

Rainbow DQN for Planning

Hakan Yilmaz, Maryna Kapitonova, Theresa Eimer
Machine Learning for Robotics Lab Course

DQN aims to train a neural net to approximate the value function of a given problem. We used the maze planning task to showcase various extensions of the DQN method that are collectively known as Rainbow DQN (deepmind paper citation here).

Methods

Double DQN

To avoid maximization bias for the target calculation, we use two nets: the agent chooses the action, the target net predicts its Q-value (Insert new target formula).

a series of time steps and is at the same time self-annealing as the net will learn to ignore the noise at some point. This will result in less overfitting and greater robustness.

Prioritized Experience Replay

To learn the most, we need to learn from the transitions with the most information. Assuming those are the ones with the biggest error, we used ranked prioritization to assign transitions with higher errors a higher probability of being sampled (Insert probability formula). To avoid a sampling bias, we scale the error according to the transition weights (Insert weight scaling formula).

Target network

The target network enables us to decouple the learning targets from out net. It's gradually updated towards to network (Insert factor).

Noisy nets

Introducing noise will force the network to learn to explore over

Multi-step learning

To evaluate each action better, we don't use the immediate reward but the reward accumulated after the action (Insert formula).

Results

One loss graph, one step graph and two or three sentences about performance?