DeepMDP

Learning Latent Space Continuous Models for Representation Learning

Carles Gelada, Saurabh Kumar, Jacob Buckman, Ofir Nachum, Marc G. Bellemare
Simple Representations for RL
DeepMDP

Latent Space Model:

\[ \tilde{\mathcal{M}} = \langle \tilde{S}, \tilde{A}, \tilde{R}, \tilde{P}, \gamma \rangle \]

\[ \mathcal{M} = \langle S, A, R, P, \gamma \rangle \]

& trained via the following two losses:
Reward Loss

\[ \phi(s) \xrightarrow{\overline{R}} \overline{R}(\phi(s), a) \]

\[ s, a \xrightarrow{R} R(s, a) \]
Transition Loss

\[
\phi(s) \xrightarrow{\mathcal{P}} \tilde{\mathcal{P}}(\cdot | \phi(s), a) \quad \xrightarrow{L_D(s, a)} \quad \phi \mathcal{P}(\cdot | s, a)
\]

\[
\mathcal{P}(\cdot | s, a) \quad \xrightarrow{\phi} \quad \phi \mathcal{P}(\cdot | s, a)
\]
Tractable Losses

\[ L^\xi_{\bar{R}} = \mathbb{E}_{s,a \sim \xi} |\mathcal{R}(s,a) - \bar{\mathcal{R}}(\phi(s),a)| \]

\[ L^\xi_D = \mathbb{E}_{s,a \sim \xi} \left[ \mathcal{D}(\phi \mathcal{P}(\cdot | s,a), \bar{\mathcal{P}}(\cdot | \phi(s),a)) \right] \]
Deep Policies

\[ \bar{\Pi} \subset \Pi \]

\[ \bar{\pi}(a|s) := \bar{\pi}(a|\phi(s)) \]
Representation $\phi$ Quality
Only Discards:

\[ \pi(\text{Previous State}) = \text{Down} \]

\[ \pi(\text{Current State}) = \text{Up} \]

Phi as a Representation

\[ \mathbb{E}_{s,a \sim \xi_{\pi}} |Q^\pi(s, a) - \bar{Q}^\pi(\phi(s), a)| \leq \frac{\left( L_{\xi\pi}^{\mathcal{R}} + \gamma K_{\bar{\nu}} L_{\xi\pi}^{\mathcal{D}} \right)}{1 - \gamma} \]
Donut World
DeepMDP on Donut World

2D latent space + DeepMDP losses

\[ \bar{R}(\phi(s), a) \quad \bar{P}(\cdot | \phi(s), a) \]

\[ \phi(s) \in \mathbb{R}^2 \]
DeepMDP on Donut World

Visualization of latent distance
DeepMDP
Auxiliary Task

Base C51 agent + DeepMDP losses
DeepMDP
Auxiliary Task

Base C51 agent
+ DeepMDP losses
• DeepMDPs as Models of the Environment

• Norm-MMD Metrics and their Associated Smoothness
Thanks For Listening

Poster #108