixos dependem de decisões em níveis mais altos
- Um agente que escolhe qual (outro) agente irá agir dependendo da situação
- Um agente que observa a situação completa e envia estratégias para outros agentes que observam de forma parcial



Treinando de forma Hierárquica

- É difícil propagar o gradiente através de muitos níveis
- A recompensa geralmente depende das ações dos níveis mais baixos

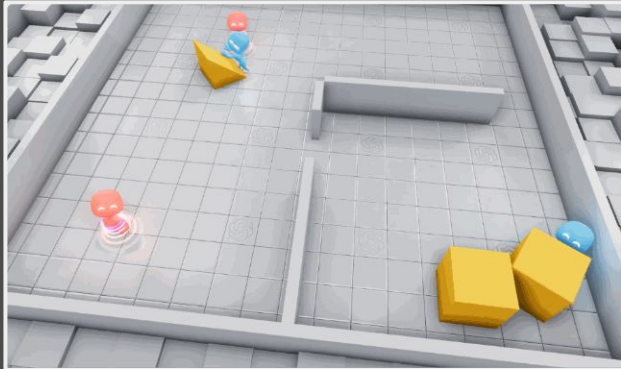




Multi-Agent

E se eu tiver vários agentes?

- Agentes interpretam uns aos outros como parte do ambiente
- Há vários tipos de ambientes multiagente
- Tarefas cooperativas ou competitivas
- Agentes simétricos ou assimétricos
- Atingem comportamentos coordenados





Self-Play

Jogando contra si mesmo

- Um modelo pode jogar contra si mesmo em ambientes com certas características
 - Ambientes Simétricos (capacidades iguais* dos dois lados)
 - Ambientes Competitivos
- **Por Que?**
- O ambiente não possui um adversário padrão
 - Ex: Xadrez





AlphaStar



OpenAI Five



Sua vez . . .