Action and Perception as Divergence Minimization

Danijar Hafner, Pedro Ortega, Jimmy Ba, Thomas Parr, Karl Friston, Nicolas Heess

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

\[ \phi \]

\( \tau(x) \) parameterized belief (incl actions)

Input Sequence

\( z \) set of agent latents

Perception

\( x \) lifetime trajectory of inputs

Action

3 Joint KL Minimization

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

\[ \min KL(p(x,z) \| \tau(x,z)) \]

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

\( \tau(x | z) \tau(z) \)

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

5 Information Bounds

Minimizing joint KL to an expressive target...

- realises 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

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

7 Types of Latents

Latent Variable

Past Infomax

Future Infomax

Actions

N/A

past actions are observed

Empowerment & MaxEnt RL

Skills

N/A

past skills are observed

Skill Discovery

State Estimation

VAE, DVB, SOLAR, PlanNet

State Estimation Gain

N/A

VAE, DVB, SOLAR, PlanNet

State Information Gain

NDIGO, DVB-LM

Dynamics Parameters

System Identification

PETs, Bayesian PlanNet

Dynamics Information Gain

VIME, MAX, PlanExplore

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

\[ \min KL(p(x,z) \| \tau(x,z)) \]

9 Niche Seeking

Minimizing a joint divergence also brings the marginals together

\[ KL(p(x,z) \|| \tau(x,z)) \geq KL(p(x) \|| \tau(x)) \]

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