Neural K-Nearest Neibhorhood Training

Guannan Hu

March 12, 2018

Abstract

1 Introduction

Until the AlphaGo defeated the 18-time world champion Lee Sedol, the reinforcement learning have a rapid development with deep neural network. such as play atari game without human knowledge [1], and AlphaZero [4] who is the extension version have been trained without human knowledge and self-play. Although the deep reinforcement learning have a great performance, there exist a lot of problems in deep reinforcement learning algorithms. [2]

- Stochastic gradient optimisation requires the use of small learning rates. Due to the global approximation nature of neural networks, high learning rates cause catastrophic interference. Low learning rates mean that experience can only be incorporated into a neural network slowly.
- Environments with sparse reward signal can be difficult for a neural network to model as there may be very few instances where the reward is non-zero. This can be viewed as a form of class imbalance where low-reward samples outnumber high-reward samples by a unknown number. Consequently, the neural network disproportionally underperforms at predicting larger rewards, making it difficult for an agent to take the most rewarding actions.
- Reward signal propagation by value-bootstrappiDeep Q-Learning Networkng techniques, such as Q-learning, results in reward information being propagated one step at a time through the history of previous interactions with the environment. This can be fairly efficient if updates happen in reverse order in which the transitions occur. However, in order to train on randomly selected transitions, and, in order to further stabilise training, required the use of slow updating *target network* further slowing down reward propagation.

Deep Q-learning Network Mnih et.al present the first deep learning model to successfully learn control policies directly from high-dimensional sensory input using reinforcement learning.[3] The model is a convolutional neural network, trained with a variant of Q-learning, whose input is raw pixels and whose output is a value function estimating future rewards.

Double Deep Q-learning Network Double Deep Q-Learning Network

Dueling Deep Q-learning Network Dueling Deep Q-Learning Network

Experience Replay and Prioritized Experience Replay

Trusted Region Policy Optimization Trust Region Policy Optimization

Proximity Policy Optimization Proximity Policy Optimization

Neural Episodic Training [2]

2 Relate Works

3 Neural K-Nearest Neighbourhood Training

4 Experiments

5 Conclusion

References

[1] Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Alex Graves, Ioannis Antonoglou, Daan Wierstra, and Martin Riedmiller. Playing atari with deep reinforcement learning. Computer Science, 2013.

[2] Alexander Pritzel, Benigno Uria, Sriram Srinivasan, Adri Puigdomnech, Oriol Vinyals, Demis Hassabis, Daan Wierstra, and Charles Blundell. Neural episodic control. 2017.

[3] David Silver, Aja Huang, Chris J Maddison, Arthur Guez, Laurent Sifre, George Van Den Driessche, Julian Schrittwieser, Ioannis Antonoglou, Veda Panneershelvam, Marc Lanctot, et al. Mastering the game of go with deep neural networks and tree search. nature, 529(7587):484–489, 2016.

[4] David Silver, Julian Schrittwieser, Karen Simonyan, Ioannis Antonoglou, Aja Huang, Arthur Guez, Thomas Hubert, Lucas Baker, Matthew Lai, Adrian Bolton, et al. Mastering the game of go without human knowledge. Nature, 550(7676):354, 2017.