DL Lab Course: Final Project

Yufeng Xiong

University of Freiburg

Feb 17th, 2017

Project Purpose

Deep Q-Network (DQN) equation for updating the target value:

$$Q_{target} = r + \gamma * \max_{a'} Q(s', a', w)$$

Problems:

- Use the same network to select and evaluate the action
- Lead to overestimation of the actions
- Harm performance and stability

Therefore, Double DQN is proposed.

Double Deep Q-Network (DDQN)

In DDQN, a target network is proposed to remove the upward bias caused by $\max_a Q(s, a, w)$.

- Primary network (w) is used to select an action
- Target network (w^-) is used to evaluate the target Q-value

DDQN equation for updating the target value:

$$Q_{target} = r + \gamma * Q(s', arg\max_{a'} Q(s', a', w), w^-)$$

Then the MSE loss equation becomes:

$$I = (r + \gamma * Q(s', arg\max_{a'} Q(s', a', w), w^-) - Q(s, a, w))^2$$

Implemention details

I implement the DDQN for the simple-maze task in assignment 4 and compare the results with DQN.

Network architecture:

- Convolutional layer 1
- Convolutional layer 2
- Convolutional layer 3
- Fully connected layer 1
- Fully connected layer 2

Key parameters:

- training steps: 100,000
- target Q-network update: every 1000 steps
- test steps: 30,000
- epsilon: 0.2

Performance Comparisons

The comparative results are as follows:

Network type	Learning rate	Test accuracy	Average steps
DQN	1.E-04	loss explodes	-
	1.E-05	99.97%	8.8
DDQN	1.E-04	100.00%	5.7
	1.E-05	99.79%	7.5

In conclusion, by decoupling the action choice from the target Q-value generation, the improved DQN can substantially reduce the overestimation, train faster and be more stable.

References

- Van Hasselt, H., Guez, A. and Silver, D., 2016, March. Deep Reinforcement Learning with Double Q-Learning. In AAAI (pp. 2094-2100).
- *icml.cc/2016/tutorials/deep_rl_tutorial.pdf*
- https://medium.com/@awjuliani/simple-reinforcement-learning-with-tensorflow-part-4-deep-q-networks-and-beyond-8438a3e2b8df.f0bkjj8u5