Complex Locomotion Generation with Diverse Skill Learning

JIN NAKAZAWA LABORATORY

d-hacks
B3 mioto
Mentor: B4 sean

Background

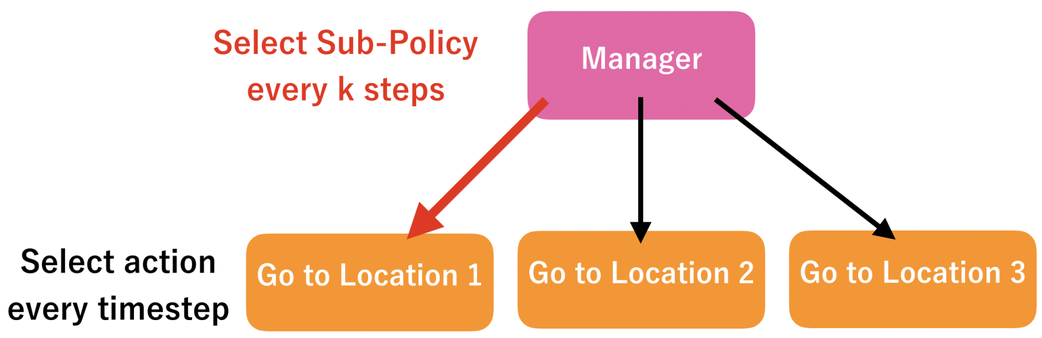
Robotic control is an important application of **reinforcement learning** algorithms because of its adaptability to state changes in the environment. Such systems will be applied to situations with bumpy landscapes such as the aftermath of a natural disaster.

Problem

There are difficulties in learning robotic control that traverse through landscapes with complex details such as stair-like bumps. Learning such tasks can be challenging because of the vast number of possible action sequences.

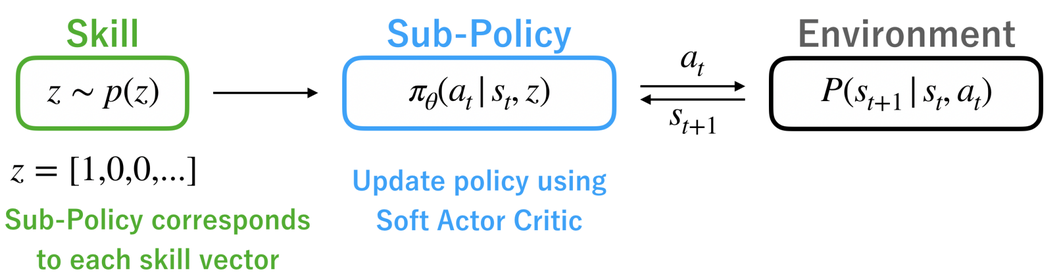
Approach

We believe that an effective approach is to use **hierarchical reinforcement learning**, where complex tasks are reinterpreted as sequences of smaller simpler tasks, each depicted by sub-policies. An additional high level policy reorders the sub-policies to depict a complex task. For the generation of diverse complex locomotion, a large number of diverse sub-policies must be acquired.



Preliminary Work

Diversity Is All You Need (DIAYN) is an algorithm for acquiring diverse skills by optimizing its policy, which inputs both environment states and a **skill vector**, a one-hot vector sampled from a **categorical distribution**, to output actions that correspond to the vector. The policy is updated using Soft Actor Critic.



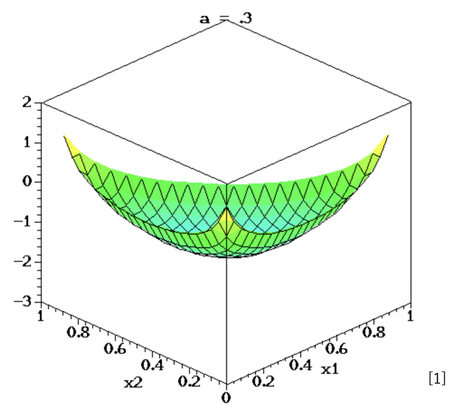
Although DIAYN is shown to be able to perform well on complex landscapes, a problem lies in the distribution the skills are sampled from, which limits the number of acquireable skills to the dimension of the vector set beforehand.

Dirichlet-DIAYN

Our proposal is to use **Dirichlet distributions** for representing **continuous skill vectors**. The shape of a Dirichlet distribution can be set to resemble different distributions by altering its concentration α. Using this property, by setting the initial α to 0, resembling a categorical distribution, and gradually increasing it to 1, resembling a continuous uniform distribution, during training, the policy can adapt to learning with continuous skill.

$$\alpha(t) = \min(\gamma + (1-\gamma)\frac{t}{\tau}, 1)$$

$$z \sim Dirichlet(\alpha(t))$$



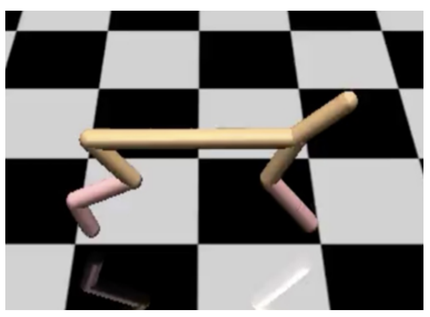
[1]

Our Previous Work

We implemented DIAYN to **HalfCheetah**, a virtual environment with a 2 dimensional cheetah-type robot on an infinite flat plane.

Future Experiments

We will evaluate Dirichlet-DIAYN and DIAYN by applying them to hierarchical reinforcement learning, where the manager outputs a skill every k steps. Each agent will walk across bumpy landscapes, and their performances will be tested and compared.



Eysenbach, Benjamin, et al. "Diversity is all you need: Learning skills without a reward function." arXiv preprint arXiv:1802.06070 (2018).

Haarnoja, Tuomas, et al. "Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor." arXiv preprint arXiv:1801.01290 (2018).