Summary 10: Human Level Control Through Deep Reinforcement Learning

Susan Cherry

Reinforcement learning is difficult: in order for reinforcement learning to be successful, agents must be able to use high dimensional inputs to derive representations of the environment and then use these inputs to generalize from their past experiences. This paper uses advances in deep neural networks to develop a new artificial agent (which is called a deep Q-network) that is able to learn policies directly from high dimensional sensory inputs using end-to-end reinforcement learning.

The authors begin by presenting the deep-Q network (DQN) which combines reinforcement learning with deep neural networks. Specifically, they use deep convolutional network which uses hierarchical layers of tiled convolution filters to mimic receptive fields. Consider the tasks in which an agent interacts with an environment through sequences of observations actions, and rewards. The authors use deep convolution neural networks to approximate the optimal action-value function, which is the maximum sum of rewards discounted by γ at time t, achievable after by a behavior policy after making an observation and taking an action.

Next, the authors describe how the Q-learning addresses the instabilities that are typically present in reinforcement learning: They use experience replay to remove correlations in observation sequence and use an iterative update to adjust action values towards target values that are updated only periodically. This reduces the correlations with the target. Unlike other stable methods, the algorithm presented here is efficient enough to be used with large neural networks. The authors describe how to perform experience replay and apply Q-learning updates. They present the Q-learning update at iteration i, which uses the loss function presented at the bottom of the first page.

The authors evaluate the DQN using the Atari 2600 games that offer diverse and challenging tasks. They demonstrate that the DQN successfully learn policies and that they are able to train large neural networks using signal and stochastic gradient descent in a stable manner. The DQN outperformed existing reinforcement learning methods on 43 out of the 49 of the games and performed at the level of a professional human on all 49 games. They find that disabling the replay memory, separate target Q network, and deep convolution network architecture greatly reduce performance. Furthermore, they use a technique for the visualization of high dimensional data called 't-SNE' and find that the tSNE algorithm tends to map the DQN representation of states that are perpetually similar to nearby points. This provides evidence that the representations learned by the DQN allow it to predict state and action values accurately.

Overall, this paper shows the DQN can successfully learn policies in a wide range of diverse environments with only minimal prior information. This draws from the neurobiological evidence that reward signals during learning might influence the representations of characteristics in the visual cortex. As this model shows, combining machine learning techniques with biological mechanisms can results in effective algorithms that can learn and succeed at challenging and diverse tasks.