Independent Skill Transfer for Deep Reinforcement Learning

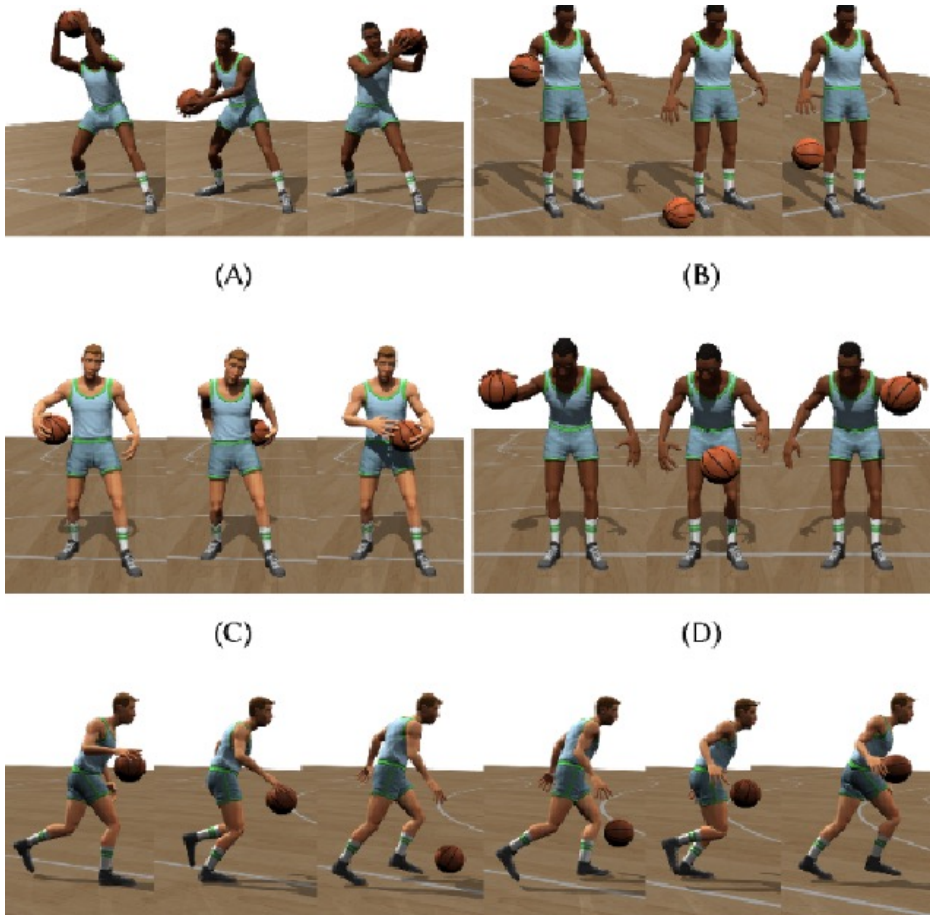
Qiangxing Tian[1,2] , GuanchuWang[2] , Jinxin Liu[1,2] , Donglin Wang[2] and Yachen Kang[2]

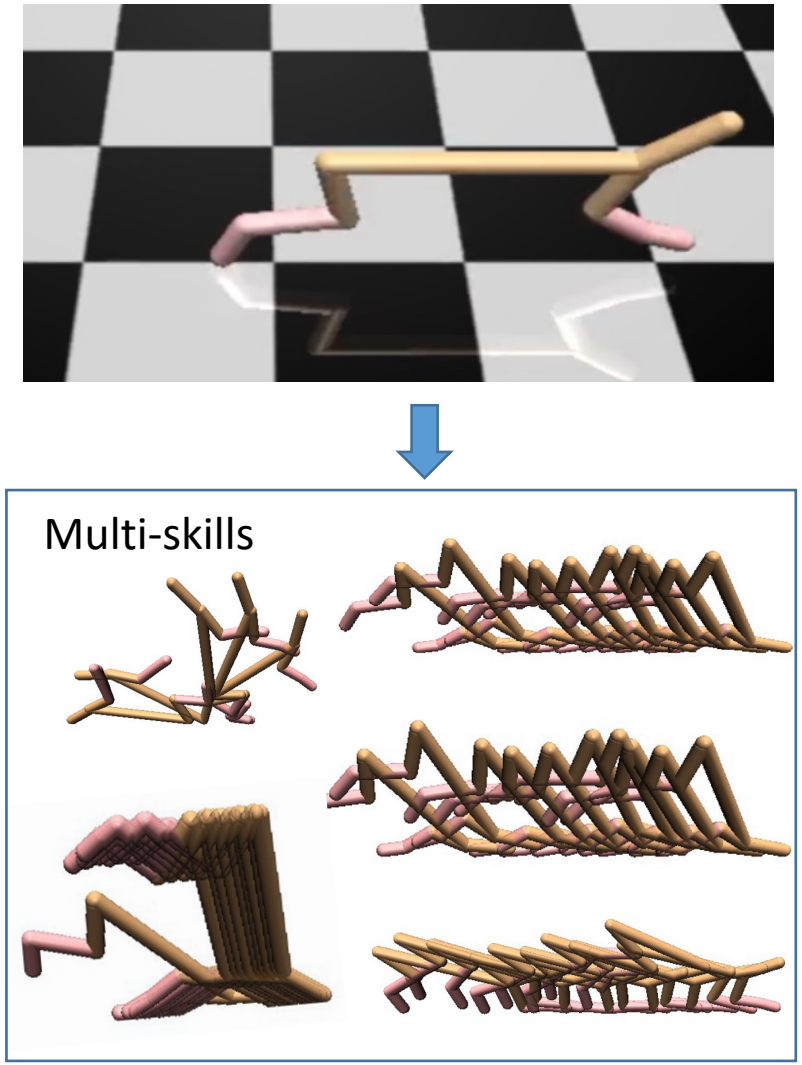
[1]Zhejiang University, Hangzhou, China

[2]School of Engineering, Westlake University, Hangzhou, China

Motivation

- Background





Motivation

- Problem 1 - Strong correlation

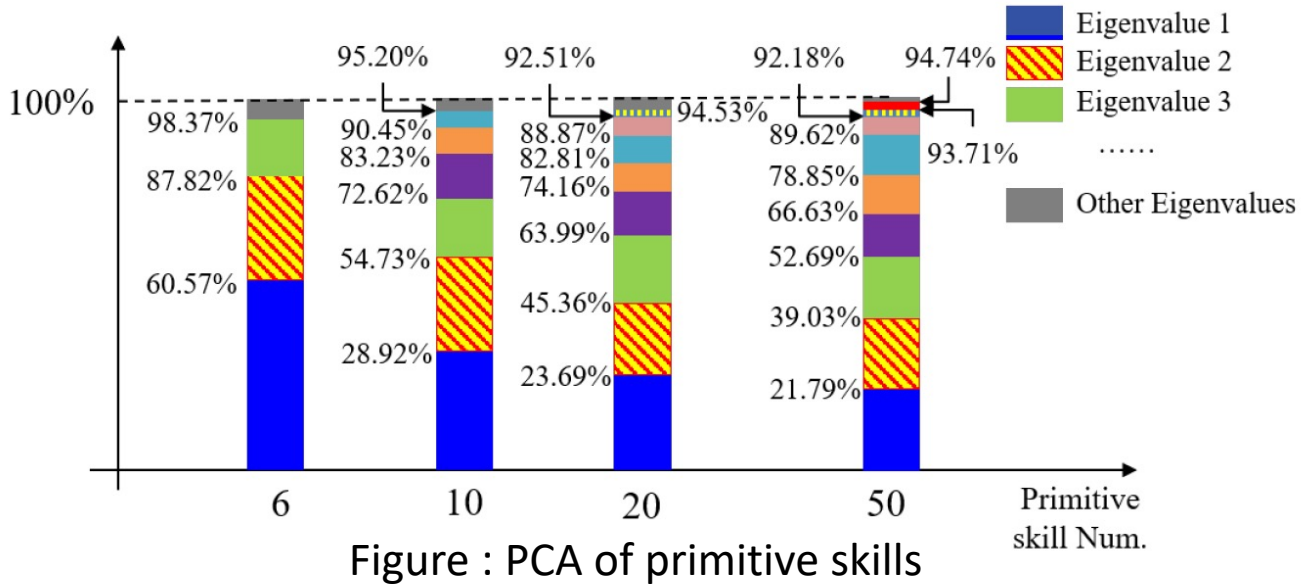


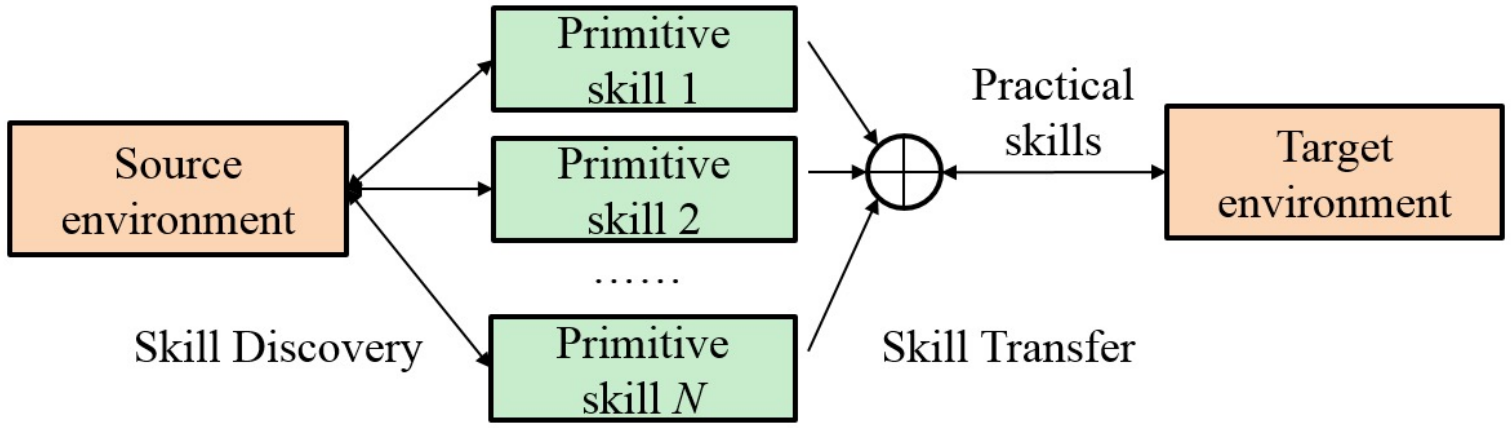
Figure : PCA of primitive skills

The normalized eigenvalue of cross-correlation denotes the percentage of each independent component.

- 6; 10; 20 and 50 primitive skills are generated by existing skill discovery algorithms.
- PCA on actions shows that there is strong correlation between these primitive skills.

Motivation

- Problem 2 - Unbalance of skill discovery and transfer



Primitive skills are learned separately.

Primitive skills are combined.

Our Method

Framework

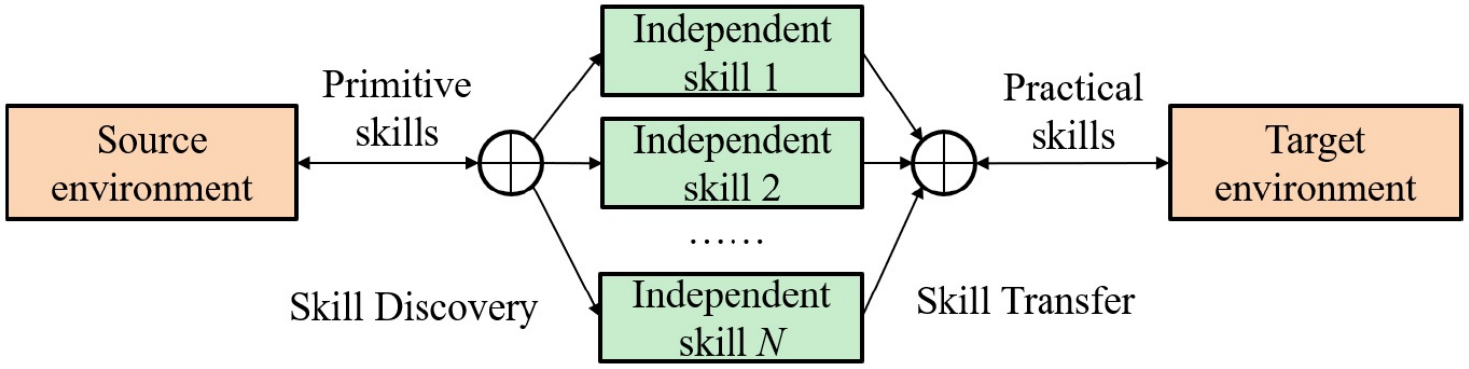


Figure: Independent Skill Transfer (IST).

- Be able to reduce the dimension of skills and enhance the efficiency of skill transfer.
- Each of primitive skills is the combination of all independent skills, which is balanced with the combined practical skill.

Our Method

Learn Independent Skills (LIS)

- Collection of Observation and Action

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

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

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

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

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

- Generation of Independent actions

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

- Generation of Independent Skills

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

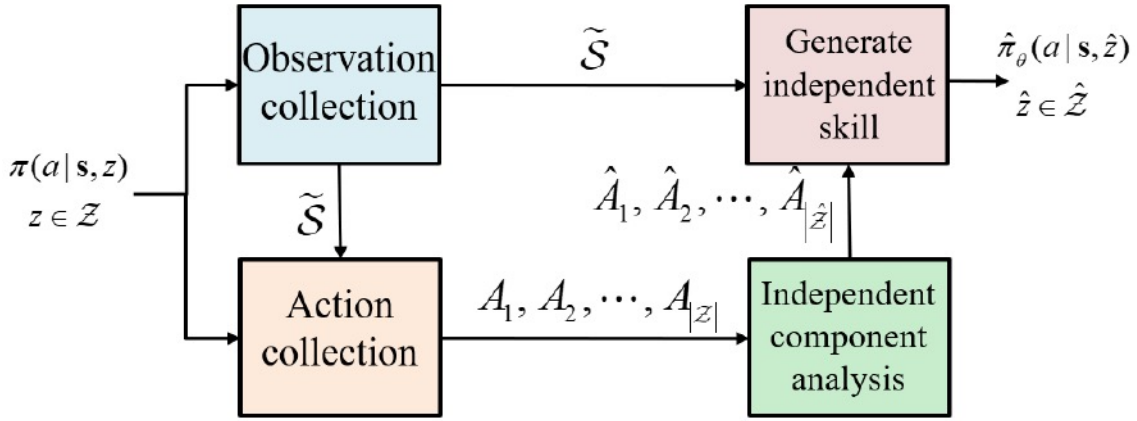


Figure: Framework of Learning Independent Skills (LIS).

Our Method

- Independent Skill Transfer (IST)

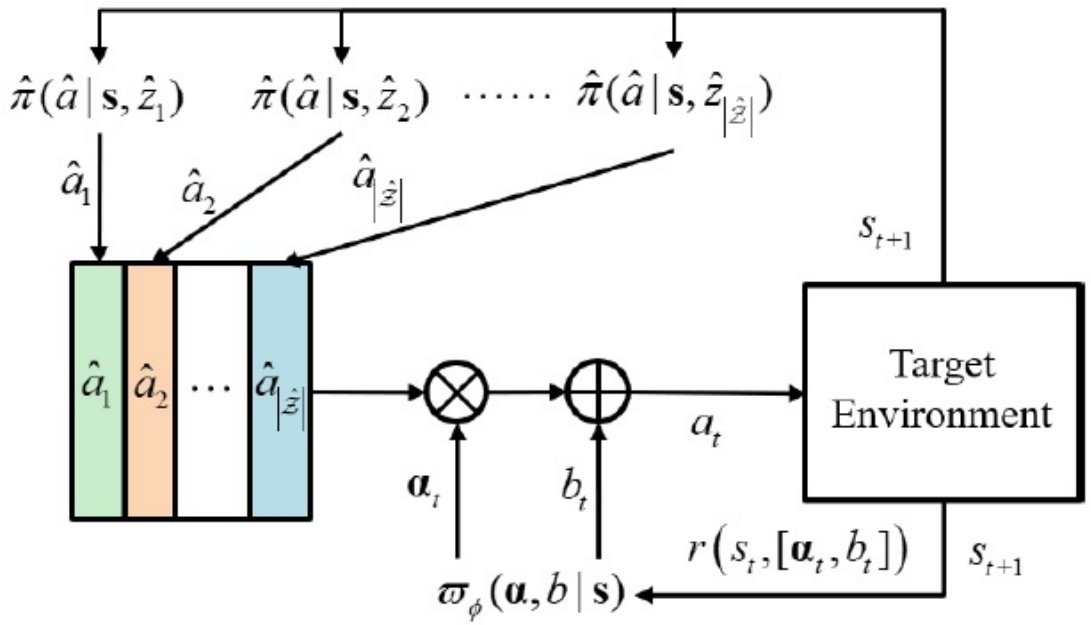


Figure: Process of Independent Skill Transfer (IST)

Independent skill:

$$\hat{\pi}_\theta(\hat{\boldsymbol{a}}|\boldsymbol{s}_t,\hat{z}_1)$$

Transfer policy:

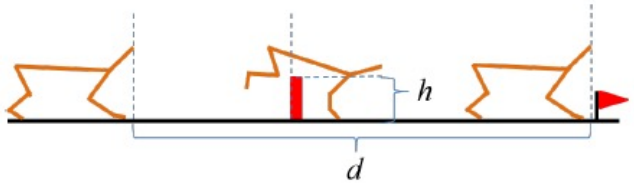
$$\varpi_\phi([\boldsymbol{\alpha},b]|\boldsymbol{s})$$

Composite action:

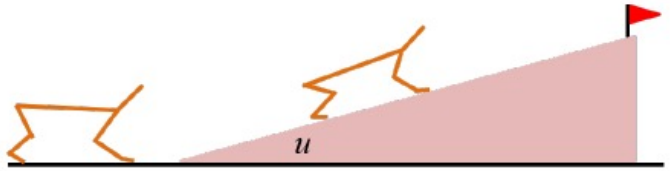
$$\boldsymbol{a}_t = \hat{\boldsymbol{a}}_1\alpha_{t,1} + \hat{\boldsymbol{a}}_2\alpha_{t,2} + \ldots + \hat{\boldsymbol{a}}_{|\hat{\boldsymbol{z}}|}\alpha_{t,|\hat{\boldsymbol{z}}|} + \mathbf{1}\otimes b_t$$

Experiment

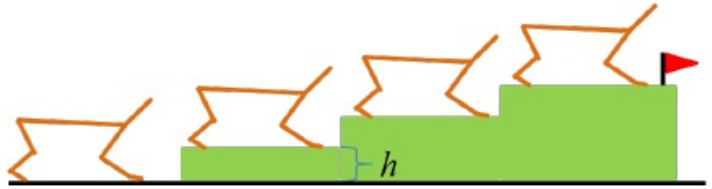
Environment setting



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



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



HCU -- The agent goes upstairs to reach the destination.

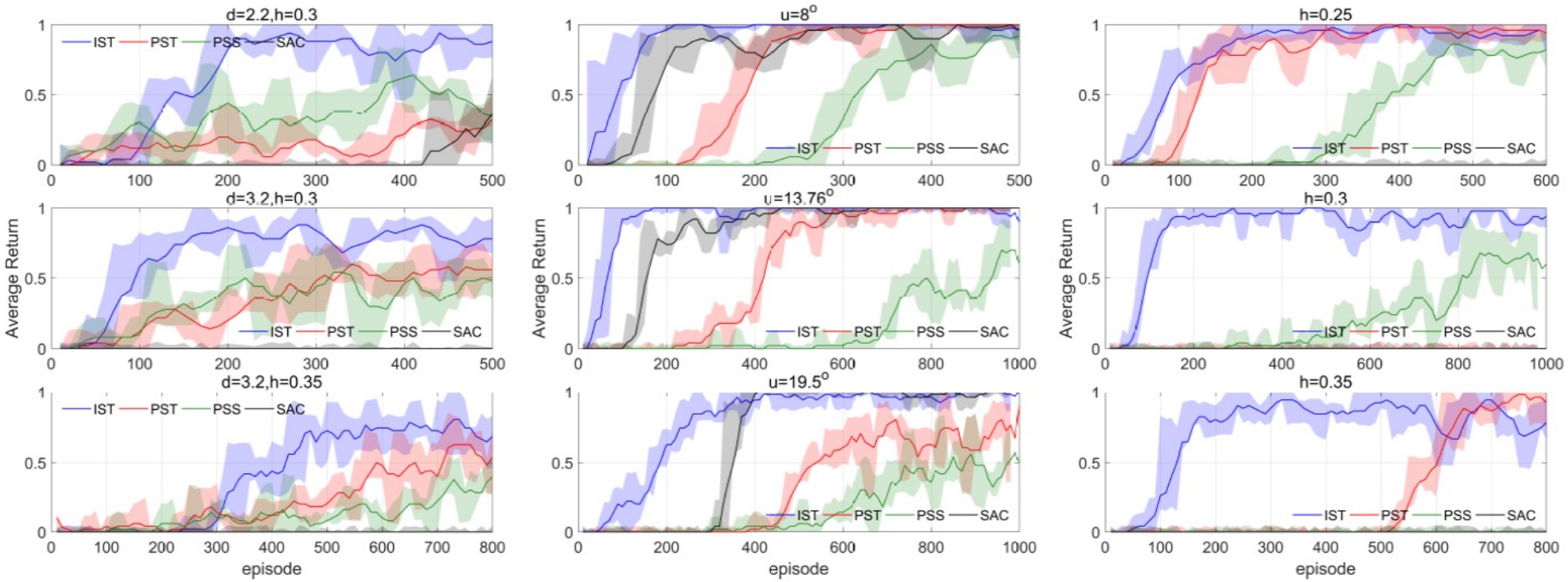
Baselines

- Primitive skill transfer PST [1].
- Primitive skill selection (PSS) [2].
- Conventional RL- SAC [3].

[1] Xue Bin Peng, 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.

Experiment

- Performance of IST



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

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

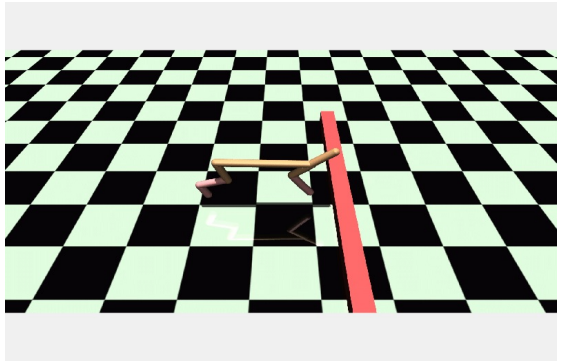
- Compared with PST.
- Compared with primitive skill selection (PSS).
- Compared with conventional RL.

Experiment

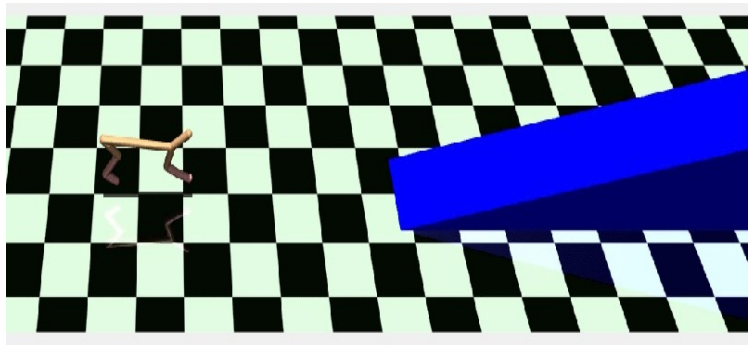
- Skill Transfer on Difficult Tasks

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%**	–	–	–

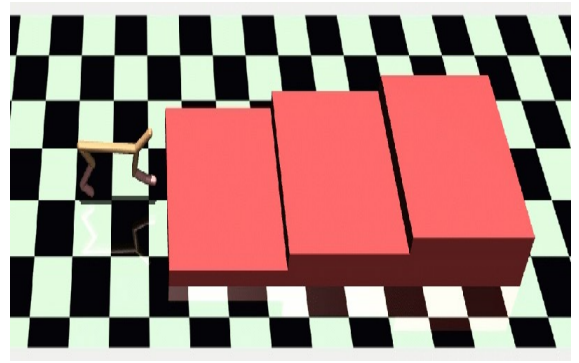
Table 1: Success rate of IST, PST, PSS and SAC over HCH, HCA and HCU within 1000 episodes.



HCH



HCA



HCU

Conclusion

- The correlation between skills can be largely deducted and a lower dimension is thus obtained to enhance the efficiency of skill transfer.

- Combination of independent skills take effects in both skill discovery and transfer, where transferring independent skills is more essential.

- Independent skills are task-independent, which can be transferred to a variety of practical skills in a target environment.

西湖大學
WESTLAKE UNIVERSITY

IJCAI-PRICAI
YOKOHAMA
2021

Thanks for attention.