Finding Friend and Foe in Multi-agent Games

Jack Serrino*, Max Kleiman-Weiner*, David Parkes, Josh Tenenbaum

Harvard, MIT, Diffeo

Poster #197
The Resistance: Avalon as a testbed for multi-agent learning and thinking

Recent progress limited to games where teams are known or play is fully adversarial (Dota, Go, Poker).

Avalon (5 Players)
- Two teams: "Spy" and "Resistance"
  - Spies know who is Spy and who is Resistance
    - Goal: plan to sabotage Resistance while hiding their own identity.
  - Resistance only know they are Resistance
    - Goal: learn who is a Spy & who is Resistance.
- Information about intent is often noisy and ambiguous and adversaries may be intentionally acting to deceive.

(Eskridge, 2012)
The Resistance: Avalon as a testbed for multi-agent learning and thinking

Recent progress limited to games where teams are known or play is fully adversarial (Dota, Go, Poker).

(Eskridge, 2012)
The Resistance: Avalon as a testbed for multi-agent learning and thinking

Recent progress limited to games where teams are known or play is fully adversarial (Dota, Go, Poker).

Avalon (5 Players)
- Two teams: “Spy” and “Resistance”
  - **Spies** know who is Spy and who is Resistance
    - Goal: plan to sabotage Resistance while hiding their own identity.
  - **Resistance** only know they are Resistance
    - Goal: learn who is a Spy & who is Resistance.

(Eskridge, 2012)
The Resistance: Avalon as a testbed for multi-agent learning and thinking

Recent progress limited to games where teams are known or play is fully adversarial (Dota, Go, Poker).

Avalon (5 Players)
- Two teams: “Spy” and “Resistance”
  - **Spies** know who is Spy and who is Resistance
    - Goal: *plan* to sabotage Resistance while hiding their own identity.
  - **Resistance** only know they are Resistance
    - Goal: *learn* who is a Spy & who is Resistance.

Information about intent is often noisy and ambiguous and adversaries may be intentionally acting to deceive.

(Eskridge, 2012)
Combining counterfactual regret minimization with deep value networks

- Approach follows DeepStack system developed for NL poker (Moravcik et al, 2017).

*Main contributions:*

- Actions themselves are only partially observed:
  - Deduction required in the loop of learning
- Unconstrained value networks are slower and less interpretable:
  - Develop an interpretable win-probability layer with better sample efficiency.

(Johanson et al, 2012)
Combining counterfactual regret minimization with deep value networks

- Approach follows DeepStack system developed for NL poker (Moravcik et al, 2017).

Main contributions:
- Actions themselves are only partially observed:
  - Deduction required in the loop of learning
Combining counterfactual regret minimization with deep value networks

- Approach follows DeepStack system developed for NL poker (Moravcik et al, 2017).

**Main contributions:**
- Actions themselves are only partially observed:
  - Deduction required in the loop of learning
- Unconstrained value networks are slower and less interpretable:
  - Develop an interpretable win-probability layer with better sample efficiency.

(Johanson et al, 2012)
Deductive reasoning enhances learning when actions are not fully public

```plaintext
procedure CALCTERMINALBELIEF(h, b, \pi_1...p)
   for \rho \in b do
      1. \text{b}_\text{term}[\rho] \leftarrow b[\rho] \prod_i \pi_i(I_i(h, \rho))
      2. \text{b}_\text{term}[\rho] \leftarrow \text{b}_\text{term}[\rho](1 - \mathbb{1}\{h \vdash \neg \rho\}) \quad \triangleright \text{Zero beliefs that are logically inconsistent}
   end for
   return \text{b}_\text{term}
end procedure
```

1. Calculate joint probability of assignment given the public game history

2. Zero out assignments that are impossible given the history.

2) is not necessary in games like Poker, with fully observable actions!
The Win Layer

\[ V(I, \pi^\sigma) \in \mathbb{R}^{n \times |P|} \]

\[ |P| := \text{number of assignments to roles, } \rho \]
\[ n := \text{number of players} \]

Previous approaches:

\[ NN(I, \pi^\sigma) \approx V(I, \pi^\sigma) \]

Our approach:

\[ \tilde{w}(I, \pi^\sigma) = \begin{bmatrix} P(\text{good win}|I, \pi^\sigma, \rho_1) \\ \vdots \\ P(\text{good win}|I, \pi^\sigma, \rho_{|P|}) \end{bmatrix} \in [0, 1]^{|P|} \]

\[ V(I, \pi^\sigma) = f(\tilde{w}(I, \pi^\sigma)) \]
\[ NN(I, \pi^\sigma) \approx \tilde{w}(I, \pi^\sigma) \]

- In 5-player Avalon, 300 values to estimate!
- 60 values to estimate (via sigmoid)
- Correlations are learned imperfectly.
- Correlations are exact.
The Win Layer enables faster + better NN training
DeepRole wins at higher rates than: vanilla-CFR, MCTS, heuristic algorithms

(Wellman, 2006; Tuyls et al 2018)
DeepRole played online in mixed teams of human and bot players w/o communication (1,500+ games)
DeepRole outperformed humans playing online as both a collaborator and competitor.

<table>
<thead>
<tr>
<th></th>
<th>Adding DeepRole or a Human to 4 DeepRole</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>+DeepRole</td>
</tr>
<tr>
<td></td>
<td>Win Rate (%)</td>
</tr>
<tr>
<td>Overall</td>
<td>46.9 ± 0.6</td>
</tr>
<tr>
<td>Resistance</td>
<td>34.4 ± 0.7</td>
</tr>
<tr>
<td>Spy</td>
<td>65.6 ± 0.9</td>
</tr>
</tbody>
</table>
DeepRole outperformed humans playing online as both a collaborator and competitor

<table>
<thead>
<tr>
<th></th>
<th>Adding DeepRole or a Human to 4 DeepRole</th>
<th></th>
<th>Adding DeepRole or a Human to 4 Human</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>+DeepRole</td>
<td>+Human</td>
<td>+DeepRole</td>
</tr>
<tr>
<td></td>
<td>Win Rate (%)</td>
<td>(N)</td>
<td>Win Rate (%)</td>
</tr>
<tr>
<td>Overall</td>
<td>46.9 ± 0.6</td>
<td>(7500)</td>
<td>38.8 ± 1.3</td>
</tr>
<tr>
<td>Resistance</td>
<td>34.4 ± 0.7</td>
<td>(4500)</td>
<td>25.6 ± 1.5</td>
</tr>
<tr>
<td>Spy</td>
<td>65.6 ± 0.9</td>
<td>(3000)</td>
<td>57.8 ± 2.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>60.0 ± 5.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>51.4 ± 8.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>67.4 ± 7.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>48.1 ± 1.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>40.3 ± 1.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>59.7 ± 1.9</td>
</tr>
</tbody>
</table>
DeepRole make rapid accurate inferences about human roles during play and observation.
Finding Friend and Foe in Multi-agent Games

Jack Serrino*, Max Kleiman-Weiner*, David Parkes, Josh Tenenbaum

Harvard, MIT, Diffeo

Poster #197

Play online: ProAvalon.com