Random Expert Distillation For Imitation Learning

Ruohan Wang, Carlo Ciliberto, Pierluigi Amadori, Yiannis Demiris

ICML 2019
Imitation Learning

- Policy learning from a limited set of expert demonstrations
- Intuitive & efficient skills transfer
- Captures styles & preferences
Inverse Reinforcement Learning

- Generative Adversarial Imitation Learning (Ho et al., 2015)

- Optimization challenges
  - Training instability
  - sample inefficiency
Random Expert Distillation (RED)

- Directly learns a reward function with Random Network Distillation (RND) (Burda et al., 2018)

- Considers how “similar” is the agent to the expert, instead of how “different”
Reward Function

Over expert trajectories $D = \{s_i, a_i\}_{i=1}^{N}$ and $f_{\theta}: \mathbb{R}^m \rightarrow \mathbb{R}^n$

$$\theta^* = \min_{\theta} \|f_{\theta}(s, a) - f_{\text{rnd}}(s, a)\|_2^2.$$ 

Define the reward as

$$r(s, a) = \exp(-\sigma \|f_{\theta^*}(s, a) - f_{\text{rnd}}(s, a)\|_2^2)$$

The reward asymptotically estimates the support of the expert policy.
## Mujoco Experiments

<table>
<thead>
<tr>
<th></th>
<th>Hopper</th>
<th>HalfCheetah</th>
<th>Walker2d</th>
<th>Reacher</th>
<th>Ant</th>
</tr>
</thead>
<tbody>
<tr>
<td>GAIL</td>
<td>3614.2 ± 7.2</td>
<td>4515.7 ± 549.5</td>
<td>4878.0 ± 2848.3</td>
<td>-32.4 ± 39.8</td>
<td>3186.8 ± 903.6</td>
</tr>
<tr>
<td>GMMIL</td>
<td>3309.3 ± 26.3</td>
<td>3464.2 ± 476.5</td>
<td>2967.1 ± 702.0</td>
<td>-11.89 ± 5.27</td>
<td>991 ± 2.6</td>
</tr>
<tr>
<td>RED</td>
<td>3626.0 ± 4.3</td>
<td>3072.0 ± 84.7</td>
<td>4481.4 ± 20.9</td>
<td>-10.43 ± 5.2</td>
<td>3552.8 ± 348.7</td>
</tr>
</tbody>
</table>

Image ref: [https://creativestudio2019spring.files.wordpress.com/2019/02/openaigym.png](https://creativestudio2019spring.files.wordpress.com/2019/02/openaigym.png)
Training Stability & Sample Efficiency

Hopper

Reacher
## Driving Task

<table>
<thead>
<tr>
<th></th>
<th>Average</th>
<th>Best</th>
</tr>
</thead>
<tbody>
<tr>
<td>BC</td>
<td>1033 ± 474</td>
<td>1956</td>
</tr>
<tr>
<td>GAIL</td>
<td>795 ± 395</td>
<td>1576</td>
</tr>
<tr>
<td>GMMIL</td>
<td>2024 ± 981</td>
<td>3624</td>
</tr>
<tr>
<td>RED</td>
<td>4825 ± 1552</td>
<td>7485</td>
</tr>
<tr>
<td>Expert</td>
<td>7485 ± 0</td>
<td>7485</td>
</tr>
</tbody>
</table>
Reward function penalizes dangerous driving
Summary

- Random Expert Distillation is a new framework for imitation learning, using the estimated support of the expert policy as reward.

- Our results suggest that RED is viable, robust and attains good performance.

- Future works: combining different sources of expert information for more robust algorithms.
Thank you

- Code: https://github.com/RuohanW/RED

- Check out our poster:
  Pacific Ballroom #39
  6:30 to 9:00 pm today