Improvements since Natural Deep Q-Learning

Qi Cao

Mar. 28 2018

DQN combines reinforcement learning and deep learning, inspiring plenty of follow-up researches. Deep Double Q-learning network, Prioritized replay and Dueling network are three of the most famous improvements since natural DQN.

- **Deep Double Q-Learning Network**

Although DQN algorithm combines Q-learning with a deep neural network, it still suffers from substantial overestimations in some games. Such overestimation occurs when we use max operation to choose and evaluate an action. Deep Double Q-Learning Network (DDQN) reduces the overestimations by decomposing the max operation in the target into action selection and action evaluation. According to DQN, there are two networks: online network and target network. The update to the target network stay unchanged for several steps and remains a periodic copy of the online network. Researchers propose to evaluate the greedy policy according to the online network, but using the target network to estimate its value. The empirical results show that DDQN improves over DQN both in value accuracy and policy quality.

- **Prioritized Experience Replay**

Experience replay lets online reinforcement learning agents remember and reuse experiences from the past transitions. Through prioritizing the transitions to be replayed, we could make experience replay more efficient and effective. The central component of prioritized replay is how to measure the importance of each transition. TD-error indicates how unexpected the transition is. The transition with the largest absolute TD error is replayed from the memory. Meanwhile, new transitions without TD-error would equipped with maximal priority. To overcome some issues like focusing on a small subset of the experience, researchers introduce a stochastic sampling method, which combines pure greedy prioritization and uniform random sampling. DQN with prioritized experience replay has better performance on most Atari experiments.

- **Dueling Network Architectures**

Dueling Network is equipped with two separate estimators: one for the state value function and the other for the state-dependent action advantage function. These two functions share a common convolutional feature learning module and the two streams are aggregated on a special aggregating layer to generate value function Q. Experiments show that dueling architecture could identify the correct action faster during policy evaluation compared with single-stream network, especially when the number of actions increase.

- **Spotlight Question**

Is it possible to combine some of above architectures to improve the performance of DQN?

- **Reading**

[1] Van Hasselt, H., Guez, A., & Silver, D. (2016, February). Deep Reinforcement Learning with Double Q-Learning. In AAAI (Vol. 16, pp. 2094-2100).

[2] Schaul, T., Quan, J., Antonoglou, I., & Silver, D. (2015). Prioritized experience replay. arXiv preprint arXiv:1511.05952.
[3] Wang, Z., Schaul, T., Hessel, M., Van Hasselt, H., Lanctot, M., & De Freitas, N. (2015). Dueling network architectures for deep reinforcement learning. arXiv preprint arXiv:1511.06581.