Multi-task Reinforcement Learning with Soft Modularization

Ruihan Yang, Huazhe Xu, Yi Wu, Xiaolong Wang

project page: https://rchalyang.github.io/SoftModule
Motivation

Most RL: specialized

How?   Multi-task RL

Goal: Apply to real-world

Picture source: https://robotage.guru/robot-chef-robotic-kitchen/
Current Multi-task RL Benchmark: Meta-World

Containing dozens of robotics manipulation tasks.

Yu, T et al. Meta-World: A Benchmark and Evaluation for Multi-Task and Meta Reinforcement Learning, 2019
Meta-World

Success Rate

MT10

MT50

Over 60%

Multi-Task SAC  Multi-Task Multi-Head SAC  Single-Task SAC

Haarnoja, T et al. Soft Actor-Critic: Off-Policy Maximum Entropy Deep Reinforcement Learning with a Stochastic Actor, 2018
Failed Examples

Baseline failed to generalize to different tasks

Open Door  Close Drawer  Insert Peg
General Multi-Task Learning

Computer Vision: Detection + Segmentation

Robot Learning: Pushing + Grasping + Poking > Grasping

He, K et al. Mask R-CNN, 2017
Pinto, L et al. Learning to Push by Grasping: Using multiple tasks for effective Learning, 2017
Two Challenges in Multi-Task Reinforcement Learning

Avoid negatively interference between irrelevant tasks

Reuse shared components across similar tasks

Reach
Open Door
Open Window
Modularization

Previous Modular network for multi-task RL

In hierarchical manner

Andreas, J., Klein, D., and Levine, S. Modular multi-task reinforcement learning with policy sketches, 2017
Soft Modularization

Base Policy Network + Routing Network

Differentiable
Soft Modularization

Base Policy Network

Routing Network

Network Module

Dot Product

\( s_t \)

\( z_T \)
Soft Modularization

Base Policy Network

Routing Network

Network Module

Dot Product

Base Policy Network

Routing Network

Network Module

Dot Product

S_t

Z_T

Reweight with Probabilities
Soft Modularization

Reweight with Probabilities

Base Policy Network

Routing Network

Network Module

Dot Product

Weighted Sum as Input

\[ s_t \]

\[ z_T \]
Soft Modularization

Base Policy Network

Routing Network

Network Module

Dot Product

Reweight with Probabilities

$S_t$

$Z_T$
Soft Modularization

Base Policy Network

Routing Network

Reweight with Probabilities

Network Module

Dot Product

$S_t$

$Z_T$
Soft Modularization

- Base Policy Network
- Routing Network
- Weighted Sum as Input

Network Module
Dot Product

\( s_t \) \( z_T \)

Reweight with Probabilities
Soft Modularization

Base Policy Network

Routing Network

Network Module

Dot Product

Reweight with Probabilities
Balance Multi-task learning in SAC

Temperature weight for multi-task RL:

\[ J_\pi(\phi) = \mathbb{E}_{T \sim p(T)}[w_T \cdot J_{\pi,T}(\phi)] \]  

(1)

Temperature adjustment in SAC:

\[ J(\alpha) = E_{a_t \sim \pi_{\phi}}[-\alpha \log \pi_{\phi}(a_t|s_t) - \alpha \overline{H}] \]

(2)

Temperature weight for multi-task RL:

\[ w_i = \frac{\exp(-\alpha_i)}{\sum_{j=1}^{M} \exp(-\alpha_j)} \]

(3)
Soft Modularization
Module Sharing: Knowledge Sharing.

Push

Close Window
Module Sharing : Knowledge Sharing.
Module Sharing: Knowledge Sharing.
Experiments

Original Meta-World:
   Fixed goal.
   MT10-Fixed / MT50-Fixed

More realistic and more challenging:
   Goal conditioned
   MT10-Conditioned / MT50-Conditioned
Experiments: Over All

MT10

- Fixed: Ours (80.00%) vs. Multi-Task Multi-Head SAC (40.00%)
- Conditioned: Ours (60.00%) vs. Multi-Task Multi-Head SAC (30.00%)

MT50

- Fixed: Ours (70.00%) vs. Multi-Task Multi-Head SAC (30.00%)
- Conditioned: Ours (60.00%) vs. Multi-Task Multi-Head SAC (30.00%)
Experiments : MT10

MT10-Fixed

MT10-Conditioned

- Ours
- Multitask SAC
- Multihead Multitask SAC
Experiments: MT50

MT50-Fixed

MT50-Conditioned
Effects on Network Capacity

Performed on MT50-Fixed

Larger network can not solve multi-task RL.
Effects on Network Capacity

**Success Rate**
- MT-50-Fixed

**Params**
- Ours
- Multi-Task Multi-Head SAC-4
- Multi-Task Multi-Head SAC-4-Wide
- Multi-Task Multi-Head SAC-6
Comparison with Single Task Policy

Open Door

Open Window

- Ours
- Multi-task Multi-head SAC
- Single Task SAC
Comparison with Single Task Policy

- Ours
- Multi-Task Multi-Head SAC
- Single-task SAC
Routing Visualization

Different Task: Separated

Similar Task: Closer
Routing Visualization

Different Task:
Separated

Similar Task:
Closer
Thanks!

Our project page: https://rchalyang.github.io/SoftModule