[12.1.2021]

[Policy Transfer with Strategy Optimization]

Summary

This paper proposed to train a family of potential policies by augmenting the policy input to also include dynamic parameter. When transferring to the target environment, it uses evolutionary strategy to pick the best polices → strategy.

Some key points of this paper:

- Not implement on the real robot → maybe one possible solution.

Major Analysis and Comparison

- Two main approaches in order to cross the sim to real gap:
    - Fidelity improvement
    - Robust → a family of policies,among all dynamics, it will implicitly/explicitly select one and output the best action
        - A similar idea is to train an adaptive policy with the current and the past observations as input
            - ANYmal sci learning paper is not the same, it includes historical data as observation.
            - I do not understand yet, what does training an adaptive policy mean here? Is the input here the same as action?
- To answer the previous question, the policy here is not o-→a, instead, the dynamic parameter is also included: o,\mu → a . So Every time at the beginning of a rollout, the dynamics parameter is picked randomly and then we train the policy
    - when you need to pick the best policy that suits for the target environment, need to define a evaluation matrix to find the best policy

$$\mu^* = \arg\max_{\mu} J_{\mathcal{M}^t}(\pi_\mu).$$

    - problem is we still need samples from the target environment to pick the best policy –> strategy.