Task-relevant State-focused Unsupervised Skill Discovery

Hyeyun Kim, Junyong Seo, Shinsuk Park*
Korea Univ., Korea Univ., *Korea Univ.

{khyoon714, junyong9939, *drsspark}@korea.ac.kr

Abstract

There have been many studies on unsupervised skill discovery that enables learning diverse movements and alleviates the time-consuming manual tuning of reward functions in reinforcement learning. Conventional skill discovery methods based on mutual information (MI) aim to maximize the information transmission between entire states and skills. With this approach, it can be challenging for agents to learn skills that involve specific state manipulation during robotics tasks. To overcome the limitation, this study proposes an objective function with a MI term that takes account of task-relevant states so that agents can learn to manipulate the target states. The proposed method was evaluated in a loco-manipulation task using the Mujoco simulation. The results show that the agent trained by the proposed method can explore more states for a given robotic task, compared to conventional unsupervised skill discovery methods.

Ⅰ. Introduction

Reinforcement learning (RL) is considered similar to the learning methods of other intelligent creatures and is widely used in various machine learning fields. Recently, deep RL, which combines RL and Deep neural networks, has been applied to the robotic control field [1] and has demonstrated its effectiveness. However, RL requires substantial time for manual supervision to define and adjust reward functions for desired behavior. Additionally, RL's limitation to learning behaviors that satisfy specific conditions in reward functions presents a challenge for robots to acquire a diverse range of behaviors for robotic control.

In contrast to the aforementioned drawbacks, actual intelligent creatures learn various useful skills without supervision and sometimes apply previously acquired skills when learning new ones. Considering these characteristics, several unsupervised skill discovery approaches have been proposed [2]. These methods help reduce the need for manually defining and adjusting rewards and can provide combinable actions for subsequent operations.

One widely used method in unsupervised skill discovery maximizes the Mutual information (MI) between skills and states. However, applying previous MI-based methods to robotic control has limitations, especially in tasks requiring specific state transitions, such as adjusting object positions in manipulation tasks. Prior MI-based unsupervised skill discovery methods struggle to induce transitions for specific states because agents learn skills while giving equal consideration to both task-relevant states and non-relevant states. Therefore, emphasizing specific state transitions is crucial for effectively applying MI-based skill discovery to robotic control tasks.

To overcome the limitations of existing general MI-based approaches, we propose a method called Task-relevant State-focused Unsupervised Skill Discovery. This method aims to discover skills that involve specific state transitions, which is challenging in conventional MI-based skill discovery methods. This makes it easier to use MI-based approaches in robotic control for tasks that require specific state transitions. By taking into account MI related to the desired states, we enable the robot to learn task-relevant skills more effectively.

This paper has two main contributions. First, our method helps agents learn skills related to specific task-relevant states while also considering other states. This makes MI-based approaches effective for robotic control tasks. Second, our method showed higher state space coverage in targeted states compared to previous unsupervised skill discovery methods for a loco-manipulation task in the Mujoco simulation environment.

Ⅱ. Method

A. Mutual Information

MI-based unsupervised skill discovery methods learn skills in a manner that maximizes the MI between latent skill vector Z and state S. The MI equation can be expressed as follows:

$$I(Z;S) = H(Z) + H(S) - H(Z,S) \quad (1)$$
$$= H(S) - H(S|Z) \quad (2)$$
$$= H(Z) - H(Z|S) \quad (3)$$

$I(S;Z)$ represents the MI between Z and S, and $H(\cdot)$ represents the Shannon entropy. Equations (2) and (3) correspond to the forward-MI approach and the reverse-MI approach, respectively. The reverse-MI approach is used in this paper.

B. Task-relevant State-focused Unsupervised SD

Now describing task-relevant state-focused unsupervised skill discovery, which aims to discover skills focusing on specific states while still considering the overall states. The objective function $\mathcal{J}$ is proposed as follows:

$$\mathcal{J} = I(S;Z) + \alpha I(S_{focus};Z) \quad (4)$$
$$= H(Z) - H(Z|S) + \alpha\left[H(Z) - H(Z|S_{focus})\right] \quad (5)$$

In Equation (4), the first term, $I(S;Z)$, is the MI between S and Z, and calculates the dependency between two variables, S and Z, as in previous MI-based unsupervised skill discovery methods. $I(S_{focus};Z)$ represents the MI between S_{focus} and Z, where S_{focus} is the user-selected task-relevant states intended to be highly related to learned skills. The relative weight of $I(S_{focus};Z)$ to $I(S;Z)$ is adjusted through the hyperparameter α.

The weighted term, $I(S_{focus};Z)$, is incorporated into the objective function $I(S;Z)$ of conventional MI-based skill discovery methods. This approach not only considers all states like the previous methods but also focuses on S_{focus} while learning skills. Therefore, it is possible to learn skills that focus on controlling S_{focus}, making it easier to perform robotic tasks where specific state manipulation is required.

Ⅲ. Experiments

The proposed method was assessed via a loco-manipulation task in the Mujoco simulation environment. The goal is to train the agent to acquire skills to move the object to various positions. The object position was chosen as the S_{focus}, the task-relevant target states in our objective function.

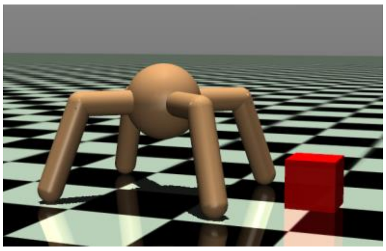

Figure 1. The simulation was performed in the Mujoco Ant environment[5] with a 10kg, 15cm cubic object of uniform density.

CSD[4] was used to approximate the MI terms in Equation (4). Mapping functions and distance functions in [4] were trained with respect to the first and second MI terms of Equation (4), respectively. The hyperparameters can be found in Table 1. Values not listed in the table are equal to those in the manipulation environment described in [4]. The performance of skills learned using the objective function in Equation (4) was compared to those learned using the conventional objective function without $I(S_{focus};Z)$.

Hyperparameter	Value
α	2.5
$Intrinsic\ reward\ coefficient$	500
$Discount\ factor$	0.99
$Episode\ length$	100

Table 1. Hyperparameters for simulation.

Figure 2. shows the experimental results with and without $I(S_{focus};Z)$. For quantitative comparison, state coverage was used as a metric, defined as the number of 0.05×0.05 and 0.1×0.1 sized regions in the state space visited at least once by the object and robot CoM position, respectively. For both with and without $I(S_{focus};Z)$, the agent state coverage converged to similar values. However, the converged object state coverage was greater with $I(S_{focus};Z)$. This indicates that our method helped the agent to discover skills specifically focused on the object position, S_{focus}, while also learning skills related to states other than S_{focus}.

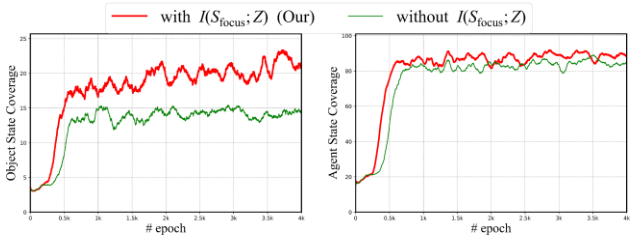

Figure 2. Comparison of object (left) and agent (right) state coverage in the loco-manipulation task with and without the $\boldsymbol{I(S_{focus};Z)}$.

Ⅳ. Conclusion

In this paper, we presented a novel MI-based unsupervised skill discovery method that learns skills by focusing on specific state changes. The proposed objective function includes a weighted MI term for focused states. This aims to address the limitation of previous MI-based methods, where skills learned with equal weights for all states can be restricted for use in manipulating specific states. Through the Mujoco loco-manipulation task, the proposed method outperformed previous MI-based methods in object state coverage by focusing on task-relevant object positions. In the future, we plan to apply our method to real robot tasks.

ACKNOWLEDGMENT

This work was supported by Korea Institute for Advancement of Technology (KIAT) grant funded by the Korea Government (MOTIE) (P0020535, The Competency Development Program for Industry Specialist) and the National Research Foundation of Korea(NRF) Grant funded by the Korean Government(MSIT) (RS-2023-00208052)

References

[1] S. Gu, E. Holly, T. Lillicrap, and S. Levine, "Deep reinforcement learning for robotic manipulation with asynchronous off-policy updates," IEEE Xplore, May 01, 2017.

[2] K. Gregor, Danilo Jimenez Rezende, and Daan Wierstra, "Variational Intrinsic Control," *OpenReview*, Feb. 17, 2017.

[3] B. Eysenbach, J. Ibarz, A. Gupta, and S. Levine, "Diversity is all you need: Learning skills without a reward function," in 7th International Conference on Learning Representations, ICLR 2019, 2019.

[4] S. Park, K. Lee, Y. Lee, and P. Abbeel, "Controllability-aware unsupervised skill discovery," in Proceedings of the 40th International Conference on Machine Learning, 2023, pp. 27 225–27 245.

[5] E. Todorov, T. Erez, and Y. Tassa, "MuJoCo: A physics engine for model-based control," *IEEE Xplore*, Oct. 01, 2012.