Utilizing Simulation for Reinforcement Learning and Curiosity Driven Exploration in Robotics[⋆]

Hampus Åström
Volker Krüger
Elin Anna Topp

Department of Computer Science, Lund University, Lund, Sweden

Reinforcement learning (RL) is powerful when one does not have any, or sparsely labeled data to guide the learning of relevant behaviour. RL does, however, require a lot of training time, and for robotic training real operation for extended time is often impractical or impossible.

The natural alternative to physical run time is to do training in simulation. However, for many tasks simulation accuracy can be lacking both in terms of the robotic model and the model of its environment, and a system trained on a faulty model will not in general perform well in the real world. A solution to this problem is to train the robot in simulation while varying parameterization, domain randomization [2], to make the system learn a robotic behaviour that is robust to noise and systematic variations in the physical behaviour of both the robot itself and the environment it interacts with.

Further issues that are common to reinforcement learning are producing a good reward signal and efficiently exploring the state space. In constrained physical or simulated environments these problems could be alleviated by using curiosity driven exploration for a robotic agent [3], in order to let it find its own capabilities and the properties of the environment on its own. This could be a productive way to develop basic robotic skills, without an explicit reward signal, that could then be applied to more complex tasks.

The curiosity driven approach could be complemented by or combined with reinforcement learning for open ended tasks like "build as high as possible with the objects you can reach". The hypothesis is that such tasks always give some reward, and even simple things like tearing down or constructing a pile will give some reward related feedback from which the system can learn and improve its behaviour.

Initial plan of action

In applying the principle of domain randomization, as found in [2], there are several important decisions to be made. Firstly a task and robot needs to be selected, and the proposal is to start by using a ABB YuMi or KUKA iiwa to solve block building tasks. Secondly a simulation environment that allows for a high level of domain randomization needs to be selected. This is yet to be

[⋆] This work was partially supported by the Wallenberg AI, Autonomous Systems and Software Program (WASP) funded by Knut and Alice Wallenberg Foundation.

determined, but MuJoCo [4] or Gazebo [1] are two options under consideration. Thirdly the input and output space for the learning algorithm has many options. Initially the aim will be to use higher abstraction for the learning input space and a fairly low level action space, similar to [2]. The system will also need some kind of vision system to provide these abstract inputs once it runs outside of simulation, and for that task RGBD video is the primary option.

Once this system can perform in some capacity it might be possible to expand the investigations to curiosity driven exploration in more advanced tasks, as well as investigating how the abstraction level for the input and action space influence outcome.

References

1. Koenig, N., Howard, A.: Design and use paradigms for gazebo, an open-source multi-robot simulator. In: 2004 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) (IEEE Cat. No.04CH37566). vol. 3, pp. 2149–2154 vol.3 (Sep 2004). https://doi.org/10.1109/IROS.2004.1389727
2. OpenAI, Andrychowicz, M., Baker, B., Chociej, M., Józefowicz, R., McGrew, B., Pachocki, J.W., Pachocki, J., Petron, A., Plappert, M., Powell, G., Ray, A., Schneider, J., Sidor, S., Tobin, J., Welinder, P., Weng, L., Zaremba, W.: Learning dexterous in-hand manipulation. CoRR **abs/1808.00177** (2018), http://arxiv.org/abs/1808.00177
3. Pathak, D., Agrawal, P., Efros, A.A., Darrell, T.: Curiosity-driven exploration by self-supervised prediction. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). pp. 488–489 (July 2017). https://doi.org/10.1109/CVPRW.2017.70
4. Todorov, E., Erez, T., Tassa, Y.: Mujoco: A physics engine for model-based control. In: 2012 IEEE/RSJ International Conference on Intelligent Robots and Systems. pp. 5026–5033 (Oct 2012). https://doi.org/10.1109/IROS.2012.6386109