Deep Reinforcement Learning for Atari Games Aided with Human Guidance

Kshitiz Tripathi
Stanford University

Objectives
1. Apply Deep Reinforcement Learning techniques to train an agent to play video games in a generic manner without hand-crafted feature set.
2. Develop an approach for enabling the agent to be guided by a human teacher.

Results

Figure 1: Video Games: pong(left), breakout, tetris

Figure 2: Pong training curve
Figure 3: Breakout training curve

Human Guidance

• Human guidance can play an essential role in helping the bot learn game strategy involving multiple steps. Example: Picking a key in a game to unlock door or a simple one is tunneling strategy in breakout.

• One approach is to train the agent using supervised learning explicitly with samples containing the desired strategy collected from human teacher. Cons: requires too many samples and this may diverge the model from the optimal policy.

RL - Policy Gradient

• Explicit Policy $\pi_\theta(a_t|s_t)$ approximated by neural network.
• Policy Gradient is given by

$$g = E_{t=0}^{\infty} \Psi_t \nabla_\theta \log \pi_\theta(a_t|s_t)$$

where,

$$\Psi_t = \sum_{t'=t}^{\infty} \gamma^{t'-t} r_{t'}$$

(Neural Network)

Figure 6: Policy Network with 2 layers Fully Connected Net.

<table>
<thead>
<tr>
<th>parameters</th>
<th>value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hidden layer Neurons</td>
<td>1000</td>
</tr>
<tr>
<td>Learning Rate</td>
<td>0.0005</td>
</tr>
<tr>
<td>Discount Factor</td>
<td>0.99</td>
</tr>
<tr>
<td>Optimizer</td>
<td>RMSProp</td>
</tr>
<tr>
<td>RMSProp Decay Rate</td>
<td>0.95</td>
</tr>
<tr>
<td>Update Batch Size</td>
<td>10</td>
</tr>
</tbody>
</table>

Figure 7: HyperParameters.

Conclusion

• Trained agents able to beat hard-coded computer player with a mean score > 10 and score 25+ mean score in breakout.
• Results awaited for reward shaping analysis to make agent learn tunneling strategy in breakout.

References


Contact Information

Web: https://github.com/kshitiz8
Email: ktripathi8@gmail.com