Project Report for CIS 419/519
The Comparison of Different Reinforcement Learning in Atari Games

kun Qian KUNQIAN@SEAS.UPENN.EDU
Yuxiang Qiao QIAOYX@SEAS.UPENN.EDU
Yuchen Yang YYUC@SEAS.UPENN.EDU

1. Objective

In the real world, we care about how robots interact with the environment and perform a task under typical setting. However, Some environments are difficult to set up because setting up a scenario costs a lot of money or Some physical conditions are hard to reach(such as high speed, high temperature). Therefore, how to implement this realistic situation into a virtual scene and then test the performance of our robot in this setting is of vital importance. We care about whether an algorithm works perfectly under such a scenario.

Our goal is to implement agents based on Reinforcement learning to play Super Mario or other Atari games in OpenAI Gym(Brockman et al., 2016) automatically. We will then compare the accuracy and complexity for different learning algorithms. Then we will try to improve our models by modification of environment definition, reconstruction of replay buffer and other related strategies.

2. Related Work

Reinforcement learning (RL) is a branch of machine learning that is concerned with making sequences of decisions(Brockman et al., 2016).

There are three main categories of algorithms for RL: Value-Based: Q-learning(Watkins & Dayan, 1992) and Deep Q-learning(DQN)(Mnih et al., 2013) are the most popular algorithm, Policy-Based: Policy Gradient(Sutton et al., 1999) and other classical algorithm. The third categories of algorithms combine Value-Based and Policy-Based methods, called Actor-Critic(AC), which includes Based Actor-Critic(Sutton & Barto, 1998), Advantage Actor Critic(A2C)(Sutton & Barto, 1998), Trust Region Policy Optimization (TRPO)(Schulman et al., 2015), Proximal Policy Optimization(PPO)(Schulman et al., 2017) and so on.

Some DQN and related methods have been applied to play computer games ((Mnih et al., 2015);(Schaul et al., 2015)), especially Atari games.

3. Data

We will use the environment provided by OpenAi gym, and choose several games like Mario or Tetirs. It is easy for us to monitor the whole training process and get different values of the training models thanks to the convenience of OpenAi.

4. Method

Initially, we will implement the game of Super Mario based on DQN. then we will try other RL strategies such as policy-based strategy, to play Mario automatically.

After that, we will compare those different algorithms with dimension: accuracy which we may use cross-validation, learning curve or other methods to show; complexity (both time complexity and space complexity)

Finally, we will try to improve these models by changing the definition of environment, action or other variables or improve models by changing RL algorithms. There are two axes that can be improved: performance and complexity.

Acknowledgments

If you did this work in collaboration with someone else, or if someone else (such as another professor) had advised you on this work, your report must fully acknowledge their contributions. If you received external help or assistance on this project, you must cite these sources here in the acknowledgements section. If you do not have anything to list in this section, write simply “None.”

References

Brockman, Greg, Cheung, Vicki, Pettersson, Ludwig, Schneider, Jonas, Schulman, John, Tang, Jie, and Zaremba, Wojciech. Openai gym. *ArXiv*, abs/1606.01540, 2016.

Mnih, Volodymyr, Kavukcuoglu, Koray, Silver, David, Graves, Alex, Antonoglou, Ioannis, Wierstra, Daan, and

Riedmiller, Martin. Playing atari with deep reinforcement learning. 2013. URL http://arxiv.org/abs/1312.5602. cite arxiv:1312.5602Comment: NIPS Deep Learning Workshop 2013.

Mnih, Volodymyr, Kavukcuoglu, Koray, Silver, David, Rusu, Andrei A., Veness, Joel, Bellemare, Marc G., Graves, Alex, Riedmiller, Martin A., Fidjeland, Andreas K., Ostrovski, Georg, Petersen, Stig, Beattie, Charles, Sadik, Amir, Antonoglou, Ioannis, King, Helen., Kumaran, Dharshan, Wierstra, Daan, Legg, Shane, and Hassabis, Demis. Human-level control through deep reinforcement learning. *Nature*, 518:529–533, 2015.

Schaul, Tom, Quan, John, Antonoglou, Ioannis, and Silver, David. Prioritized experience replay. 11 2015.

Schulman, John, Levine, Sergey, Abbeel, Pieter, Jordan, Michael I., and Moritz, Philipp. Trust region policy optimization. In *ICML*, 2015.

Schulman, John, Wolski, Filip, Dhariwal, Prafulla, Radford, Alec, and Klimov, Oleg. Proximal policy optimization algorithms. *ArXiv*, abs/1707.06347, 2017.

Sutton, Richard S. and Barto, Andrew G. *Reinforcement Learning: An Introduction*. MIT Press, 1998. ISBN 0262193981. URL http://www.cs.ualberta.ca/%7Esutton/book/ebook/the-book.html.

Sutton, Richard S., McAllester, David, Singh, Satinder, and Mansour, Yishay. Policy gradient methods for reinforcement learning with function approximation. In *Proceedings of the 12th International Conference on Neural Information Processing Systems*, NIPS'99, pp. 1057–1063, Cambridge, MA, USA, 1999. MIT Press.

Watkins, Chris and Dayan, Peter. Q-learning. *Machine Learning*, 8:279–292, 1992.