BandTec DIGITAL SCHOOL

Uma abordagem de algoritmos de reinforcement learning para atingir performance próxima do nível humano em jogos de atari



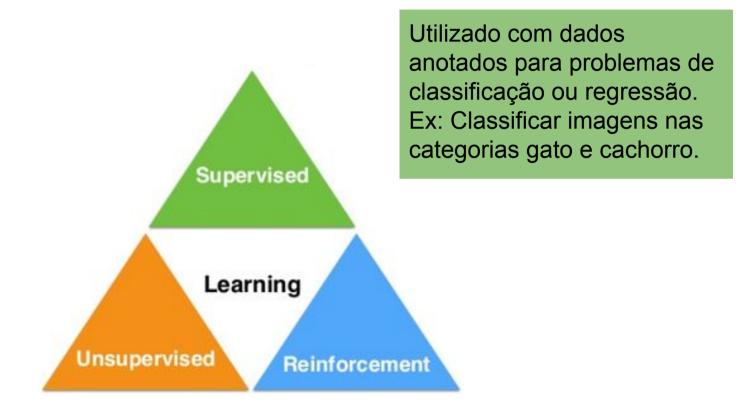
Agenda

- Introdução ao tema
- Objetivo
- Metodologia
- Apresentação dos resultados
- Conclusão
- Bibliografia





Introdução ao tema -Algoritmos de Machine Learning

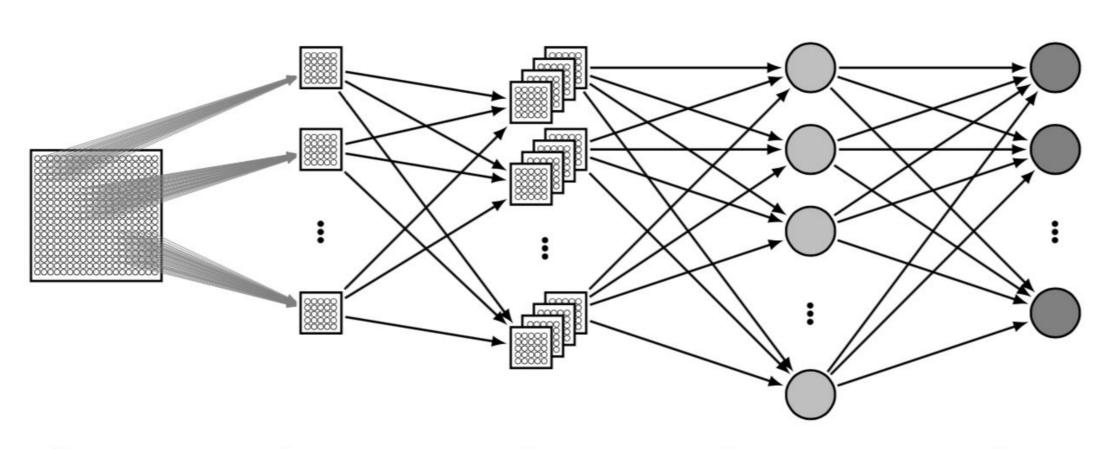


Encontrar padrões em uma estrutura oculta sem dados anotados e sem feedback. Ex: Encontrar segmentos de potenciais clientes

Utilizado sem dados anotados e aprende através de recompensas positivas ou negativas. Ex: algoritmo de trading de bolsa de valores.



Introdução ao tema -Redes Neurais Convolucionais

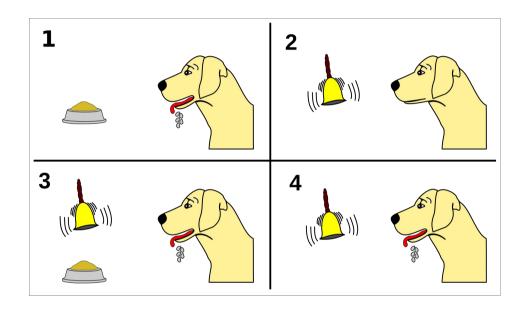


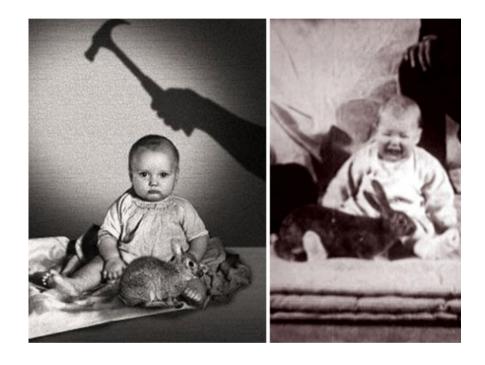
Camada de Entrada Camada Convolucional Camada Convolucional Camada Densa

Camada de Saída



Introdução ao tema -Ligação com a biologia



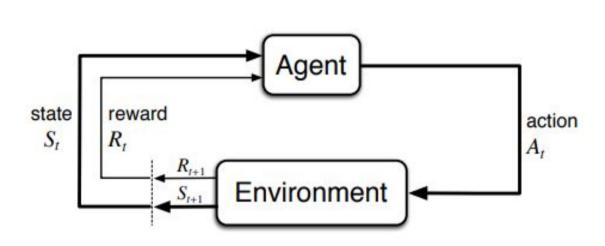


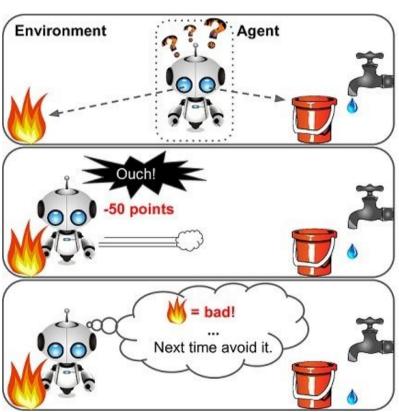
Experimento de Pavlov (cão de Pavlov)

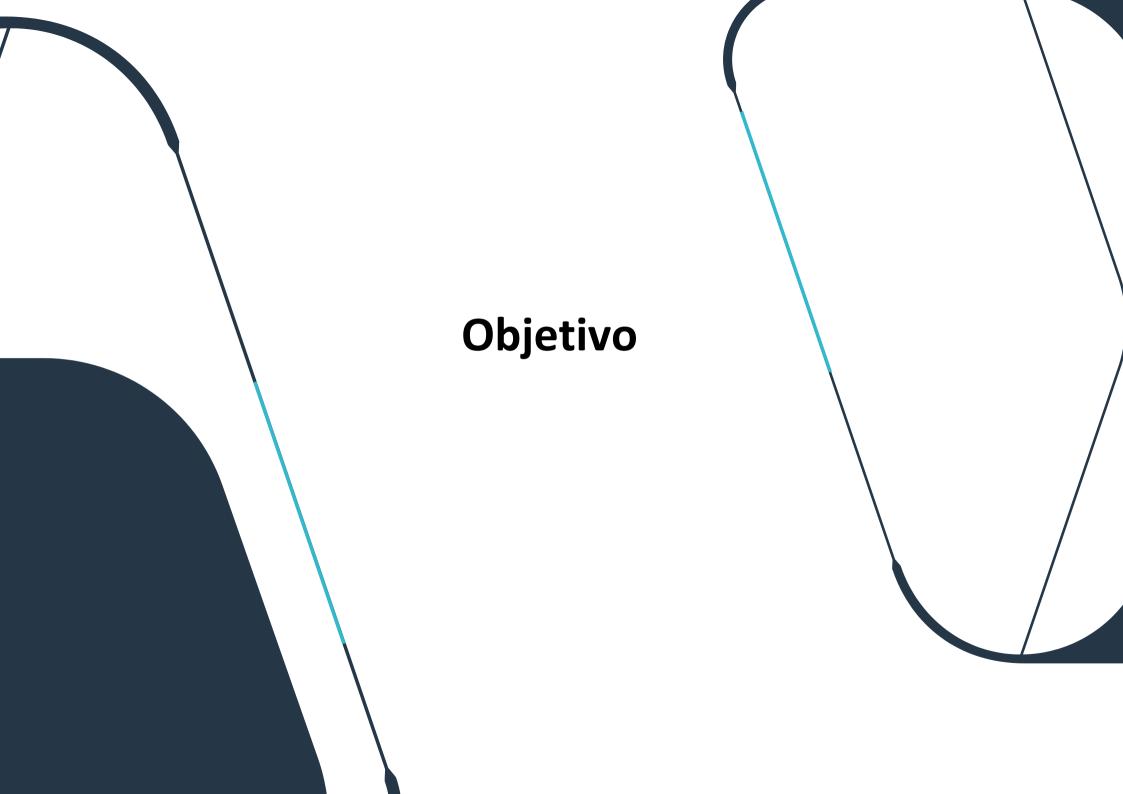
Experimento de Watson (little Albert)



Introdução ao tema -Markov Decision Process



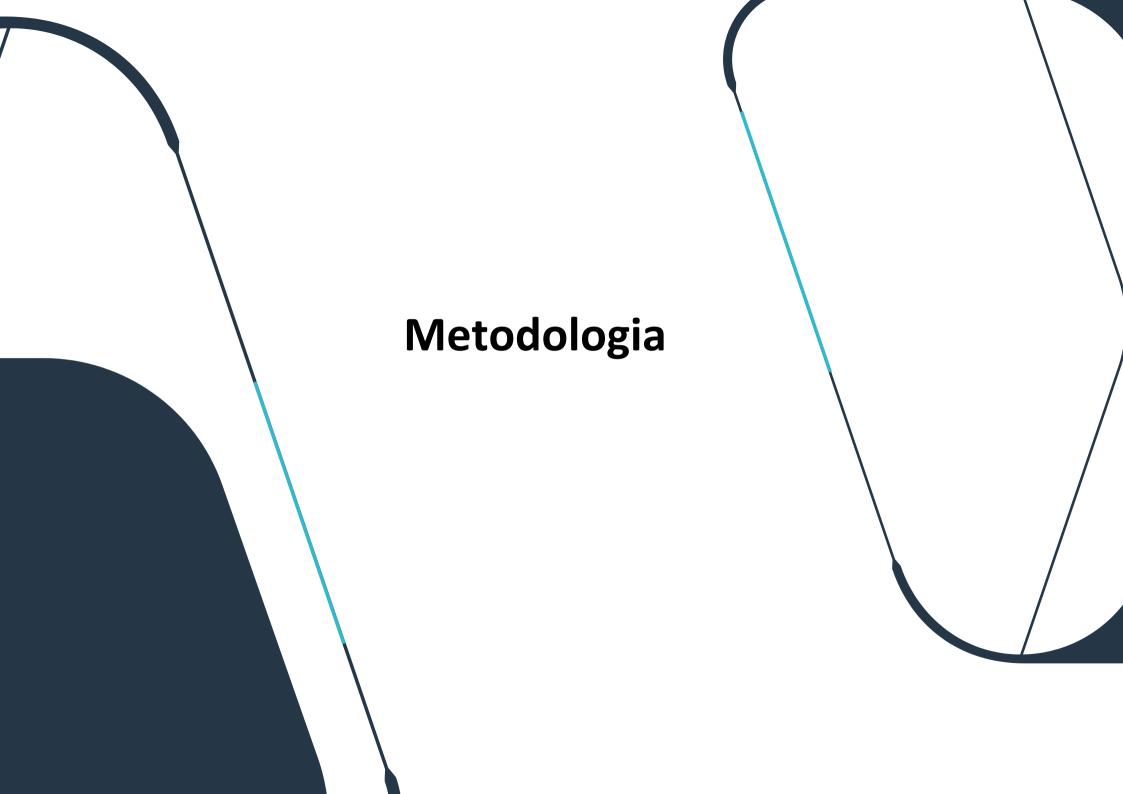






Objetivo

Demonstrar a capacidade de algoritmos de reinforcement learning de aprender tarefas complexas, como jogos de atari, sem ter antes dados anotados como acontece na categoria de algoritmos supervisionados.





Metodologia - Ferramentas



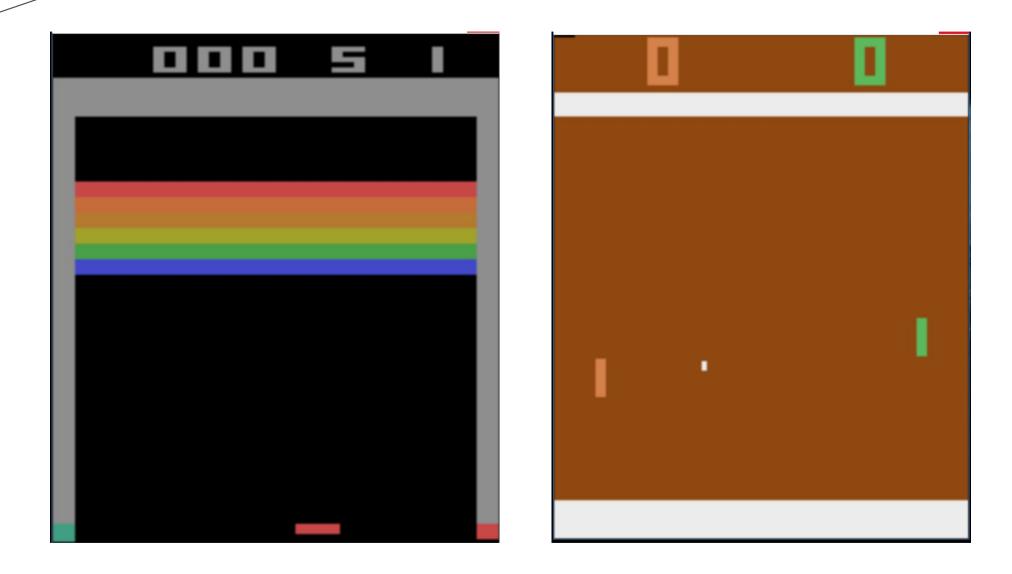








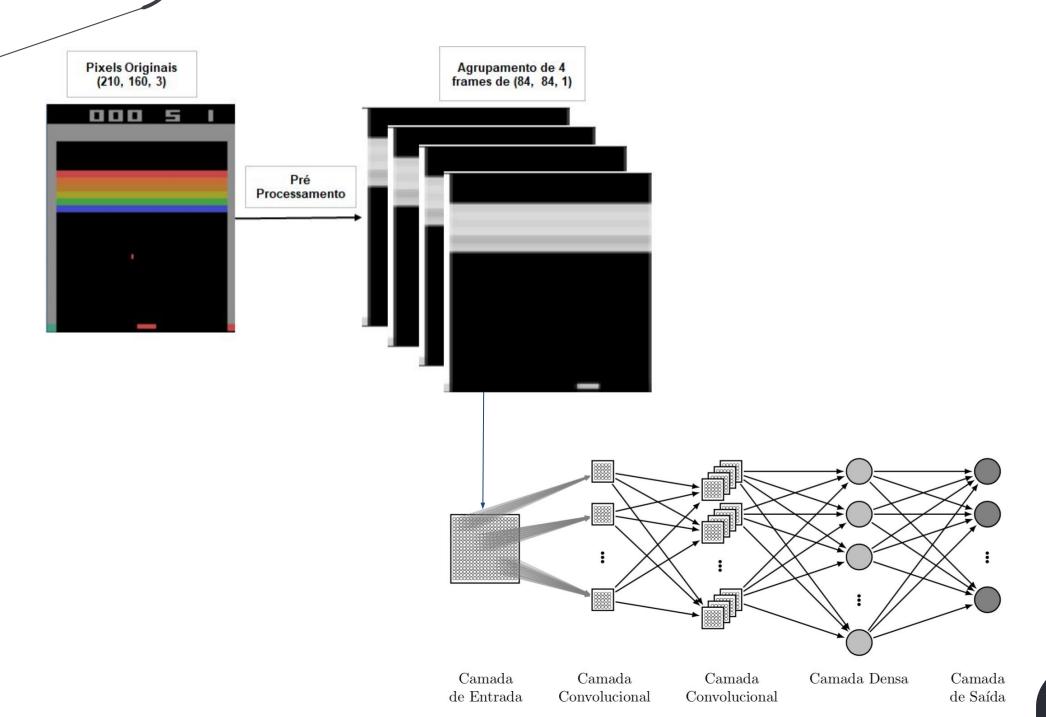
Metodologia - Ambiente



Jogos Breakout e Pong utilizando a ferramenta gym

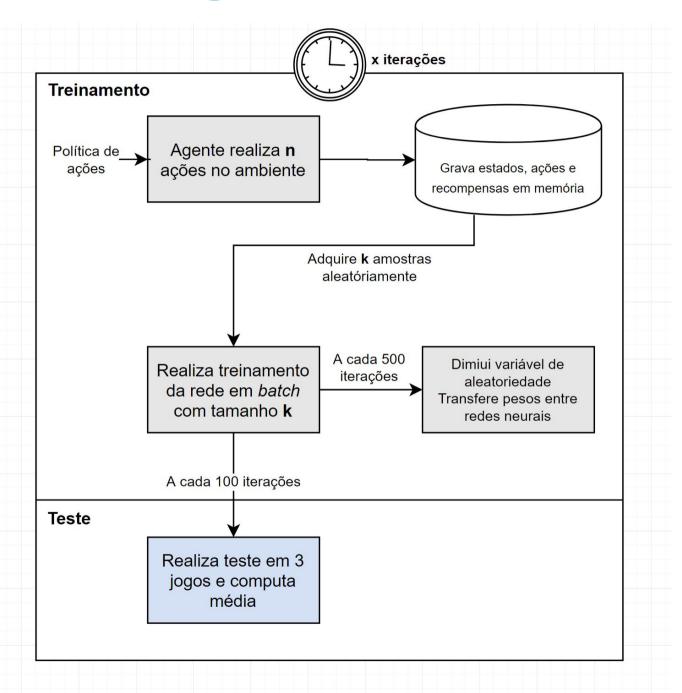


Metodologia - Pré Processamento





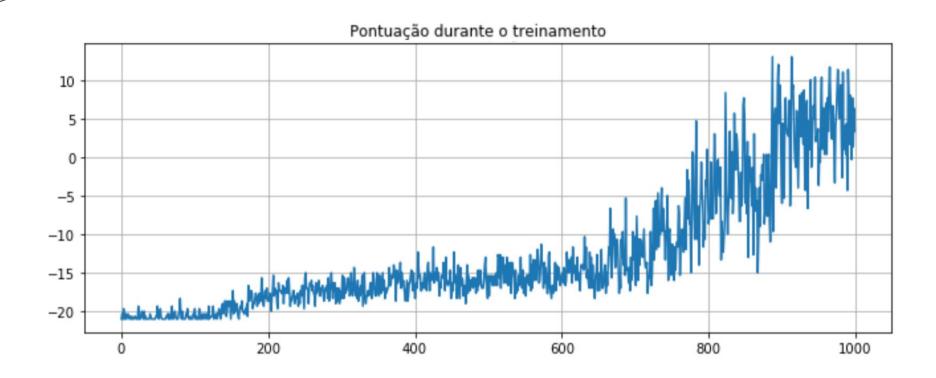
Metodologia - Treinamento







Resultados - Pong



Após treinamento:

Dados de 500 jogos

Jogo	Média	Desvio	Máximo	Mínimo
Pong	19.51	0.88	21.00	14.00

Benchmarks:

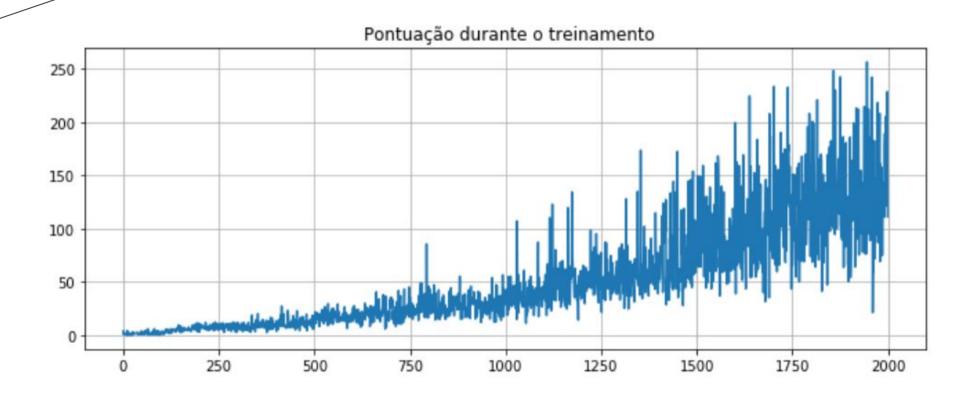
Média da pontuação humana durante 2 horas de jogo:

-3 (Minih et al. 2013)

Pontuação máxima do jogo: 21



Resultados - Breakout



Após treinamento:

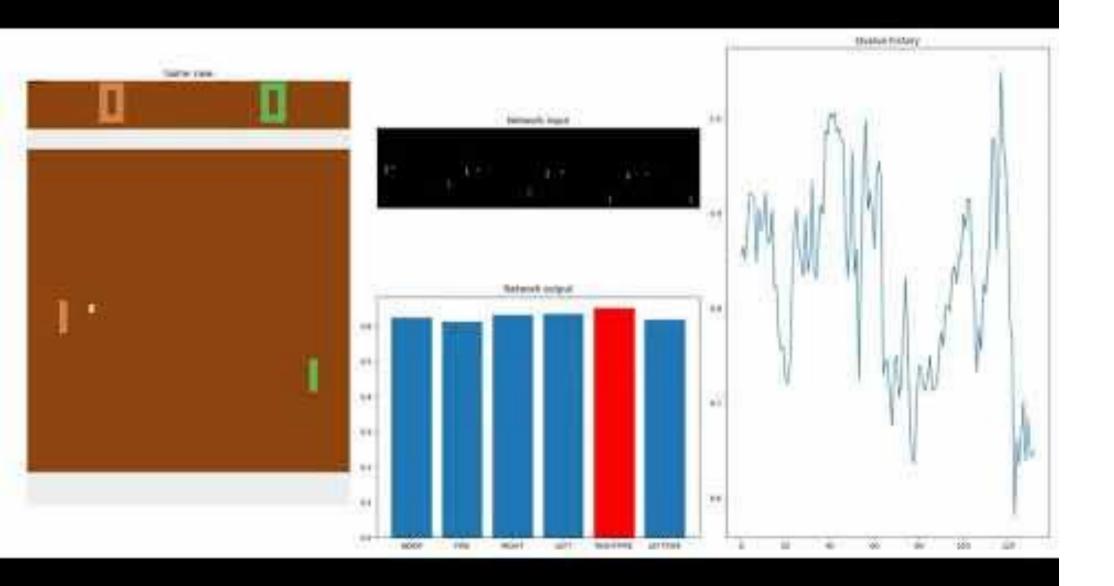
Dados de 500 jogos

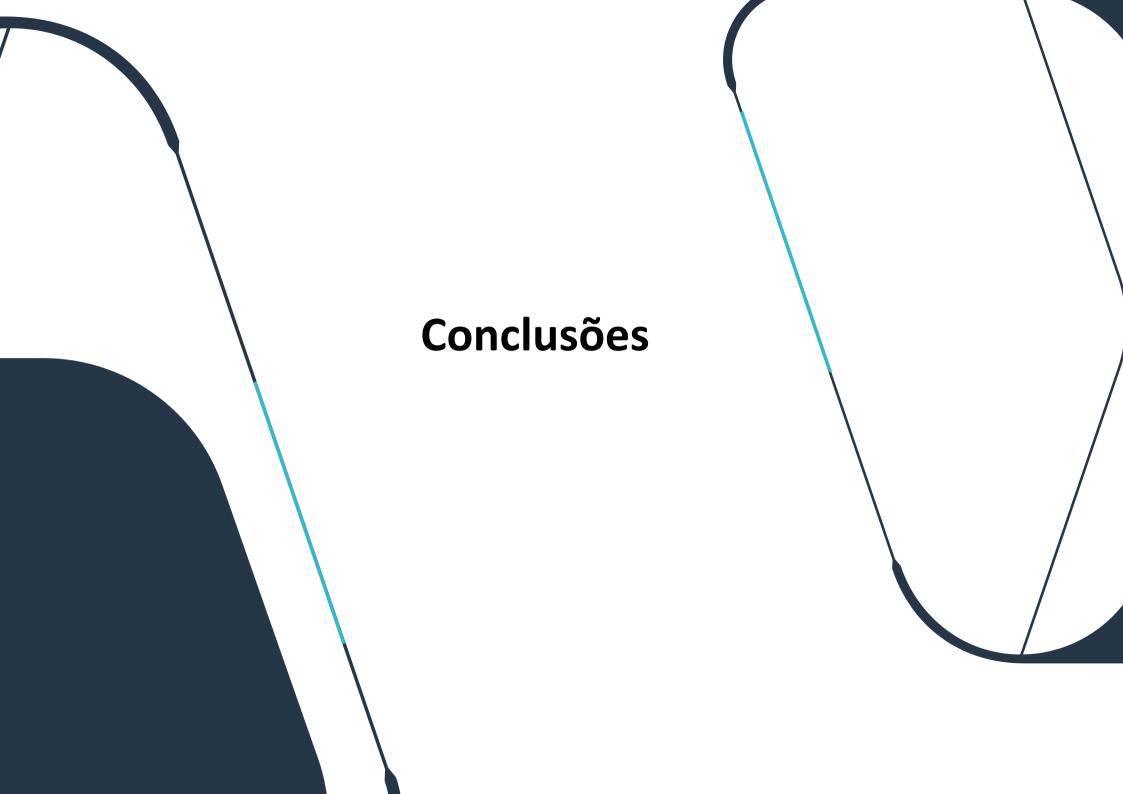
Jogo	Média	Desvio	Máximo	Mínimo
Breakout	208.48	59.89	361.00	41.00

Benchmarks:

Média da pontuação humana durante 2 horas de jogo: 31 (Minih et al. 2013)

Pontuação máxima do jogo: 448







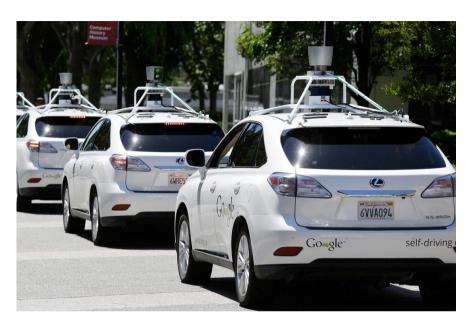
Conclusões

- Foi capaz de realizar tarefas complexas a partir de suas experiências.
- Performar bem nos dois jogos.
- Desafio de treinamento.
- Muito sensível a hiperparâmetros.
- Esparsidade das recompensas.
- Dificuldade na execução do algoritmo.



Conclusão - Outras aplicações











Bibliografia

- [1] FORMIGONI, M. Neurobiologia: Mecanismos De Reforço E Recompensa E Os Efeitos Biológicos E Os Efeitos Comuns Às Drogas De Abuso. Aberta, 2017.
- [2] LORICA, B. **Practical applications of reinforcement learning in industry**. Disponível em https://www.oreilly.com/ideas/practical-applications-of-reinforcement-learning-in-industry Acesso em: 20 setembro 2018.
- [3] SUTTON, R.; BARTO, A. Reinforcement Learning: an Introduction, second edition, The MIT Press, 2018.
- [4] ALPHAYDIN, E. Introduction To Machine Learning, The MIT Press, 2014.
- [5] SILVER, D. **Lecture 2: Markov Decision Processes**. Disponível em http://www.aberta.senad.gov.br/medias/original/201704/20170424-094615-001.pdf Acesso em: 20 setembro 2018.
- [6] MINIH, V. et al. Playing Atari with Deep Reinforcement Learning, DeepMind Technologies, 2013.
- [7] LIN, L. **Self-Improving Reactive Agents Based On Reinforcement Learning,** Planning and Teaching, Kluwer Academic Publishers, 1992.
- [8] HASSELT, H. GUEZ, A. SILVER, D. **Deep Reinforcement Learning with Double Q-learning,** Google DeepMind, 2016.
- [9] SILVER, D. Hassabis, D. **AlphaGo Zero: Learning From Scratch**. Disponível em https://deepmind.com/blog/alphago-zero-learning-scratch/> Acesso em: 25 setembro 2018.

