Summary & Goals

This work explores how simple strategies in the game of Leduc Hold’em can be used to beat a sophisticated poker AI, DeepStack. We first analyze, under unbiased training, how significantly DeepStack outperforms most traditional poker-playing strategy profiles employed by humans.

We then consider the ability of an opponent to bias the training phase such that DeepStack is optimized to play against a particular strategy profile. Finally, by allowing for this biasing, we show that DeepStack can be defeated by a subset of strategy profiles if the player can change their strategy post-training. While DeepStack achieves nearly super-human performance, we conclude that DeepStack is susceptible to training poisoning.

Leduc Hold’em Poker

HAND STRENGTHS (BEST TO WORST):
Table Card Match
King
Queen
Jack

6-Card Deck
(2 Jacks, 2 Queens, 2 Kings)

1 Private Card
dealt per player

RISKY
Always Raises
Random Bluffer
Smart Bluffer

INTELLIGENT
Always Calls

Strong Adaptive Rocks
Mild Adaptive Rocks

CONSERVATIVE
Passive Rocks

Rocks Player Game Tree

Input
Layer
Parametric
ReLU
Linear
Parametric
ReLU
Output
Layer

DeepStack Architecture

Summary & Goals

DeepStack Architecture

Training Poisoning in DeepStack

Results

<table>
<thead>
<tr>
<th>Player</th>
<th>Unbiased</th>
<th>Biased</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Mild Adaptive Rocks</td>
<td>-11.4 ± 12</td>
<td>63.6 ± 28</td>
</tr>
<tr>
<td>2 Passive Rocks</td>
<td>-43.5 ± 20</td>
<td>-14.8 ± 38</td>
</tr>
<tr>
<td>3 Strong Adaptive Rocks</td>
<td>-1.7 ± 9.4</td>
<td>-1.9 ± 29</td>
</tr>
<tr>
<td>4 Rocks</td>
<td>-1.5 ± 5.8</td>
<td>4.3 ± 18</td>
</tr>
<tr>
<td>5 Random Bluffer</td>
<td>-55.6 ± 30</td>
<td>6 ± 34</td>
</tr>
<tr>
<td>6 Smart Bluffer</td>
<td>-35.8 ± 21</td>
<td>25.6 ± 32</td>
</tr>
</tbody>
</table>

Table 2. Average Chips Per Game on Biased vs. Unbiased Training. The table reports mean and 95% confidence intervals.

<table>
<thead>
<tr>
<th>Player</th>
<th>Unbiased</th>
<th>Biased</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Mild Adaptive Rocks</td>
<td>192</td>
<td>457</td>
</tr>
<tr>
<td>2 Passive Rocks</td>
<td>328</td>
<td>620</td>
</tr>
<tr>
<td>3 Strong Adaptive Rocks</td>
<td>152</td>
<td>468</td>
</tr>
<tr>
<td>4 Rocks</td>
<td>93.9</td>
<td>260</td>
</tr>
<tr>
<td>5 Random Bluffer</td>
<td>492</td>
<td>544</td>
</tr>
<tr>
<td>6 Smart Bluffer</td>
<td>391</td>
<td>517</td>
</tr>
</tbody>
</table>

Table 3. Standard Deviation of Chips Won

Source code: github.com/rawls238/LeducTrainingPoisoning/