Sub-policy pruning in Meta Learning Shared Hierarchies

Qing HONG[1, 2*] Yusuke TANIMURA[2,1] Hidemoto NAKADA[2,1]
[1]University of Tsukuba
[2]National Institute of Advanced Industrial Science and Technology

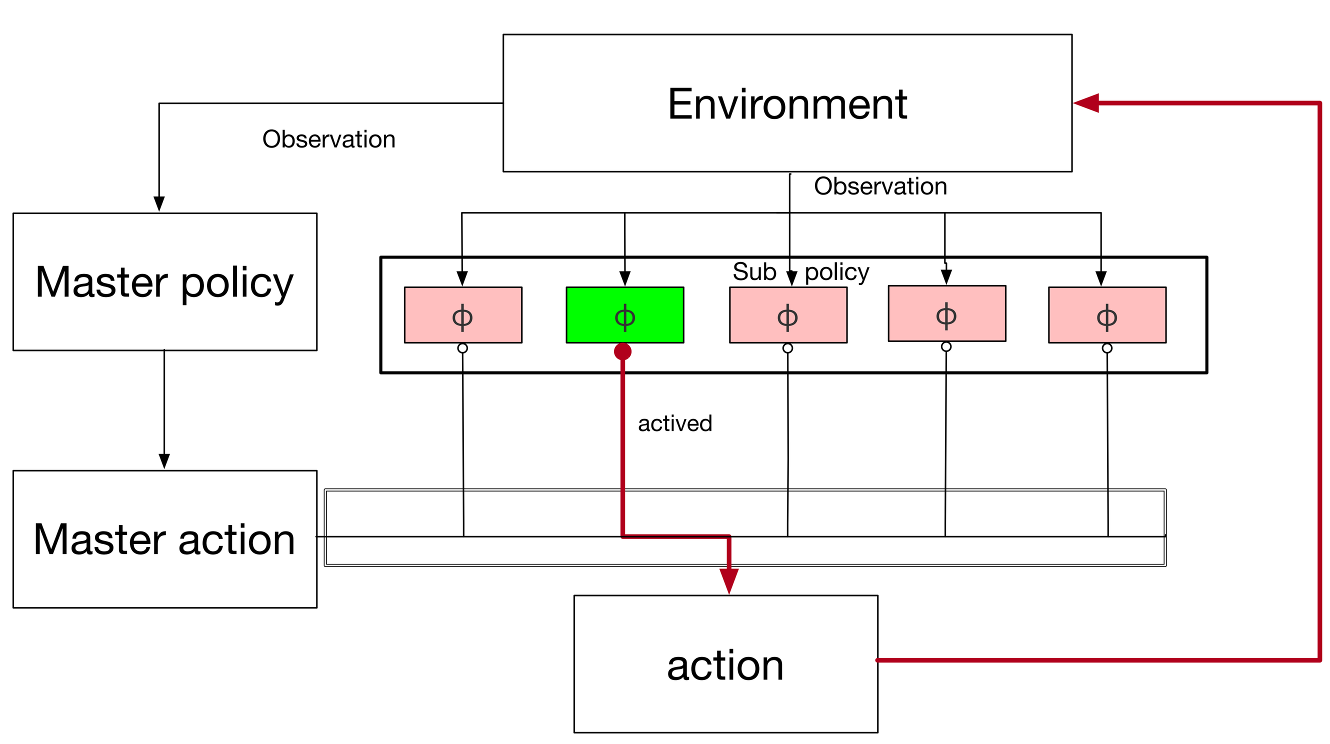
Introduction

In Reinforcement learning, It is a big challenge to quickly reach high reward within a distribution of tasks. (e.g. a set of similar rule tasks can share information between each other)
MLSH is a meta learning method to solve such problems. But it is not suitable for complicated environments and it is very difficult to train them properly. We propose a method to effectively prune excessive sub-policies to give better chance the other sub-policies to be trained.

MLSH

MLSH can quickly reach high reward with the help of pre-trained sub-policies.

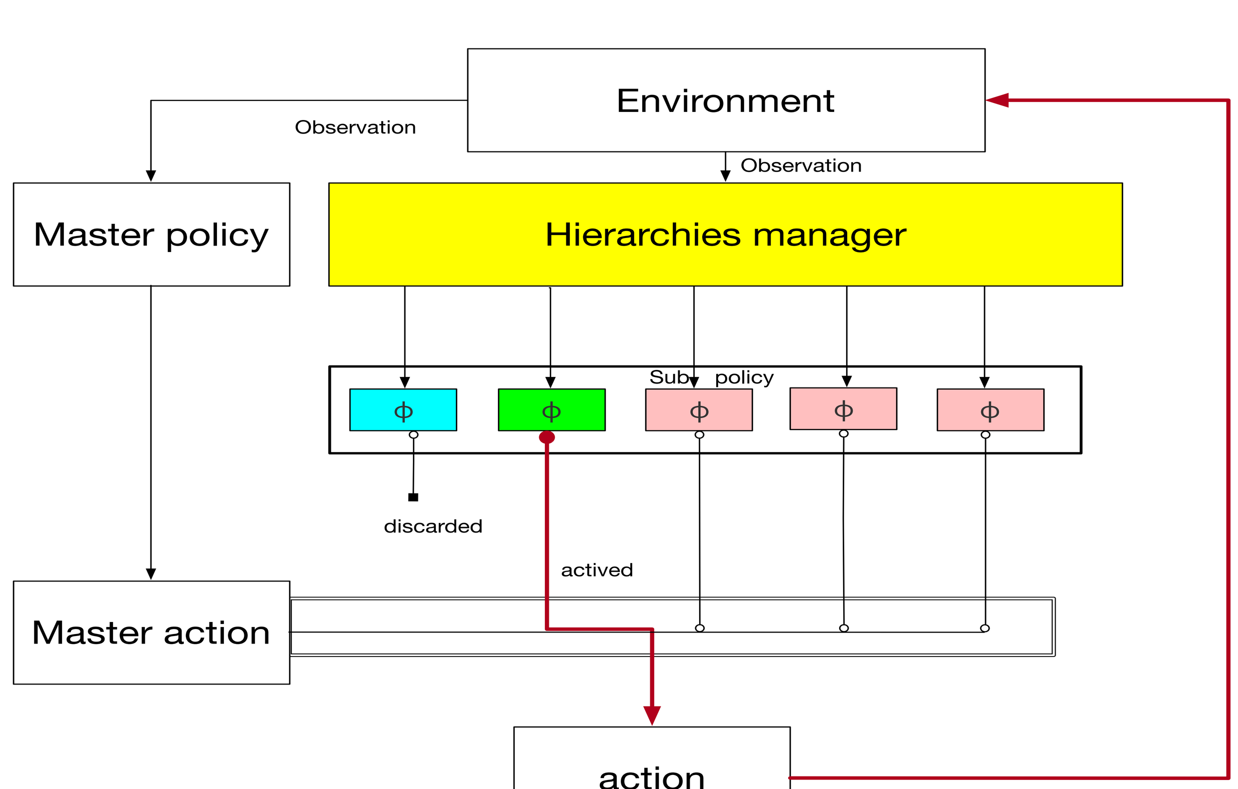
1. 'Hierarchical' structure with one master policy and multiple sub-policies.
2. Sub-policies responsible for specific situations.
3. Master policy to 'select' one of the sub-policies.



Pruning Method

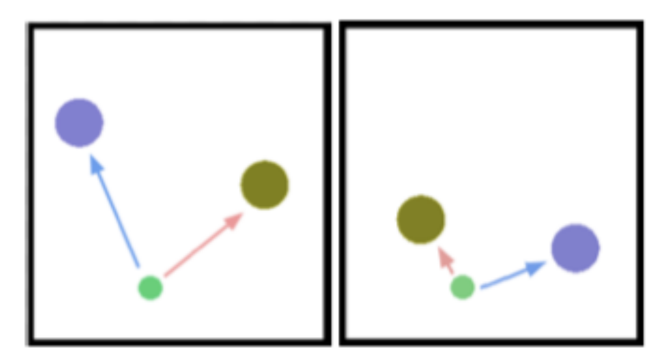
Purpose: To find a proper number of sub-policies.

- Enough number of sub-policies.
- Prune 'excessive' policies.
- The hierarchies manager is responsible for pruning.

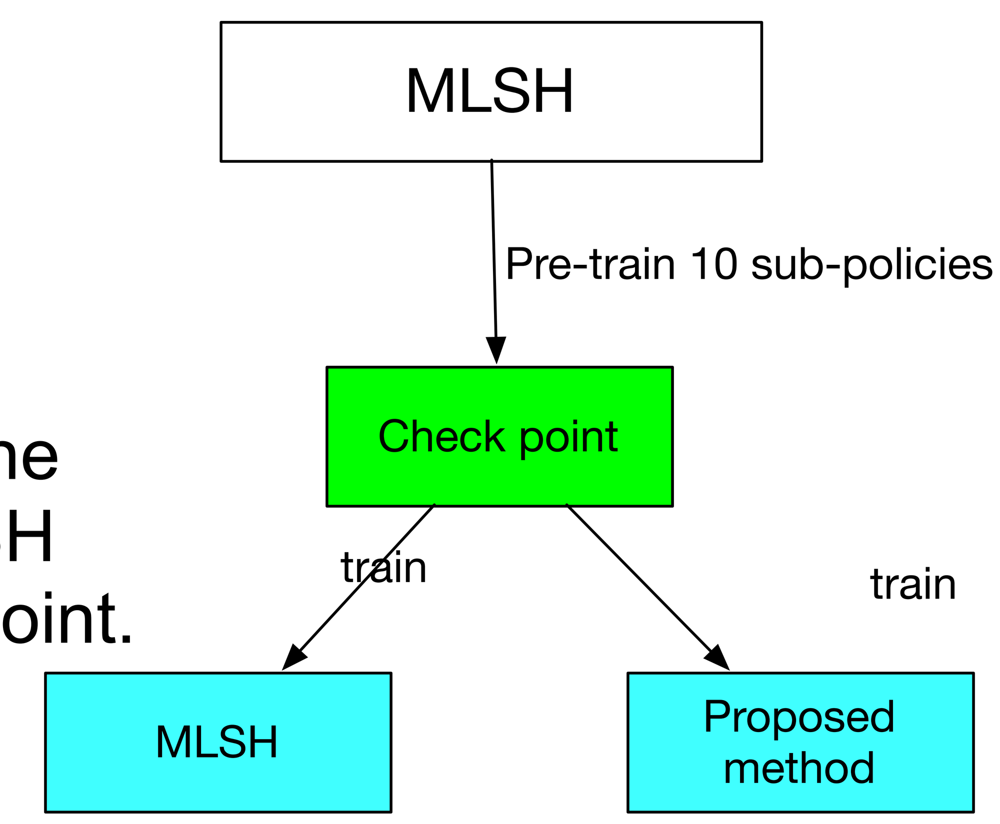


Experiment

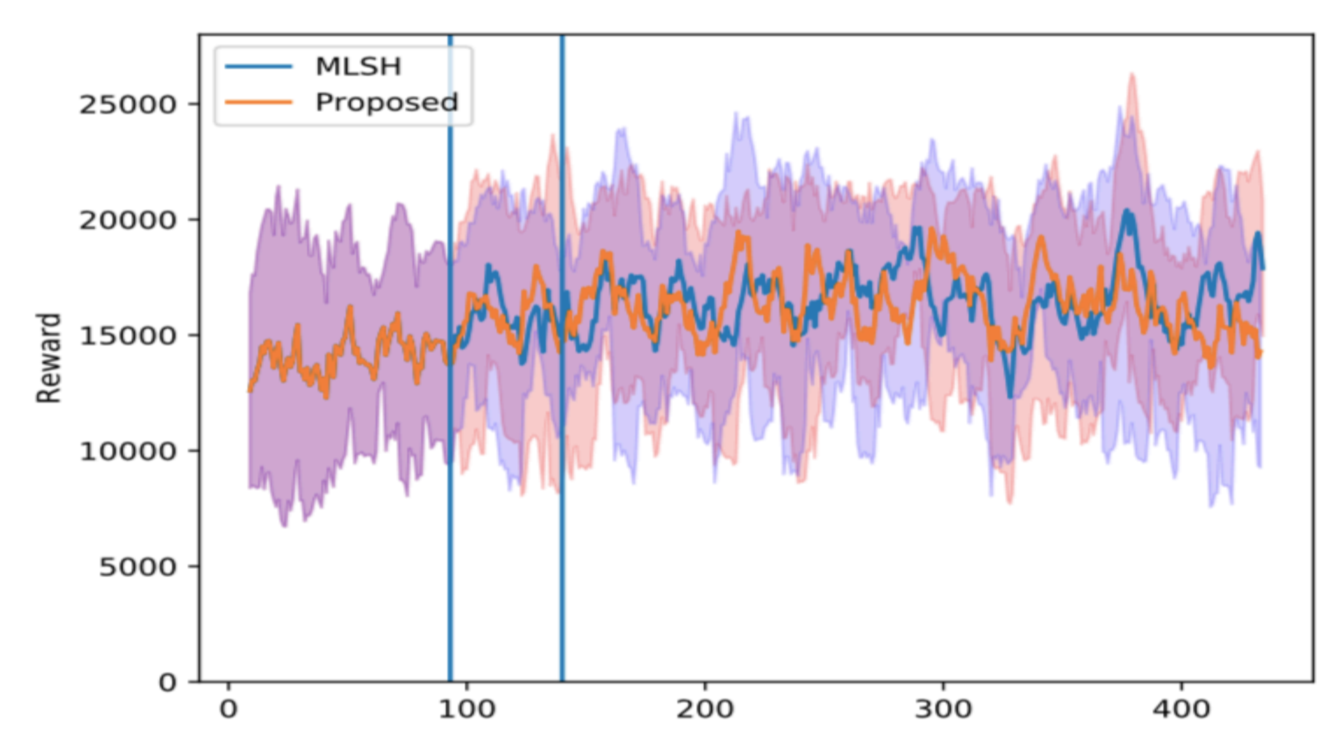
Simple 2D-moving bandits:
The agent(small dot)'s goal is to get as close as possible to the other dots.



- Checkpoint: Pre-train the sub-policies using MLSH and save same checkpoint.
- A pair of experiments



Result



The vertical line denotes the time point we prune the policies. We prune 30% and 10% sub-policies in this two time point.

Future Work

- Detect the sub-policies which always perform identically.
- Performs further pruning in order to reach the minimum number of sub-policies.

References

[1] Frans, K. , et. al, Meta learning shared hierarchies. CoRR, abs/1710.09767, 2017.
[2] Bacon,P.L., et. al, The option-critic architecture. CoRR, abs/1609.05140, 2016.
[3] Mnih,V., et. al, Playing atari with deep reinforcement learning. CoRR, abs/1312.5602, 2013.
[4] Kulkarni,T.D., et. al, Hierarchical deep reinforcement learning: Integrating tempo- ral abstraction and intrinsic motivation. CoRR, abs/1604.06057, 2016.

Acknowledgement

- ***This poster is based on results obtained from a project commissioned by the New Energy and Industrial Technology Development Organization (NEDO).***
- ***This work was supported by JSPS KAKENHI Grant Number JP16K00116.***