Karlsruhe Institute of Technology

Autonome Lernende Roboter (ALR)

Prof. Gerhard Neumann

Project Type

- [x] Master Thesis
- [ ] Bachelor Thesis
- [ ] Research Project

Supervisors

Denis Blessing

denis.blessing@kit.edu

Onur Celik

celik@kit.edu

Philipp Becker

philipp.becker@kit.edu

Difficulty

Algorithmic

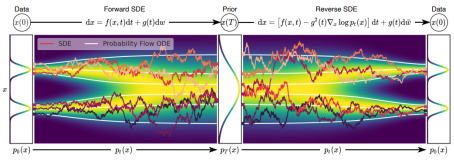
Math

Application

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

[1] Abbas Abdolmaleki, Jost Tobias Springenberg, Yuval Tassa, Remi Munos, Nicolas Heess, and Martin Riedmiller. Maximum a posteriori policy optimisation. *arXiv preprint arXiv:1806.06920*, 2018.

[2] Fabio Pardo. Tonic: A deep reinforcement learning library for fast prototyping and benchmarking. *arXiv preprint arXiv:2011.07537*, 2020.

[3] John Schulman, Sergey Levine, Pieter Abbeel, Michael Jordan, and Philipp Moritz. Trust region policy optimization. In *International conference on machine learning*, pages 1889–1897. PMLR, 2015.

[4] Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. *arXiv preprint arXiv:2011.13456*, 2020.

[5] Yuval Tassa, Yotam Doron, Alistair Muldal, Tom Erez, Yazhe Li, Diego de Las Casas, David Budden, Abbas Abdolmaleki, Josh Merel, Andrew Lefrancq, et al. Deepmind control suite. *arXiv preprint arXiv:1801.00690*, 2018.