Variational Empowerment as Representation Learning for Goal-based Reinforcement Learning

Jongwook Choi Archit Sharma Honglak Lee Sergey Levine Shixiang (Shane) Gu

University of Michigan
Google Research

ICML 2021

Introduction

- **Goal-Conditioned RL** (GCRL): learn optimal policies that control some states to desired goal states
- **Empowerment** (VIC, DIAYN, DADS): by maximizing the mutual information (MI) between state and latent code (skill or goal), we can learn diverse skills or goal representations and reward functions for goal-reaching **without reward**

Our Contributions:

- We view variational MI as a principled framework for representation learning in goal-based RL, through an **unifying perspective** for GCRL and variational empowerment (VE) algorithms
- GCRL as Variational Empowerment:
  - We derive novel variants of GCRL such as adaptive variance and linear-mapping GCRL
  - We find that **regularization** of the posterior is important (e.g. spectral normalization)
- Variational Empowerment as GCRL**:**
  - We extend HER (Hindsight experience Replay) to **Posterior HER** for MI-based RL
  - We propose **Latent Goal Reaching (LGR)** metric for evaluating VE algorithms

Unification of GCRL and Variational Empowerment

- The Barber-Agakov lower bound of MI:

$$\mathcal{I}(s,z) = \mathcal{H}(z) - \mathcal{H}(z|s) \geq \mathcal{H}(z) + \mathbb{E}_{z,s\sim p^{\pi}_{\theta}(z,s)}\left[\log q_{\lambda}(z|s)\right]$$

- Jointly optimization w.r.t. policy and variational posterior (e.g., DIAYN, VISR):

$$\mathcal{F}(\theta,\lambda) = \mathbb{E}_{z\sim p(z), s\sim\pi_{\theta}}\left[\log q_{\lambda}(z|s)\right]$$

- Key Observation: This objective encapsulates the standard GCRL, when a fixed-variance Gaussian distribution $q_{\lambda}(z|s) = \mathcal{N}(z; s, \sigma^2 I)$ is used for the posterior:

$$F(\pi) = \mathbb{E}_{z\sim p(z), s\sim\pi_{\theta}}\left[-\frac{1}{\sigma^2}\|z-s\|^2\right]$$

- This provides a novel interpretation for GCRL as a variational empowerment algorithm with a hard-coded and fixed variational distribution: by varying expressivity through the choice of $q(z|s)$, we can interpolate between GCRL and variational empowerment

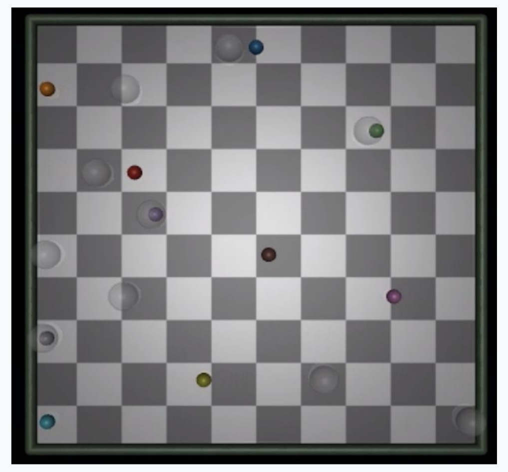
Method	Goal space	$q_{\lambda}(z\|s)$	Learnable λ	Learning π_z
GCRL (Kaelbling, 1993)	Continuous ($\mathbb{R}^d$)	$\mathcal{N}(s,\sigma^2 I)$	-	(HER)
aGCRL (**ours**)	Continuous ($\mathbb{R}^d$)	$\mathcal{N}(s,\Sigma)$	Σ	HER
linGCRL (**ours**)	Continuous ($\mathbb{R}^d$)	$\mathcal{N}(As,\sigma^2 I)$	A	P-HER
InfoGAIL* (Li et al., 2017)	Discrete	Categorical	q_{λ}	-
DIAYN (Eysenbach et al., 2019)	Discrete	Categorical	q_{λ}	-
DIAYN (continuous)	Continuous ($\mathbb{R}^d$)	$\mathcal{N}(\mu(s),\Sigma(s))$	$\mu(\cdot)$	-
DISCERN (Warde-Farley et al., 2019)	$= \mathcal{S}$ (e.g. image)	Non-parametric	Embedding$(\cdot)$	HER
VISR (Hansen et al., 2020)	Continuous ($\mathbb{R}^d$)	vMF$(\mu(s),\kappa)$	$\mu(\cdot)$	SF
VGCRL	Any	Any	Any	P-HER or SF

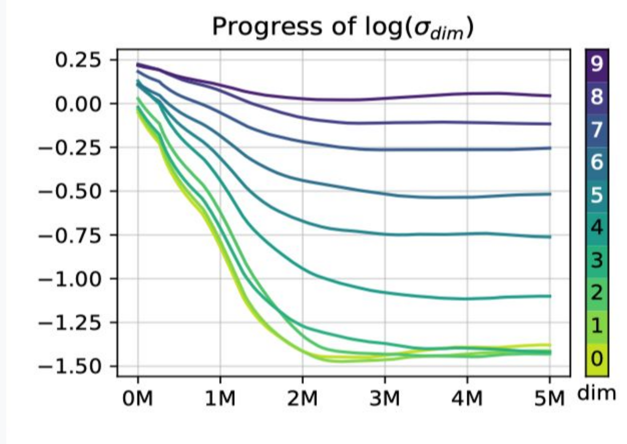
Algorithm 1 Latent Goal Reaching Metric

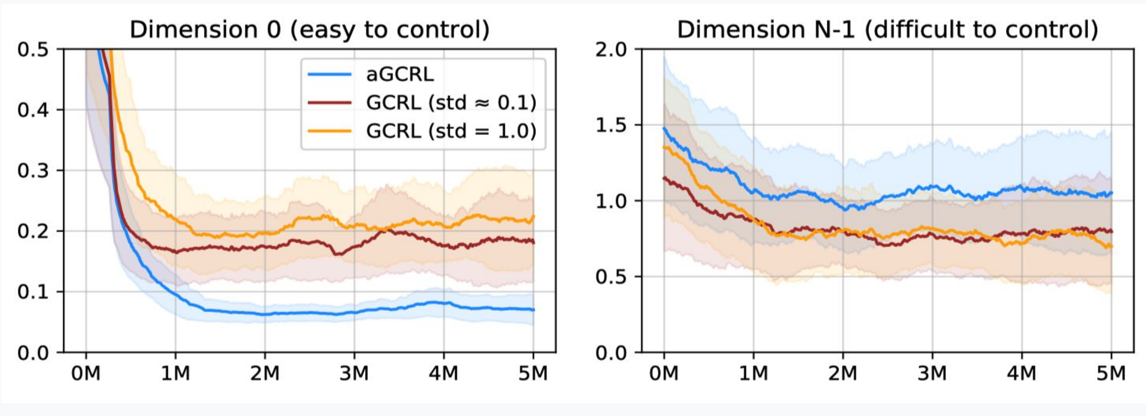
Input: target states $s^{1:N}$, trained $\pi_{\theta}(a|s,z), q_{\lambda}(z|s)$
Output: average goal distance $\bar{d}$, average latent distance $\bar{z}$
$\bar{d} \leftarrow 0, \bar{z} \leftarrow 0$
for $i \leftarrow 1$ **to** N **do**
  Embed target $z^i \leftarrow \mathbb{E}\left[q_{\lambda}(\cdot|s^i)\right]$
  Run $\pi(\cdot|\cdot, z^i)$ for T time steps, observe final state s^i_T
  Embed reached state $z^i_T \leftarrow \mathbb{E}\left[q_{\lambda}(\cdot|s^i_T)\right]$
  $\bar{d} \leftarrow \bar{d} + d(s^i, s^i_T)/N, \bar{z} \leftarrow \bar{z} + d(z^i, z^i_T)/N$
end

GCRL as Variational Empowerment

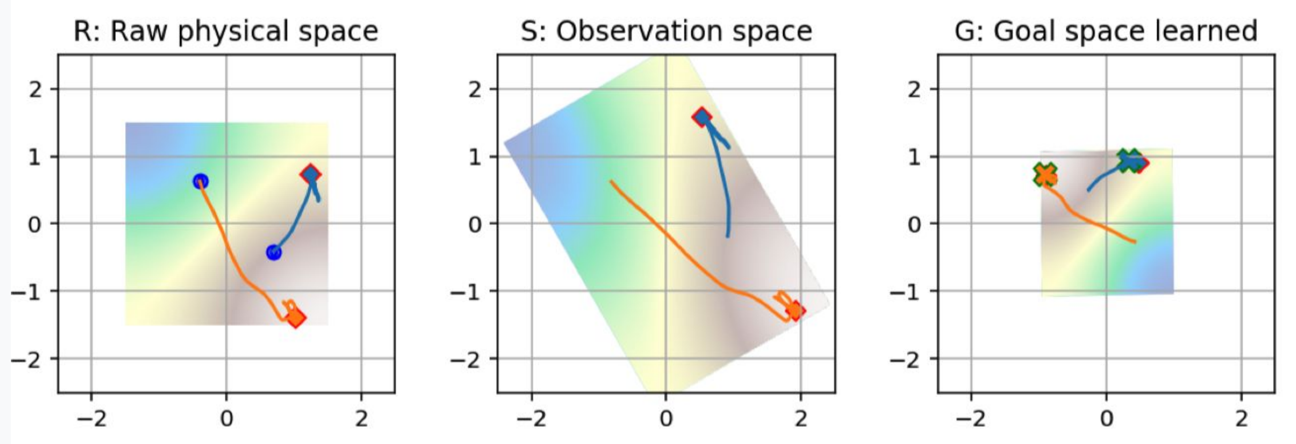
- Adaptive-variance GCRL (aGCRL) can prioritize goal-reaching in more controllable dimensions, similar to automatic relevance determination (ARD): e.g., $q_{\lambda}(z|s) = \mathcal{N}(s,\Sigma)$

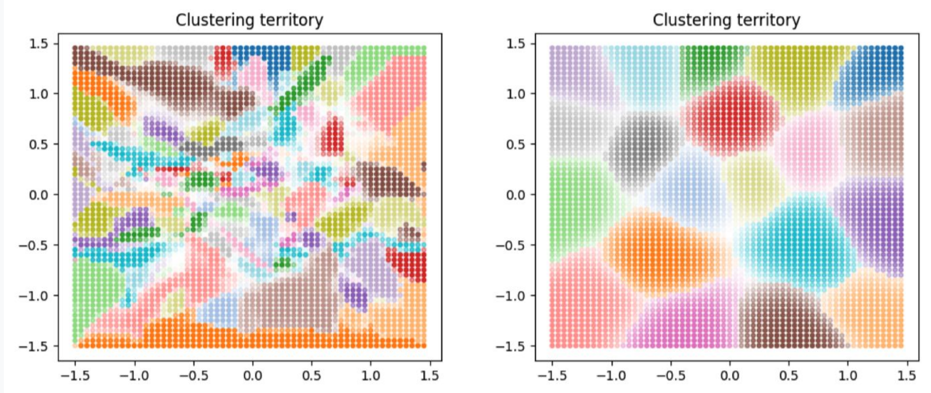






- Linear-mapping GCRL: Recovering intrinsic dimensions of variations, e.g., $q(z|o) = \mathcal{N}(Ao, \Sigma)$





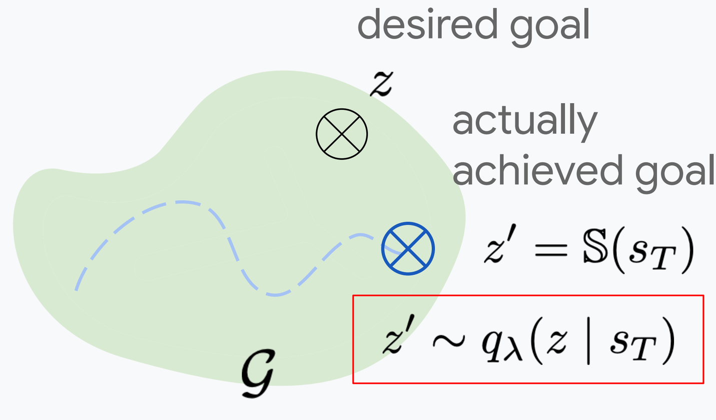
(a) Without SN. (b) With SN.

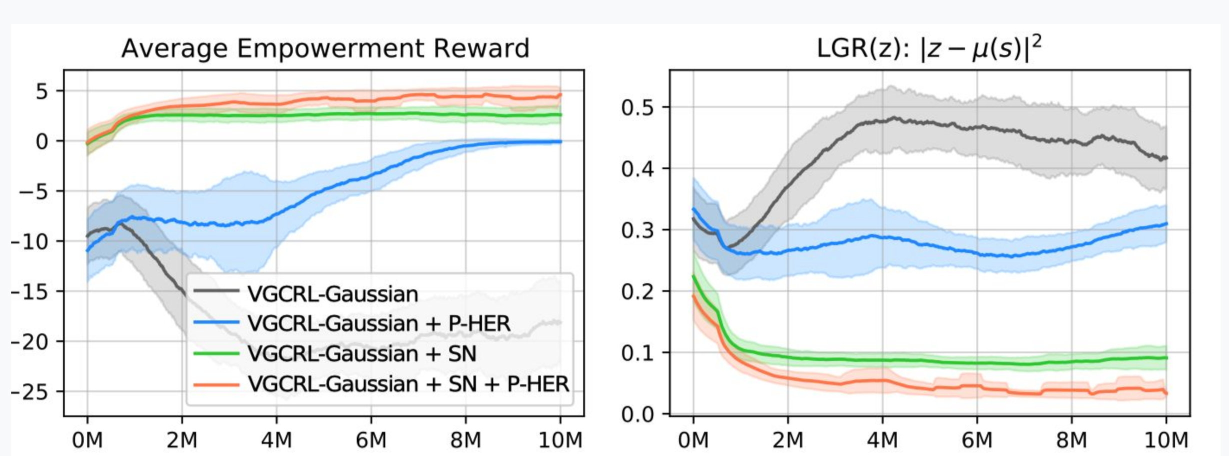
- However, high expressivity may not necessarily mean better performance for latent space learning and representation learning: proper regularizations such as Spectral Normalizations (SN) can play an important role for VE and VGCRL.

Variational Empowerment as GCRL

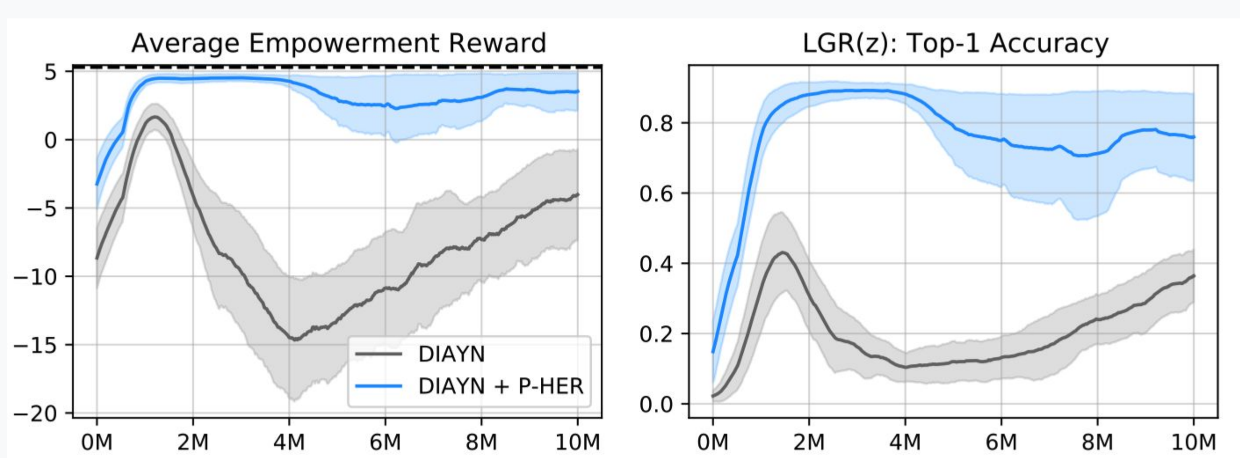
We can transfer some known techniques for GCRL to better understand VE:

- **Posterior HER** for accelerating variational empowerment algorithms: relabel the latent goal with an up-to-date estimate from the variational posterior $q_{\lambda}(z|s_T)$ when training $\pi_{\theta}(a|s,z)$
- Latent Goal Reaching (**LGR**) Metric: evaluate MI-based RL as just another goal-reaching problem given target states of interest





(d) Ant-v3 (Gaussian, $|\mathcal{G}| = 5$)



(b) Humanoid-v3 (Discrete, $|\mathcal{G}| = 200$)