

MBA - Inteligência Artificial e Big Data

**Simulando multi-agentes em jogos de tabuleiro
por turnos, utilizando aprendizagem por reforço
via Value-Decomposition Network
e Gymnasium API**

Aluno: Jorge de Melo Feldmann

Orientador: Prof. Dr. Fernando Santos Osório



Introdução

O mundo dos games é vasto, popular e em constante inovação, sendo **um dos mercados onde a inteligência artificial aflora**.

Sob esse contexto, neste trabalho foi **desenvolvido um jogo baseado em turnos e tabuleiro que simula caçador e caça**, onde foi **implementado e analisado o aprendizado por reforço de comportamentos de um sistema multi-agente** (com mais de um caçador e uma presa).

Buscou-se um aprendizado que **valorize a cooperação entre caçadores**, na captura da presa. A implementação foi realizada usando o **Simulador Gymnasium**, com os agentes treinados pelo **algoritmo Value-Decomposition Network**.

Foi **avaliado o desempenho do algoritmo**, baseado em técnicas de Aprendizado de Máquina, treinado através do **aprendizado por reforço e auto-treino**, através do relato de um experimento, como se dá o processo de preparação, execução e avaliação da curva de aprendizado mediante o algoritmo proposto.



Motivação

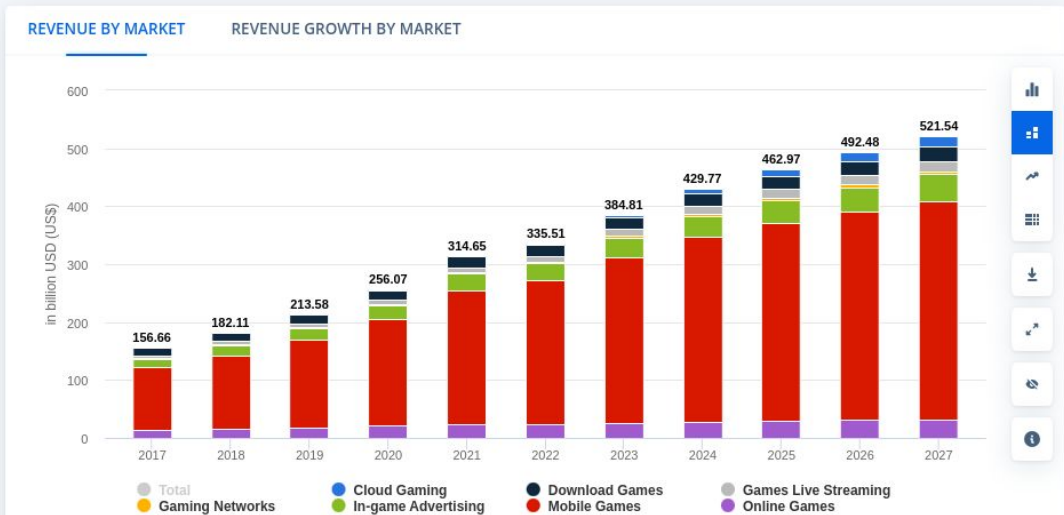


Aprendizado pessoal: videogame é um segmento pessoal.

Econômico: Dota International: 2,75 milhões de espectadores e com prêmios de mais de 40 milhões de dólares.

Previsão: a “batalha de robôs”, que hoje já acontece, será levada ao digital na forma de IA vs IA, nos jogos eletrônicos.

Revenue



Notes: Data shown is using current exchange rates and reflects market impacts of the Russia-Ukraine war.

Most recent update: Aug 2023

Source: Statista Market Insights

Questões de pesquisa e objetivos

Utilizando técnicas de Aprendizado de Máquina, a plataforma Gymnasium API e o algoritmo Value-Decomposition Network, pretende-se responder às seguintes questões de pesquisa:

1. **É possível aprender uma estratégia inteligente de cooperação multi-agente em um jogo de tabuleiro em turnos, usando o aprendizado por reforço?**
2. **É possível testar e validar esse tipo de experimento com o algoritmo Value-Decomposition Network simulado na plataforma Gymnasium API ?**



Base conceitual

- Inteligência artificial
- Redes neurais artificiais
- Redes neurais profundas
- Aprendizado de máquina
- Aprendizado por reforço
 - Multi-agentes
 - Self-play
- Value-Decomposition Network
- Gymnasium API ("OpenAI Gym")



Estado da arte



Subscribe

Latest Issues

SCIENTIFIC
AMERICAN

Sign

ARTIFICIAL INTELLIGENCE

AI's Victories in Go Inspire Better Human Game Playing

Famed AI wins in Go let human players rethink their moves in a whole new way

By Emily Willingham on March 13, 2023

South Korean professional Go player Lee Sedol reviews a match with other professionals after the fourth match against Google's artificial intelligence program AlphaGo during the Google DeepMind Challenge Match in Seoul on March 13, 2016. Credit: Google/Getty Images

In 2016 a computer named AlphaGo made headlines for defeating then world champion Lee Sedol at the ancient, popular strategy game Go. The “superhuman” artificial intelligence, developed by [Google DeepMind](#), lost only one of the five rounds to Sedol, generating comparisons to Garry Kasparov’s 1997 chess loss to IBM’s Deep Blue. Go, which [involves players facing off by moving black and white pieces called stones](#) with the goal of occupying territory on the game board, had been viewed as a [more intractable challenge](#) to a machine opponent than [chess](#).

READ THIS NEXT

SPONSORED

How next-gen CRISPR could transform medicine and agriculture

WEATHER

Hurricane Hilary Brings Major Flood Risks to U.S. Southwest

Andrea Thompson

Estado da arte



OpenAI
Five



TechTalks

HOME

BLOG ▾

TIPS & TRICKS ▾

WHAT IS ▾

INTERVIEW

Home > Blog > AI defeated human champions at Dota 2. Here's what we learned.

Blog

AI defeated human champions at Dota 2. Here's what we learned.

By Ben Dickson - April 17, 2019

👍 Curtir 41



Facebook



Twitter



Reddit



LinkedIn

OpenAI Five, a team of five neural networks, defeated world champions at a best-of-three match of popular strategy game Dota 2 (Image source: [Twitch](#))

AI made history on Saturday as neural networks defeated human world champions in a best-of-three contest at Dota 2, a popular and complex online strategy game. OpenAI Five, the AI agent developed by the namesake research lab, managed to perform the feat after **losing in a match-up against professional Dota 2 players** in a famous tournament in August.

Estado da arte



UNITE.AI

AI TOOLS ▾ CERTIFICATIONS ▾ EVENTS ▾ NEWS ▾ INTERVIEWS ▾ THOUGHT LEADERS

ARTIFICIAL INTELLIGENCE

DeepMind's AI Reaches Highest Rank of StarCraft II



Updated on December 9, 2022
By Alex McFarland

DeepMind's AlphaStar, an artificial intelligence (AI) system, has reached the highest level in StarCraft II, an extremely popular and complex computer game. The AI outperformed 99.8% of all registered human players.

It took the AI system 44 days of training to be able to reach the level. It used recordings of some of the best human players, and it learned from them until eventually going up against itself.

"AlphaStar has become the first AI system to reach the top tier of human performance in any professionally played e-sport on the full unrestricted game under professionally approved conditions," said David Silver, a researcher at DeepMind.

"Ever since computers cracked Go, chess and poker, the game of StarCraft has emerged, essentially by consensus from the community, as the next grand challenge for AI," Silver said. "It's considered to be the game which is most at the limit of human capabilities."

The work was published in the scientific journal Nature.



Estado da arte




nature

[Explore content](#) [About the journal](#) [Publish with us](#)

[nature](#) > [articles](#) > article

Article | [Open Access](#) | [Published: 07 June 2023](#)

Faster sorting algorithms discovered using deep reinforcement learning

[Daniel J. Mankowitz](#) , [Andrea Michi](#), [Anton Zhernov](#), [Marco Gelmi](#), [Marco Selvi](#), [Cosmin Paduraru](#), [Edouard Leurent](#), [Shariq Iqbal](#), [Jean-Baptiste Lespiau](#), [Alex Ahern](#), [Thomas Köppe](#), [Kevin Millikin](#), [Stephen Gaffney](#), [Sophie Elster](#), [Jackson Broshear](#), [Chris Gamble](#), [Kieran Milan](#), [Robert Tung](#), [Minjae Hwang](#), [Taylan Cemgil](#), [Mohammadamin Barekatin](#), [Yujia Li](#), [Amol Mandhane](#), [Thomas Hubert](#), ... [David Silver](#) [+ Show authors](#)

[Nature](#) **618**, 257–263 (2023) | [Cite this article](#)

352k Accesses | **1** Citations | **1155** Altmetric | [Metrics](#)

Abstract

Fundamental algorithms such as sorting or hashing are used trillions of times on any given day¹. As demand for computation grows, it has become critical for these algorithms to be as performant as possible. Whereas remarkable progress has been achieved in the past², making further improvements on the efficiency of these routines has proved challenging for both human scientists and computational approaches. Here we show how artificial intelligence can go beyond the current state of the art by discovering hitherto unknown routines. To realize this, we formulated the task of finding a better sorting routine as a single-player game. We then trained a new deep reinforcement learning agent, AlphaDev, to play this game. AlphaDev discovered small sorting algorithms from scratch that outperformed previously known human benchmarks. These algorithms have been integrated into the LLVM standard C++ sort library³. This change to this part of the sort library represents the replacement of a component with an algorithm that has been automatically discovered using reinforcement learning. We also present results in extra domains, showcasing the generality of the approach.



Desenvolvimento

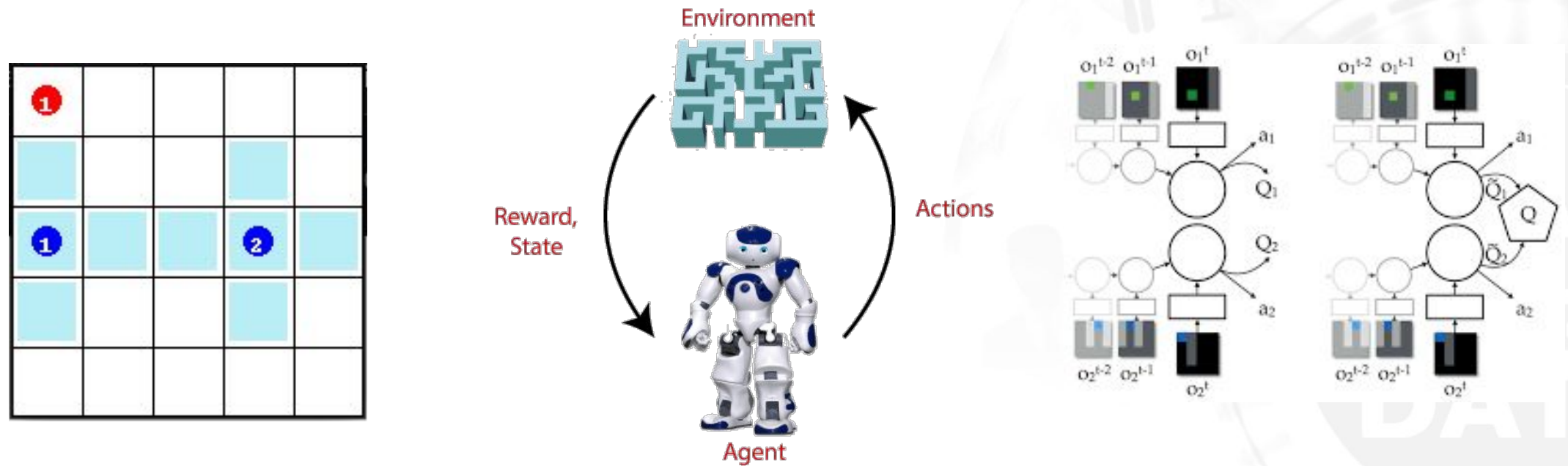
Caça e caçador



1				
1			2	



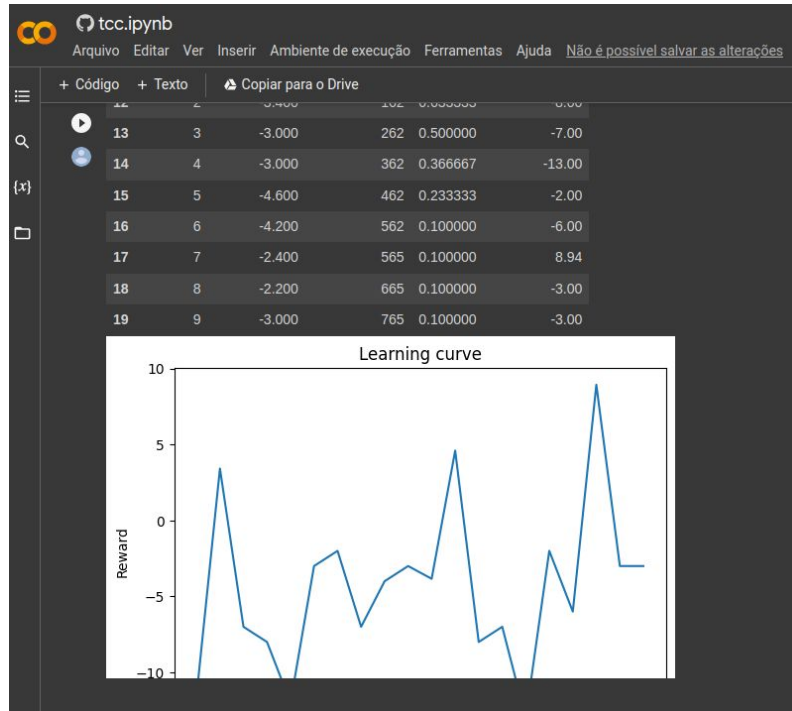
Desenvolvimento



 Gymnasium



Desenvolvimento



Experimento:

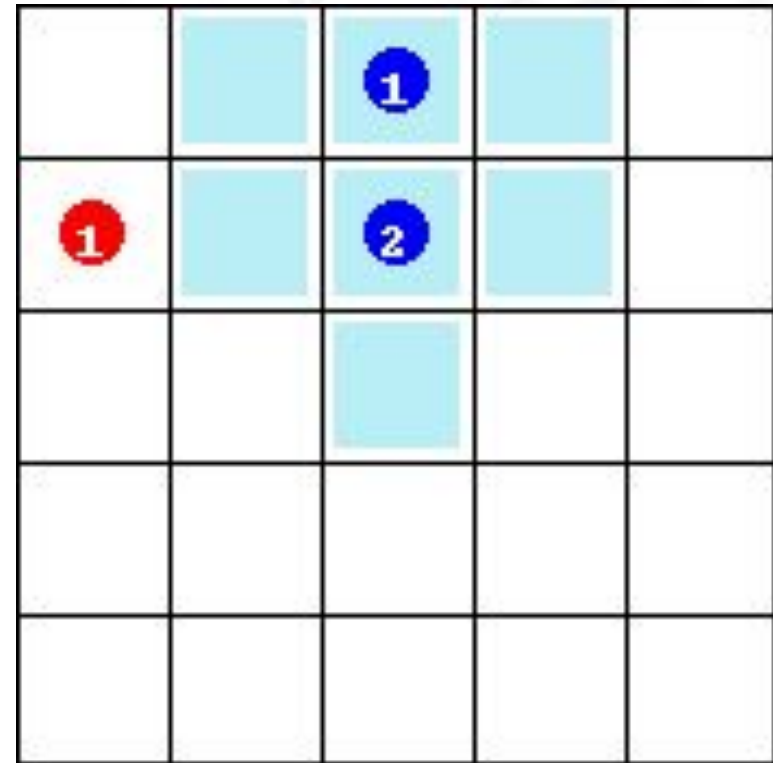
<https://github.com/jotafeldmann/tcc-ia-big-data>



Análise e discussão dos dados

Primeiro terço do experimento, detectada exploração e baixa recompensa.

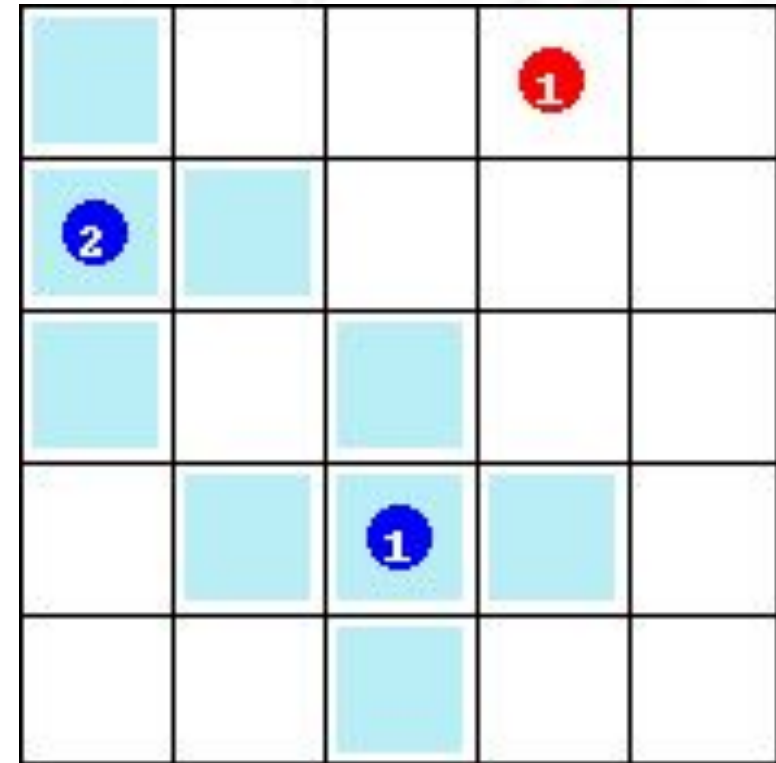
A presa (em vermelho) anda livremente.



Análise e discussão dos dados

Segundo terço do experimento, início do processo de cooperação.

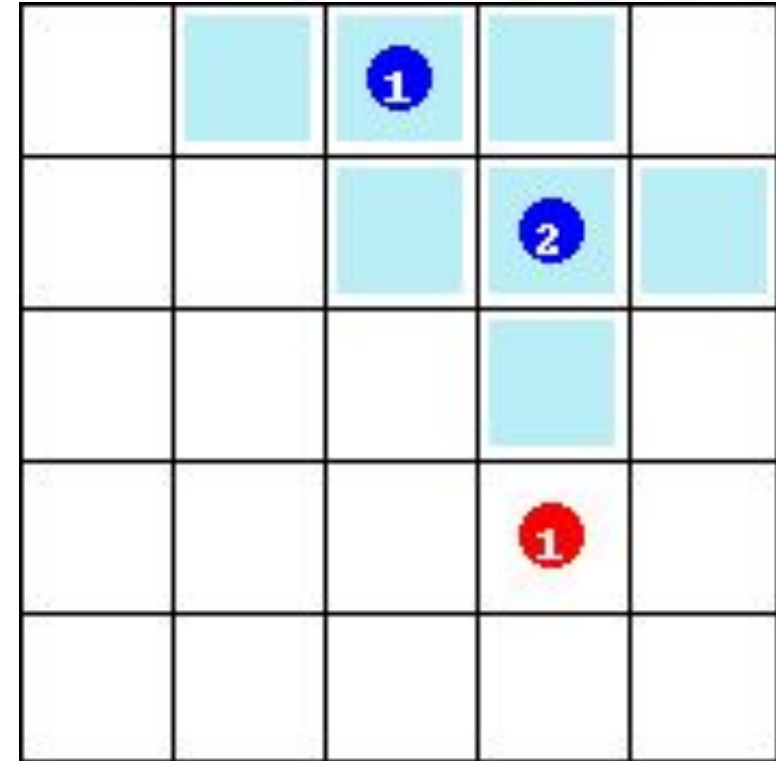
Ainda ocorrem vitórias da presa, porém crescem as vitórias dos caçadores.



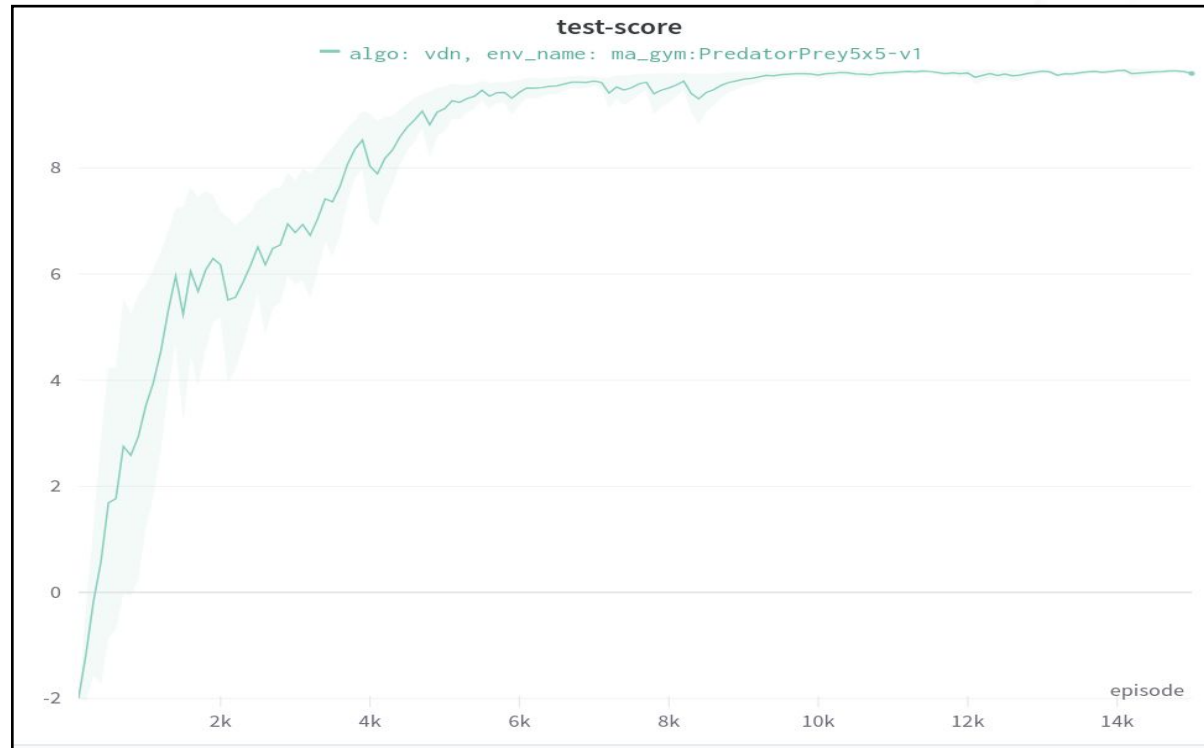
Análise e discussão dos dados

Último terço do experimento, detectada cooperação total, aumento expressivo da curva de recompensa.

Vitória total dos caçadores, demonstrando que agem de forma a cercar a presa.



Análise e discussão dos dados



Análise e discussão dos dados

Etapas	Tempo total de execução	Total de episódios	Média das recompensas
Treino	43m 3s	15000	9,683
Teste	32m 15s	15000	9,47
Total	1h 15m 18	30000	n/a



Conclusão

Respondendo-se à pergunta inicial, **foi possível recriar um experimento, simulando um jogo multi-agente de tabuleiro por turnos, onde os agentes foram capazes de trabalhar de forma cooperativa, vencendo a maior parte das partidas em sua fase de testes**, mediante uma equipe adversária programada por heurísticas, **e realizar uma análise de desempenho do modelo de aprendizado e entendimento de como isso pode ser feito**. Destacam-se as percepções e observações adquiridas durante a análise e as estratégias que foram detectadas visualmente nas análises durante o experimento, uma vez que não havia nada previsto a priori.

Um aspecto importante que está além do escopo deste trabalho, e foi percebido durante o processo, é a possibilidade de comparar outros modelos de aprendizado dado o ambiente padronizado que a Gymnasium API proporciona.

Por fim, foi **revelador e lúdico poder colocar em prática o conhecimento aprendido**, ao elaborar um experimento e verificar como é **possível utilizar aprendizado de máquina para simular um jogo**.



Agradeço pela atenção =)

Contato

<https://www.linkedin.com/in/jotafeldmann>

jota.feldmann@gmail.com

<https://github.com/jotafeldmann>

