Comparison of multiple Reinforcement Learning Algorithms on Atari Game

Aryan Singh
Computer and Information Science
University of Florida
Gainesville, FL 32611
aryansingh@ufl.edu

Vaibhav Kulkarni
Computer and Information Science
University of Florida
Gainesville, FL 32611
kulkarniv@ufl.edu

1 Introduction

When reinforcement learning was initially introduced in the 1950s, it was divided into two themes: one focused on establishing learning methods through trial and error, while the other gave a more theoretical framework for solving optimal control issues. Reinforcement learning emerged as a more formalized field of study and development in the 1980s when these practical and theoretical methodologies merged. For example, Richard Sutton and Andrew Barto emphasized theories like optimal control and dynamic programming at the time and crucial component ideas like temporal difference learning, dynamic programming, and function approximation. Fast forward to the 2000s, when deep learning started revolutionizing reinforcement learning by removing the need to manually construct features and allowing the use of raw sensor data (such as the pixels of an image rather than a segmented image).

But before we go any further let us define what exactly is reinforcement learning?

In contrast to supervised learning (which utilizes labeled training data) and unsupervised learning (which draws inferences from input data without labeled answers), reinforcement learning entails a system making short-term judgments using trial and error to optimize towards a long-term objective. The reinforcement learning agent learns the actions needed to maximize rewards over a longer period, whereas deep learning is employed to construct mathematical representations of relevant variables.

In recent years, there have been a lot of breakthroughs in the field of reinforcement learning due to the influx of deep learning methods. Some of the success stories include the following:

- In late 2013, DeepMind achieved a breakthrough in the world of reinforcement learning: using deep reinforcement learning, they implemented a system that could learn to play many classic Atari games with human (and sometimes superhuman) performance. The computer program has never seen this game before and does not know the rules. It learned by using deep reinforcement learning to maximize its score given only the pixels and game score as the input. They had used Deep Q Network.
- Alpha Go is a computer system developed by Google DeepMind that can play the game Go. The game of Go starts with an empty board. Each player has an effectively unlimited supply of pieces (called stones), one taking the black stones, the other taking white. The main objective of the game is to use your stones to form territories by surrounding vacant areas of the board. Google DeepMind's Challenge Match was a five-game Go match between 18-time world champion Lee Sedol and AlphaGo played in Seoul, South Korea between 9 and 15 March 2016. AlphaGo won all but the fourth game.
- The next from Deep Mind was Alpha Zero. Alpha Zero achieved within 24 hours a superhuman level of play in the games of chess and shogi (Japanese chess) as well as Go, and convincingly defeated a world-champion program in each case. Previous versions of AlphaGo were initially trained on thousands of human amateur and professional games to

35th Conference on Neural Information Processing Systems (NeurIPS 2021).

learn how to play the game Go. Alpha Zero, slips the Supervised Learning step and learns to play simply by playing against itself, starting completely from random play.

- Dota 2 is a multiplayer online battle arena video game. It is played in matches between two teams of five players, with each team occupying and defending their own separate base on the map. Open AI Five is a team of 5 neural networks that have started defeating amateur human teams at Dota 2. The program defeated a human in 2017.

Due to the reasons stated above and the advancement in the above field, we propose to build and train a reinforcement learning model on the game of Mario.

2 Importance of the papers

We propose to read 4 research papers pertaining to the field which all deal with either some new reinforcement learning algorithm or an improvement over the previous ones.

- **Playing Atari with Deep Reinforcement Learning:** The first paper (published in 2013) from the Deep Mind team introduced how the Deep Learning model can successfully learn control policies from high dimensional sensory input. The paper used CNN which was trained with a variant of Q-learning to take in raw pixels and output a value function to estimate reward. The proposed model was used on seven Atari 2600 games and the model outperformed all the previous models in six games and surpassed human experts on three.

  `https://arxiv.org/pdf/1312.5602.pdf`

- **Deep Reinforcement Learning with Double Q-learning:** This paper (published in 2015) from the Deep Mind team affirmatively pointed out the overestimation problem that DQN networks suffer. This paper introduced a variation of the DQN algorithm and showed that this algorithm not only reduces the observed overestimation but also leads to better performance in several games.

  `https://arxiv.org/pdf/1509.06461.pdf`

- **Prioritized Experience Replay:** This paper (published in 2016) again from the Deep Mind team was again an improvement over the DQN model. The paper introduced a new framework for prioritizing experience so as to replay important transitions more frequently, and therefore learn more efficiently. In prior work, experience transitions were uniformly sampled from a replay memory which would simply replay transitions at the same frequency at which they were originally experienced, regardless of their significance. This new framework achieved a new state-of-the-art, outperforming DQN with a uniform replay on 41 out of 49 games.

  `https://arxiv.org/pdf/1511.05952.pdf`

- **Proximal Policy Optimization Algorithms:** This paper (published in 2017) from the Open AI team discussed a new technique compared to what we have seen so far. The paper introduced a new family of policy gradient methods for reinforcement learning, which alternate between sampling data through interaction with the environment and optimizing a ”surrogate” objective function using stochastic gradient ascent. the advantage of this algorithm is that they are much simpler to implement, more general, and has a better sample complexity (empirically). The authors showed that PPO outperforms other online policy gradient methods, and overall strikes a favorable balance between sample complexity, simplicity, and wall-time on an experimental test of Atari game playing and simulated robotic locomotion tasks.

  `https://arxiv.org/pdf/1707.06347.pdf`

3 Algorithms and Dataset

3.1 Algorithms we plan to implement:

- Deep Q Network
- Double Deep Q Network
- Double Deep Q Network (Using Prioritized Experience Replay)
- Proximal Policy Optimization

3.2 Data Set:

- We would use Open AI Gym to create the game environment. We would use library gym-super-mario-bros 7.3.2 to create the game environment and use nes-py 8.1.8 to get the joypad controls to take actions in the environment.

We propose to compare and contrast each of the algorithms and produce its pros and cons.