Bridging the Imitation Gap by Adaptive Insubordination

Luca Weihs*, Unnat Jain*, lou-Jen Liu, Jordi Salvador, Svetlana Lazebnik, Aniruddha Kembhavi, Alexander Schwing

Privileged Experts and Mismatch With Students

Experts have access to full state
Student sees partial (egocentric) observations

Expert sees:

Intuitive Example

[Diagram of different actions and observations]

Expert recommends:

Move forward        Turn left         Turn left/right   Turn right

But student’s partial view makes these recommendations confusing...

Hence, student learns an averaged-policy:

Prop. 1: A student’s policy is the average of the teacher’s policy.

If \( \pi^* = \arg \min \mathbb{E}[\text{CrossEntropy}(\pi, \pi^{\text{exp}})] \)

Then \( \pi^*(o) = \mathbb{E}[\pi^{\text{exp}}(S) | f(S) = o] \)

Adaptive Insubordination (ADVISOR)

Idea:
1. Student estimates if it can imitate the expert
2. Accordingly, weigh IL & RL losses at each step

Schematic:

[Diagram of the ADVISOR model]

\( \mathcal{L}^{\text{ADV}}(\theta) = \mathbb{E}_\mu[w(S) \cdot CE(\pi^{\text{exp}}(S), \pi_f(S; \theta)) + (1 - w(S)) \cdot L(\theta, S)] \)

Auxiliary actor:
- Trained only by IL
- Estimates imitatibility of the current state

Outcome:

Imitable
High w(S) \( \rightarrow \) More IL

Non-Imitable
Low w(S) \( \rightarrow \) More RL

Results

Exhaustive evaluation in MiniGrid

[Graph showing performance across different methods]

Success across diverse tasks

[Diagram showing success across different tasks]

Table: Performance across different navigation tasks

<table>
<thead>
<tr>
<th>Task</th>
<th>PointGoal Navigation</th>
<th>ObjectGoal Navigation</th>
<th>Cooperative Navigation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SPL</td>
<td>@100%</td>
<td>SPL</td>
</tr>
<tr>
<td>Training runs</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RL only</td>
<td>30.9</td>
<td>54.7</td>
<td>6.7</td>
</tr>
<tr>
<td>RL only</td>
<td>30.1</td>
<td>68.7</td>
<td>3.6</td>
</tr>
<tr>
<td>RL + RL static</td>
<td>46.9</td>
<td>72.5</td>
<td>6.5</td>
</tr>
<tr>
<td>ADVISOR (ours)</td>
<td>57.7</td>
<td>77.3</td>
<td>11.9</td>
</tr>
</tbody>
</table>

https://unnat.github.io/advisor