Paper Review: **Playing Atari with Deep Reinforcement Learning**
https://www.cs.toronto.edu/~vmnih/docs/dqn.pdf

This paper from “deep mind technologies” is seminal work to apply deep learning model to computer video games. The model is applied to seven Atari 2600 games from the Arcade Learning Environment, and outperforms all previous Reinforcement Learning algorithms on six of the seven games, and surpasses an expert human player on three of the seven games. Cited 4386 times so far, work in this paper opens gate for more innovative deep learning research.

Let computer learn how to play video game is not something new. Authors surveyed the research in this field. They found previous work on TD-gammon is a big success but doesn’t work well with games like Go and checker, which leads to widespread belief that TD-gammon is a special case due to the stochasticity of dice rolls in that game. Also majority of work in reinforcement learning focused on linear function approximators. The approach in the paper uses deep learning neural network instead and particularly applies the learning end-to-end, i.e. working directly with the visual inputs.

The learning is formalized as a tuple (**Atari emulator environment**, **legal actions of Agent**, **current sate observed by the input image**). The task is partially observed, and the sequences of actions and observations fit a finite Markov decision process (MDP). The goal of the agent is to interact with the emulator by selecting actions to maximize the reward, i.e. game score here, which is a optimization problem satisfying Bellman equation. To make computation practical, this paper uses a non-liner, i.e. neural network(referred as Q-network), function approximator to estimate the action-value function. This Q-network is trained by minimizing a sequence of loss functions which is done through stochastic gradient descendent. The author claims the algorithms has two advantage: 1) model-free. Input is the raw pixels without priori estimate of the environment, and 2) off-policy.

This paper also introduces the challenges for using deep learning with reinforcement learning: 1) In stead of large set of labels, reinforcement learning learns from a scalar reward signal, and 2) the data samples are usually highly correlated rather than independent. To solve the problems of correlated data, this paper uses an “**experience replay**” mechanism which randomly samples previous transitions, stores the agent’s experience at each time-step and pools over many episodes into a replay memory. This mechanism integrates Q-learning for updating samples of experience and the resulting algorithm is called “deep Q-learning” by the authors. This approach has several advantage over standard Q-learning: 1) Greater data efficiency as each step of experience is potentially used in many weight updates. 2) Reduced variance of the updates due to the randomizing, and 3) Smooth learning and avoidance of oscillating/divergent parameters due to the experience replay. The experiment preprocessed the input video images by first gray-scaling, down-sampling, and cropping as 84 by 84 regions to reduce computational complexity, and then fed such regions to a neural network with 2 hidden layers and one output layer which is a fully connected linear layer with a single output for each valid action, like moving left or right. The author showed that despite lacking theoretic convergence guarantee, the method is able to train large neural network in a stable manner.