Combining Reinforcement Learning and Diffusions via Weighted Maximum Likelihood Estimation

Description

Maximum a Posteriori Policy Optimization [1] formalizes Reinforcement Learning (RL) as a weighted maximum likelihood estimation (MLE) problem. So far, only Gaussian distributions have been considered for representing the policy. Motivated by recent success in applying diffusion models to MLE [4], the goal of this thesis is to replace the Gaussian policy with a diffusion model to obtain a more flexible model for policy representation.

Tasks

The tasks in this project will involve:

- Implementation. Getting familiar with the Tonic RL library [2] which has a working implementation of Maximum a Posteriori Policy Optimization. Integrating diffusion models into the Tonic RL library.
- Benchmarking. Benchmark the new method on RL task suites such as the Deepmind Control Suite [5].
- Algorithmic extensions. Identify weaknesses and develop extensions such as trust regions [3].

References


