Action and Perception as Divergence Minimization

Danijar Hafner[1,2], Pedro Ortega[3], Jimmy Ba[2], Thomas Parr[4], Karl Friston[4], Nicolas Heess[3]

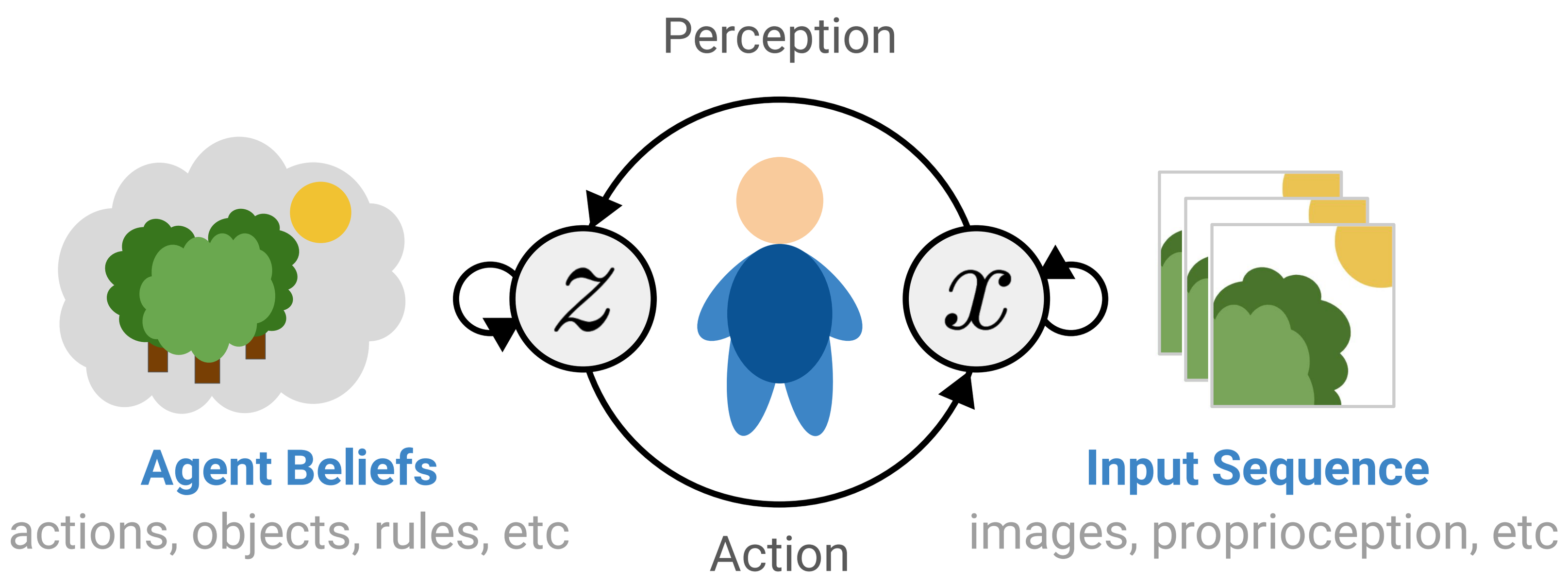
[1]Google Brain [2]University of Toronto [3]DeepMind [4]University College London

1 Overview

Action and Perception as Divergence Minimization (APD)...

1. Map of **possible agent objectives** that correspond to different latent variables, target factorizations, and divergence measures
2. **Unified perspective** on representation learning, infogain exploration, and empowerment
3. Representation learning should be **paired** with infogain exploration for a temporally consistent objective
4. **World models** as path toward adaptive infomax agents while making task rewards optional
5. **Future objectives** should be derived from a joint divergence to facilitate comparison and make target explicit

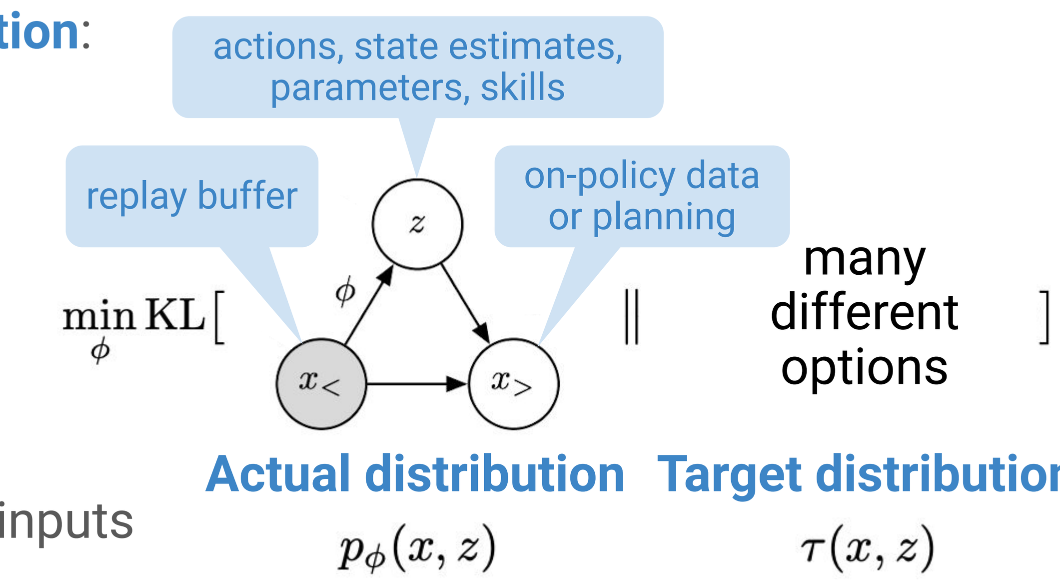
2 Agents with Latent Variables



Parameterize belief (incl actions) by ϕ

3 Joint KL Minimization

Formulate agent objective as bringing its current **actual distribution** toward a **target distribution**:



x lifetime trajectory of inputs

z set of agent latents

ϕ parameters of agent beliefs $p_\phi(z \mid x_<)$ over repr and actions

4 Target Dependencies

Factorized Targets

$$\tau(x)\tau(z)$$

Inputs and latents have zero mutual information under the target

Agent **minimizes** mutual information in the actual dist

Examples

- MaxEnt RL uses reward factor and action prior to solve the task while keeping actions as random as possible

Expressive Targets

$$\tau(x \mid z)\tau(z)$$

Target knows or learns depen. between inputs and latents

Agent **maximizes** the mutual information in the actual dist

Examples

- World models learn representations that are informative of past inputs
- Reverse predictors learn skills that maximally influence future inputs

5 Information Bounds

Minimizing joint KL to an expressive target...

- realizes the **preferences** expressed by the target
- maximizes variational bound on the **mutual information** between inputs and latents
- bound is tighter the better the target can express dependencies

$$\underbrace{\mathrm{KL}\big[p_\phi(x,z)\;\big\|\;\tau(x,z)\big]}_{\text{joint divergence}} = \underbrace{\mathrm{E}\,\overbrace{\mathrm{KL}\big[p_\phi(z \mid x)\;\big\|\;\tau(z)\big]}^{\text{simplicity}}}_{\text{realizing latent preferences}} - \underbrace{\mathrm{E}\big[\overbrace{\ln\tau(x \mid z)}^{\text{accuracy}} - \overbrace{\ln p_\phi(x)}^{\text{input entropy}}\big]}_{\text{information bound}}$$

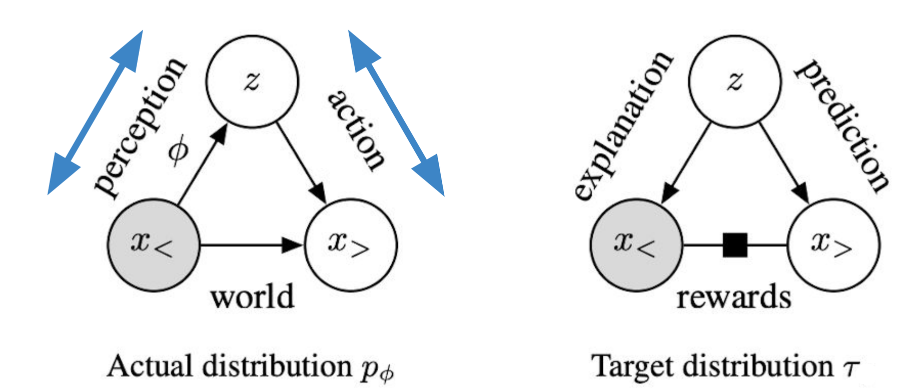
$$\underbrace{\mathrm{KL}\big[p_\phi(x,z)\;\big\|\;\tau(x,z)\big]}_{\text{joint divergence}} = \underbrace{\mathrm{E}\,\overbrace{\mathrm{KL}\big[p_\phi(x \mid z)\;\big\|\;\tau(x)\big]}^{\text{control}}}_{\text{realizing input preferences}} - \underbrace{\mathrm{E}\big[\overbrace{\ln\tau(z \mid x) - \ln p_\phi(z)}^{\text{information gain}}\big]}_{\text{information bound}}$$

6 Past and Future

Agents with expressive targets...

- infer latent representations that are **informative** of past inputs
- explore future inputs that are **informative** of the representations

$$\mathrm{I}\big[x; z\big] = \mathrm{I}\big[x_<; z\big] + \mathrm{I}\big[x_>; z \mid x_<\big]$$



(see eq 6 in the paper)

7 Types of Latents

Latent Variable	Past Infomax	Future Infomax
Actions	N/A past actions are observed	Empowerment & MaxEnt RL VIM, ACIE, EPC, SQL, SAC
Skills	N/A past skills are observed	Skill Discovery VIC, SNN, DIAYN, VALOR
State Estimates	State Estimation VAE, DVBF, SOLAR, PlaNet	State Information Gain NDIGO, DVBF-LM
Dynamics Parameters	System Identification PETS, Bayesian PlaNet	Dynamics Information Gain VIME, MAX, Plan2Explore
Policy Parameters	Belief over Policies BootDQN, Bayesian DQN	Policy Information Gain BootDQN, Bayesian DQN

8 Action Perception Cycle

Action and perception optimize the same objective but receive and affect different variables.

Under a unified target distribution...

- perception makes the **agent beliefs consistent with the world**
- actions make the **world consistent with the agent beliefs**

action

$$\min_\phi \mathrm{KL}\Big[\;\underbrace{p_\phi(z \mid x)}_{\text{beliefs}}\;\underbrace{p_\phi(x)}_{\text{inputs}}\;\Big\|\;\underbrace{\tau(x,z)}_{\text{target}}\;\Big]$$

perception

9 Niche Seeking

Minimizing a joint divergence also brings the marginals together

$$\mathrm{KL}\big[p_\phi(x,z)\;\big\|\;\tau(x,z)\big] \geq \mathrm{KL}\big[p_\phi(x)\;\big\|\;\tau(x)\big]$$

The marginal target distribution over inputs is the marginal likelihood

The agent thus converges to an **ecological niche**...

- See inputs propto how well the agent can learn to predict them
- That is large because of the information gain exploration
- That it can inhabit despite external perturbations

The agent thus seeks out a large niche that it can inhabit and understand

Models that assign high prob to **more trajectories** lead to **larger niches**

@danijarh

Project website with video: danijar.com/apd