Implementation of RL Algorithms in OpenAI Gym

Douglas Trajano, Sirleno Vidaletti
Pontifical Catholic University of Rio Grande do Sul - PUCRS
School of Technology. Porto Alegre, Brazil
{douglas.trajano, sirleno.vidaletti}@edu.pucrs.br

Abstract

Flappy Bird is an electronic game created in 2013. The objective in the game is to earn as many points as possible by controlling a bird, without letting it crash into the pipes. In this work, we will develop reinforcement learning algorithms such as Q-Learning and Deep Q-Network (DQN) to learn how to control the ship maximizing the score. The LunarLander environment is provided by OpenAI Gym, a toolkit for reinforcement learning research.

Introduction

The Flappy Bird is a game developed by Vietnamese programmer Dong Nguyen. The game, the player controls a bird, attempting to fly between columns of green pipes without hitting them. The game was released in May 2013 but received a sudden rise in popularity in early 2014 and became a sleeper hit. Flappy Bird received poor reviews from some critics, who criticized its high level of difficulty and alleged plagiarism in graphics and game mechanics, while other reviewers found it addictive. Flappy Bird was removed from both the App Store and Google Play by its creator on February 10, 2014. He claims that he felt guilt over what he considered to be its addictive nature and overuse.

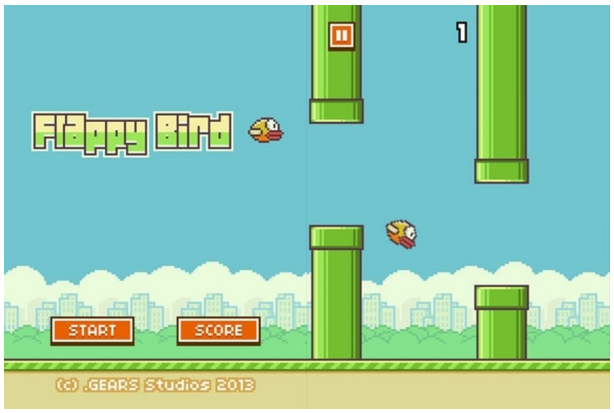

Figure 1: Flappy Bird Game

Learning to control agents directly from high-dimensional sensory inputs like vision and speech is one of the long-standing challenges of reinforcement learning (RL). The basic idea behind many reinforcement learning algorithms is to estimate the action-value function. The goal of the agent (powered by a reinforcement learning algorithm) is to interact with the emulator by selecting actions in a way that maximises future rewards.(Mnih et al. 2013)

We will develop some reinforcement learning algorithms that will be used to train our agents in the OpenAI Gym environment. OpenAI Gym is a toolkit for reinforcement learning research. It includes a growing collection of benchmark problems that expose a common interface.(Brockman et al. 2016)

Technical Approach

Our technical approach consists of two parts. The first part is the definition of the environment. The environment that will be explored in this project is provided by OpenAI Gym. The second part is the reinforcement learning algorithm, which we will be develop by ourselves.

Environment

The environment is a Python class that simulates the game and returns the observations and rewards. It uses the OpenAI Gym API.

The state observation is composed by RGB-arrays (images) representing the game's screen.

Two actions are available: do nothing and jump.

RL Algorithms

The RL algorithms will be implemented in Python. The algorithms are responsible for deciding the actions of the agent. We will develop a base class with a random policy, which will be used to define the API between the environment and the algorithms. We want to develop at least two different RL algorithms:

Q-Learning is an off-policy TD control algorithm. The learned action-value function, Q, directly approximates q_*, the optimal action-value function, independent of the policy being followed.(Sutton and Barto 2018)

Deep Q-Network (DQN) combines Q-Learning with deep neural networks to let RL work for complex, high-dimensional environments, like video games, or robotics. A critical component of DQN-style algorithms is memory buffer known as experience replay, it holds the most recent transitions collected by the policy.(Fedus et al. 2020)

Two different approaches of the experience replay will be develop and tested.

- **Experience Replay**: The most basic sampling strategy, it uses uniform sampling, whereby each transition in the buffer is sampled with equal probability.(Fedus et al. 2020)
- **Prioritized Experience Replay**: Extends experience replay function by learning to replay memories where the real reward significantly diverges from the expected reward, letting the agent adjust itself in response to developing incorrect assumptions.(Schaul et al. 2015)

At the end of the project, we will compare the performance of the developed algorithms and buffers.

Project Management

We want to validate some reinforcement learning algorithms in the Flappy Bird environment. We will use the OpenAI Gym environment to train our agents. We will work under the agile methodology of Scrum, so in each sprint we will have a planning session to define the scope of the work, and a sprint review to validate the work. The deadline for submitting papers is Dec 7, 2021. Below you can see a table with the expected sprints and the start and end dates.

Sprint	Start	End
1	Oct 25, 2021	Oct 31, 2021
2	Nov 1st, 2021	Nov 7, 2021
3	Nov 8, 2021	Nov 14, 2021
4	Nov 15, 2021	Nov 21, 2021

An estimate of tasks is provided below

- **Sprint 1**: Validate OpenAI environment and develop Deep Reinforcement Learning algorithm.
- **Sprint 2**: Develop a Temporal-Difference algorithm.
- **Sprint 3**: Evaluate the results.
- **Sprint 4**: Revision and paper development.

Conclusion

In this work, we will develop reinforcement learning algorithms and compare them with each other. We will use the OpenAI Gym environment to train our RL agents. We believe that the algorithms will be able to learn how to act in the proposed environment. At the end of the project, we will compare the performance of each algorithm.

References

Brockman, G.; Cheung, V.; Pettersson, L.; Schneider, J.; Schulman, J.; Tang, J.; and Zaremba, W. 2016. Openai gym.

Fedus, W.; Ramachandran, P.; Agarwal, R.; Bengio, Y.; Larochelle, H.; Rowland, M.; and Dabney, W. 2020. Revisiting fundamentals of experience replay. In *International Conference on Machine Learning*, 3061–3071. PMLR.

Mnih, V.; Kavukcuoglu, K.; Silver, D.; Graves, A.; Antonoglou, I.; Wierstra, D.; and Riedmiller, M. 2013. Playing atari with deep reinforcement learning. *arXiv preprint arXiv:1312.5602*.

Schaul, T.; Quan, J.; Antonoglou, I.; and Silver, D. 2015. Prioritized experience replay. *arXiv preprint arXiv:1511.05952*.

Sutton, R. S., and Barto, A. G. 2018. *Reinforcement learning: An introduction*. MIT press.