Leveraging Uncertainty-Aware Dynamics Models to Refine Actor-Critic Methods NY, USA
Advised by Dr. Hod Lipson at Columbia CS 01/2020 - ongoing

PROBLEM Model-based RL improves the sample complexity of its model-free counterparts at the cost of asymptotic performance. Imperfect models of continuous, high-dimensional dynamics introduce an approximate MDP under which even optimal policies are often useless in the real environment. Compounding errors in model prediction, also known as exposure bias, further create a bottleneck for performance of model-based RL algorithms.

ALGORITHM As in Soft Actor-Critic, the model-free training loop has an experience replay buffer filled with real data. We augment this training loop with an added planning step of Monte Carlo Tree Search (MCTS). I show fully Bayesian forward models with recurrent architectures are difficult to train, but those with stochastic latent embeddings, as in variational autoencoders, are more robust to compounding error. I introduce a novel learning curriculum to stabilize training over increasing horizons. For the first time, we use model uncertainty to penalize the rewards of states rolled out during MCTS to avoid trusting imagined trajectories that are unlikely to occur in the real environment.

RESULTS On a series of continuous control OpenAI gym tasks with varying difficulty, we consistently improve the performance of model-free RL even with imperfect models of the world.

PUBLICATION Robert Kwiatkowski, **Abhi Gupta**, Wonjun Sun, Boyuan Chen, and Hod Lipson. "Leveraging Uncertainty-Aware Dynamics Models to Refine Actor-Critic Methods." **In preparation for ICML 2021.**