Reward Dimension Reduction for Scalable Multi-Objective Reinforcement Learning

(QR code is changed. Use the one here :))

Giseung Park, Youngchul Sung

KAIST

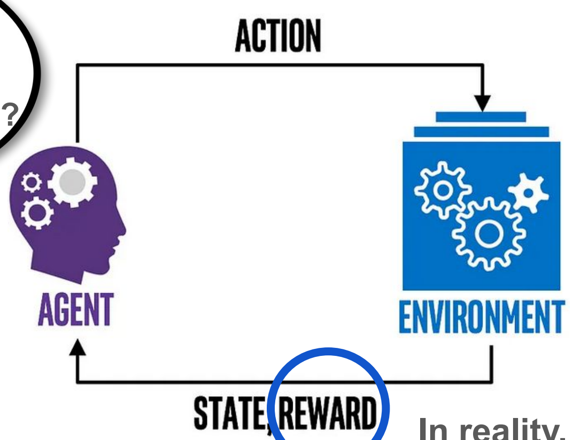
Giseung Park

Giseung Park is seeking a **postdoctoral position** in RL!

https://sites.google.com/view/giseung-park

Multi-Objective Reinforcement Learning (MORL)

How to make good decisions under multiple objectives?



In reality, reward consists of multiple quantities.

reward: vector function

MORL and Reward Uncertainty

(Scalar) Reward uncertainty set R (Regan and Boutilier, 2010)

- Strategy: conservative vs multiple-trial
- Bounded convex polytope
- Vertex functions = quantities we are interested in or able to measure.

1. Conservative strategy = max-min MORL (Park et al., 2024)
2. Multiple-trial strategy = multi-policy MORL with linear scalarization

$$\max_{\pi(\cdot|\cdot,\omega)} \omega^\top \left(\mathbb{E}_\pi \left[\sum_{t=0}^{\infty} \gamma^t \mathbf{r}_t \right] \right) = \mathbb{E}_\pi \left[\sum_{t=0}^{\infty} \gamma^t (\omega^\top \mathbf{r}_t) \right], \ \forall \omega \in \Omega.$$

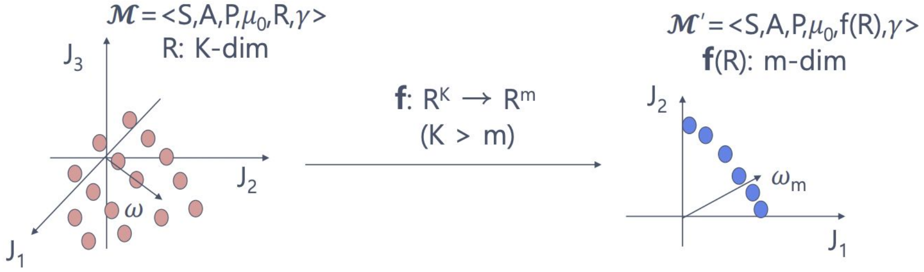
G. Park, W. Byeon, S. Kim, E. Havakuk, A. Leshem, and Y. Sung. (ICML 2024)

Multi-Policy MORL

Key Challenge

Multi-policy MORL struggles with scalability, but objectives often exhibit correlations. However, few online dimension reduction methods exist.

Solution: Online reward dimension reduction method ensuring Pareto-optimality



	Base	PCA	AE	NPCA	**Ours**
HV($\times 10^{61}, \uparrow$)	4.4 ± 6.8	0	0.007 ± 0.018	19.4 ± 15.3	$\mathbf{166.9} \pm 48.1$
SP($\times 10^{5}, \downarrow$)	1842 ± 1290	3837 ± 2164	7834 ± 3323	34.2 ± 52.3	$\mathbf{2.3} \pm 1.0$

G. Park and Y. Sung. (ICLR 2025)

Poster Information + more

- **Thu 24 April.** Hall 3 + Hall 2B #398
- **3 p.m. — 5:30 p.m.** Singapore time
- Let's chat!

Giseung Park is seeking a **postdoctoral position** in RL!

https://sites.google.com/view/giseung-park