Research Review:

Human-level control through deep reinforcement learning

This paper published in Nature on 26 Feb 2015 by DeepMind. It introduced a ground breaking method of combining reinforcement learning and deep neural networks to form a novel agent, deep Q-network (DQN), which surpassed the performance of all previous reinforcement algorithms in a complex and high dimensional game environment. Like other deep learning methods, the goal of DQN is to find a set of actions (policy) which optimizes the action-value function (known as Q function which is a form of Bellman Equation Expression) by getting feedback through the interactions between agent and environment. DeepMnid successfully trained a DQN by using two key techniques. First, they used experience replay which randomizes over the data and removes correlations in the observation sequence and smooth over changes in the data distribution. Second, they used an iterative update that adjusts the action-values (Q) towards target values in another network that are only periodically updated, thereby reducing correlations with the target. They tested the agent in the challenging domain of classic Atari 2600 games. The inputs received by the agent were only scores and pixels and they used deep convolutional network to mimic the effects of receptive fields, which was inspired by Hubel and Wiesel's seminal work on feed forward processing in early visual cortex, in order to exploit the local spatial correlations present in images, and build in robustness to natural transformations such as changes of viewpoint or scale.

DeepMnid tested DQN with the best performing methods from the reinforcement learning literature on the 49 games where results were available, with same network architecture and hyper parameter values. In addition, they also compared the performance of a professional human tester and uniformly random agent. The DQN agent outperformed the best existing reinforcement learning methods on 43 of the games without incorporating any of the additional prior knowledge about Atari 2600 games used by other approaches. Furthermore, the DQN agent performed at a level that was comparable to that of a professional human games tester across the set of 49 games, achieving more than 75%of the human score on more than half of the games. By disabling experience replay, separate target Q-network and deep convolutional network, the performance of the agent deteriorated significantly and hence it demonstrated the significance of these three features in the proposed DQN architecture. The paper also points out that the games in which DQN excels were extremely varied in their nature, from side-scrolling shooters (River Raid) to boxing games (Boxing) and three-dimensional car-racing games (Enduro). Indeed, in certain games DQN was able to discover a relatively long-term strategy. For example, in Breakout, the agent learned the optimal strategy, which was to first dig a tunnel around the side of the wall allowing the ball to be sent around the back to destroy a large number of blocks. Nevertheless, games demanding more temporally extended planning strategies such as Montezuma's Revenge still constituted a major challenge for all existing agents including DQN.