Motivation

Inspired by the human intelligence that can explore the environment and learn various skills to accomplish tasks, multi-skill DRL has been proposed as a potential solution. Our paper is focus on how to extract more effective information from these skills, and to accelerate the efficiency of skill-transfer.

Strong correlation

Figure 1 shows the eigenvalue and the corresponding percentage of principal components in primitive skills. It is observed that the actions generated by distinct primitive skills are strong correlated with each other, which indicates that the amount of skills can be reduced by eliminating this correlation so that the skill dimension is thus decreased.

Unbalance of skill discovery and transfer

In the skill discovery, the agent learns each of primitive skills separately in a source environment. However, in the skill transfer, all primitive skills are combined to instruct the agent in a target environment. As noticed, the primitive skill in skill discovery and skill transfer play a distinct role. A more balanced scheme might improve the performance of skill transfer.

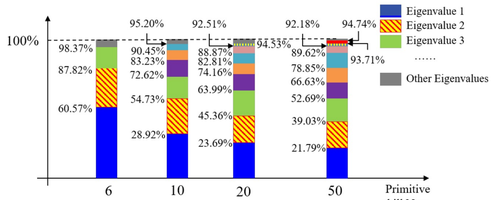


Figure 1: PCA of primitive skills.

Model

Framework

In Figure 2, diverse primitives skills are firstly cultivated in a source environment. By taking actions to represent such primitive skills, the agent then decomposes these primitive skills to acquire independent skills, where independent component analysis (ICA) is employed on those primitive skill's actions. Finally, the agent transfers independent skills into new practical skills in a target environment.

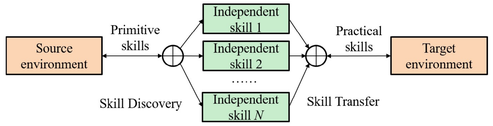


Figure 2: Framework of Independent Skill Transfer (IST).

Learn Independent Skills (LIS)

As shown in Figure 3, we first sample subset $\tilde{S}$ from the observation space S. Secondly, we sample action A_Z based on primitive skill π. We then convert the primitive actions A_Z to independent actions $\hat{A}$ through ICA. Finally, we utilize $\tilde{S}$ and $\hat{A}$ to learn independent skills $\hat{\pi}$ via supervised learning.

- Collection of Observation and Action

(1) $\tau_{z,i} = \{\boldsymbol{s}_{z,i,1}, \boldsymbol{a}_{z,i,1}, \boldsymbol{s}_{z,i,2}, \boldsymbol{a}_{z,i,2}, ..., \boldsymbol{s}_{z,i,T_i}\}, 1 \le i \le L, \pi(\boldsymbol{a}|\boldsymbol{s}, z)$

(2) $\tilde{S}_z = \{\boldsymbol{s}_{z,i,j} | 1 \le i \le L, 1 \le j \le T_i\}$

(3) $\tilde{S} = \tilde{S}_1 \cup \tilde{S}_2 \cup ... \cup \tilde{S}_{|\mathcal{Z}|}$

(4) $\boldsymbol{A}_z = [\boldsymbol{a}_1, \boldsymbol{a}_2, ..., \boldsymbol{a}_K \sim \pi(\boldsymbol{a}|\boldsymbol{s}, z) | \boldsymbol{s} \in \tilde{S}]$

(5) $A_z = \text{flatten}(\boldsymbol{A}_z)$

- Generation of Independent actions

$$\begin{bmatrix} \hat{A}_1^T \\ \hat{A}_2^T \\ \vdots \\ \hat{A}_{|\mathcal{Z}|}^T \end{bmatrix} = \boldsymbol{W}_I \boldsymbol{W}_P \begin{bmatrix} A_1^T \\ A_2^T \\ \vdots \\ A_{|\mathcal{Z}|}^T \end{bmatrix}$$

- Generation of Independent Skills

$$\hat{\boldsymbol{A}}_{\hat{z}} = [\hat{\boldsymbol{a}}_1, \hat{\boldsymbol{a}}_2, ..., \hat{\boldsymbol{a}}_K \sim \hat{\pi}_\theta(\hat{\boldsymbol{a}}|\boldsymbol{s}, \hat{z}) | \boldsymbol{s} \in \tilde{S}]; \ \min_\theta \{D_{KL}[\hat{\pi}_\theta(\hat{\boldsymbol{a}}|\boldsymbol{s}, \hat{z}) || \hat{p}(\hat{\boldsymbol{a}}|\boldsymbol{\mu}_{\boldsymbol{s},\hat{z}}, \boldsymbol{\sigma}_{\boldsymbol{s},\hat{z}})]\}$$

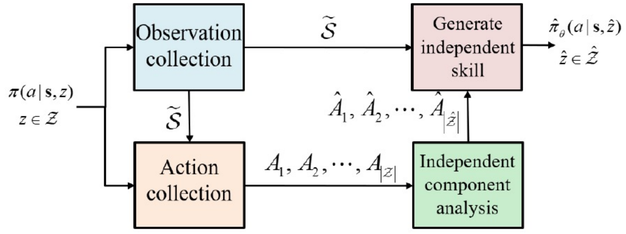


Figure 3: Framework of Learning Independent Skills (LIS).

Independent Skill Transfer (IST)

Since primitive actions are linear combination of independent actions, the transfer policy produces a weight and bias in each time step t based on observation s_t , so the composite action can be calculated by

$$\boldsymbol{a}_t = \hat{\boldsymbol{a}}_1 \alpha_{t,1} + \hat{\boldsymbol{a}}_2 \alpha_{t,2} + ... + \hat{\boldsymbol{a}}_{|\hat{\mathcal{Z}}|} \alpha_{t,|\hat{\mathcal{Z}}|} + \mathbf{1} \otimes b_t$$

After executing the composite action a_t, the agent can achieve the observation s_{t+1} and reward r from the target environment. Then, we employ SAC to update the transfer policy as shown in Figure 4.

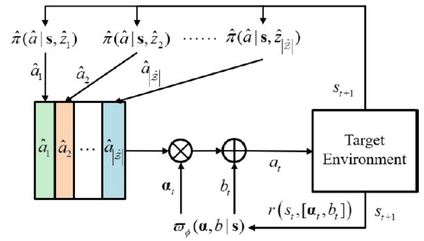


Figure 4: Process of Independent Skill Transfer (IST).

Experiment

According to the source environment HalfCheetah-v3, we set up target environments by considering extra key elements as shown in Figure 5, including HalfCheetah-Hurdle (HCH), HalfCheetah-Ascending (HCA) and HalfCheetah-Upstair (HCU).

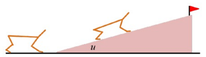
Baselines

- Primitive skill transfer PST [1]
the composite action is thought of as the linear combination of primitive actions
- Primitive skill selection (PSS) [2]
a network is trained to select an optimal skill.
- Conventional RL- SAC [3]

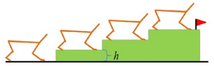
Performance of IST

The Figure 6 plots the reward collection of our method and three baselines on various tasks. Due to the independence and lower dimension of skills, it's observed that the proposed strategy collects reward more efficiently and exhibits stronger generalization ability.

(a) HCH - The agent requires to stride over a hurdle and walk to the destination.

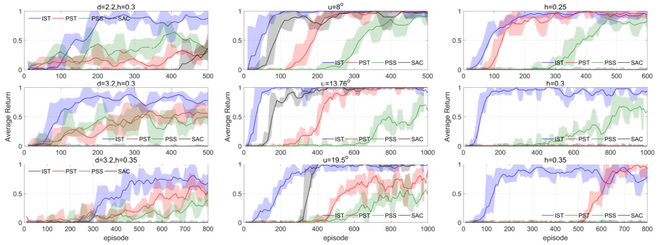


(b) HCA - The agent has to climb up to a hill to reach the destination.



(c) HCU - The agent goes upstairs to reach the destination.

Figure 5: Complex tasks.



(a) Average return in HCH. (b) Average return in HCA. (c) Average return in HCU.

Figure 6: Reward collection of IST, PST, PSS and SAC on various tasks.

Skill Transfer on Difficult Tasks

As shown in Table 1, when the given task gets harder, all skill transfer methods suffer from a degradation of performance in terms of success rate. However, the proposed IST achieves the least degradation and the best performance compared with others (PST has nearly the same degradation, but less success rate), indicating that IST is less sensitive to the difficulty level of tasks.

Environment	HCH			HCA			HCU		
Difficulty-level	d=2.2 h=0.3	d=3.2 h=0.3	d=3.2 h=0.35	u=8°	u=13.76°	u=19.5°	h=0.25	h=0.3	h=0.35
IST	**93.2 %**	**84.9 %**	**83.8%**	**100%**	99.3%	97.2%	**98.8%**	**97.4%**	**95.2%**
PST	73.5%	69%	64.7%	99.9%	99.1%	97.2%	97.3%	97.1%	94.8%
PSS	50.1%	32.4%	37.2%	95%	58.8%	45.1%	75.8%	72.3%	–
SAC	80.5%	75.4%	–	99.2%	**99.7%**	**98.1%**	–	–	–

Table 1: Success rate of IST, PST, PSS and SAC over HCH, HCA and HCU within 1000 episodes.

Conclusion

We propose to learn independent skills from primitive skills and further transfer them to high-level complex tasks. Effective observation collection and independent skills guarantee the success of low-dimension skill transfer. Experiment results show a higher learning efficiency and stronger generalization ability of our proposed method.

References

[1] Xue Bin Peng, Michael Chang, et al.. Mcp: Learning composable hierarchical control with multiplicative compositional policies. arXiv preprint arXiv:1905.09808, 2019.

[2] Archit Sharma, Shixiang Gu, Sergey Levine, et al. Dynamics aware unsupervised discovery of skills. arXiv preprint arXiv:1907.01657, 2019.

[3] Tuomas Haarnoja, Aurick Zhou, et al. Soft actor-critic algorithms and applications. arXiv preprint arXiv:1801.01290, 2018.