Team Name: badmephisto

Project: Curriculum Learning on a Rubik's Cube

Project summary

Reinforcement Learning has been shown to be a promising approach to solving complex decision problems ranging from playing Go at superhuman levels to solving control problems in robotic manipulation [1, 2]. While there are many out-of-the box reinforcement learning algorithms (Q-learning, Value Iteration, Policy Learning [3]), we ask the question:

1. Can curriculum learning [4] improve reinforcement learning for decision-making problems?

We aim to study this question under the toy problem of solving/unscrambling a Rubik's Cube [5, 6]. We hope that our findings will provide broader insights to the design of reward functions for reinforcement learning.

What you will do

We intend to write our code from scratch, including the simulator, the reward function, training loop, and evaluation infrastructure. We will use PyTorch for our models.

For our baseline, we plan to follow conventional approaches and train an epsilon greedy Deep Q network with 12 output neurons (12 moves).

For our curriculum learning setup, we will train the model to solve the puzzle from increasing lengths of scrambles (1 scramble, 2 scrambles, …).

As stretch goals, we may plug in existing RL algorithms, redesign our reward function, or investigate how much our model can be compressed through distillation/pruning.

Resources / Related Work & Papers

[1] Mastering the game of Go with deep neural networks and tree search, Silver et. al.

[2] Human-level control through deep reinforcement learning, Mnih et. al.

[3] Proximal Policy Optimization Algorithms, Schulman et. al.

[4] Curriculum Learning for Reinforcement Learning Domains: A Framework and Survey, Narvekar et. al.

[5] Solving the Rubik's cube with deep reinforcement learning and search, Agostinelli et. al.

[6] Solving the Rubik's Cube Without Human Knowledge, McAleer et. al.

Datasets

We intend to design our own simulator to train and test our networks on. We have already finished implementing our simulator, so the remaining data work relates to implementing the reward functions.

List your Group members.

- David He
- Hanlong Li
- Henry Liao
- Sriharsha Kocherla