On the Generalization Gap in Reparameterizable Reinforcement Learning

Huan Wang, Stephan Zheng, Caiming Xiong, Richard Socher

Salesforce Research

June 12, 2019
Summary

- **Reparameterize** RL to decouple randomness of the environment from the policy.

- Bring in supervised learning theory to RL thanks to reparameterization.

- Theoretical guarantees on the generalization gap in RL.

- Generalization gap is related to:
  - Number of training episodes
  - Smoothness of the environment, policy and reward
  - Discrepancy between training and test environment
  - "Complexity" of the reward and transition function class.
Reparameterization using Gumbel-max trick

- Treat each episode, $s^i = [s^i_0, s^i_1, \ldots, s^i_T] \sim D_\pi$, as a sample.
- Reparameterization using Gumbel noise $g^i_t \sim G$:
  $$s^i_{t+1} = \text{arg max } \left[ \log P(s^i_t, \pi(s^i_t)) + g^i_t \right]$$
- Empirical reward:
  $$\hat{\pi} = \text{arg max } \pi \in \Pi, s^i \sim D_\pi \frac{1}{n} \sum_i R(s^i)$$
  $$= \text{arg max } \pi \in \Pi, g^i \sim G \frac{1}{n} \sum_i R(s^i(\pi, g^i))$$
- The distribution $G$ is static and independent of learned model $\pi \rightarrow \text{as in supervised learning}$!
Generalization Gap in Episodic RL

\[
\left| \frac{1}{n} \sum_{i=1}^{n} R(s^i) - \mathbb{E}_{s \sim D'} R(s) \right| \\
\leq \left| \frac{1}{n} \sum_{i=1}^{n} R(s^i) - \mathbb{E}_{s \sim D_{\hat{\pi}}} R(s) \right| + \mathbb{E}_{s \sim D_{\hat{\pi}}} R(s) - \mathbb{E}_{s \sim D'} R(s) \\
\epsilon_{\text{intrinsic}}
\]

- Intrinsic Gap: internal randomness from the same “reparameterizable MDP”:

\[
\epsilon_{\text{intrinsic}} \leq \text{Rad}(R_{\pi}, T, I) + O \left( c \sqrt{\frac{\log(1/\delta)}{n}} \right)
\]

- External Gap: between “different reparameterizable MDPs”:

\[
\epsilon_{\text{extrinsic}} \leq L_r \zeta \sum_{t=0}^{T} \gamma^t \nu^t \frac{\nu^t - 1}{\nu - 1} + L_r \epsilon \sum_{t=0}^{T} \gamma^t \nu^t
\]
http://proceedings.mlr.press/v97/wang19o/wang19o.pdf


▶ Come see our poster!

**Wed Jun 12th 6:30 – 9:00 PM @ Pacific Ballroom #43**

▶ Contact us:

  huan.wang@salesforce.com
  stephan.zheng@salesforce.com