How to Safely Exploit Predictions in General-Sum Normal Form Games

Steven Damer
9/30/2014
Outline

- **Problem**: Given a game, and a prediction of opponent behavior, how should you act?
- **Solution**: Create a modified game which reflects your confidence in the prediction, and play a r-safe Stackelberg equilibrium
- **Content**:
  - How to calculate solution, and some sample results
  - An interesting property of the solution
  - Experimental results
- **Questions**:
  - How do you choose a parameter value?
  - Can cooperative behavior be incorporated into this?
Rock-Spock-Paper-Lizard-Scissors

1 point for winning
-1 point for losing
+.5 points for blue arrows
-.5 points for red arrows
Related Work: Safe Policy Selection

- Given a prediction, a game, and a risk value $e$
- Maximize payoff vs. the prediction, while guaranteeing payoff within $e$ of Minimax
- Adjust $e$ according to expected performance vs. opponents move
- Easy to calculate, since you’re not worried about equilibria
Safe Policy Selection Example

- Game: Rock/Spock/Paper/Lizard/Scissors
- Prediction: 100% Rock
- Safety Factor: .5
- Solution: 2/3 Spock, 1/3 Scissors
  - Payoff vs. Prediction: 5/6
  - Worst Case Payoff: -1/2 (Lizard or Spock)
Related Work: Restricted Nash Response

- Given a game, a prediction, and a weight $w$ on the prediction
- Construct a modified game in which the opponent is forced to play the prediction with probability $w$
- Play the Nash equilibrium of the modified game
- Produces a unique response in a zero-sum game
- For any $w$, there is a value of $e$ which produces the same strategy
## Restricted Nash Response Example

- **Game:** Rock/Spock/Paper/Lizard/Scissors
- **Prediction:** 100% Rock
- **Weight:** 25%
- **Solution:** .4 Spock, .6 Scissors

<table>
<thead>
<tr>
<th></th>
<th>Rock</th>
<th>Spock</th>
<th>Paper</th>
<th>Lizard</th>
<th>Scissors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rock</td>
<td>0(0),0</td>
<td>-.375(-.5),1.5</td>
<td>-1.125(-1.5),.5</td>
<td>.375(.5),-1.5</td>
<td>1.125(1.5),-.5</td>
</tr>
<tr>
<td>Spock</td>
<td>1.5(1.5),-.5</td>
<td>.375(0),0</td>
<td>0(-.5),1.5</td>
<td>-.75(-1.5),.5</td>
<td>.75(.5),-1.5</td>
</tr>
<tr>
<td>Paper</td>
<td>.5(.5),-1.5</td>
<td>1.25(1.5),-.5</td>
<td>.125(0),0</td>
<td>-.25(-.5),1.5</td>
<td>-1(-1.5),.5</td>
</tr>
<tr>
<td>Lizard</td>
<td>-1.5(-1.5),.5</td>
<td>0(.5),-1.5</td>
<td>.75(1.5),-.5</td>
<td>-.375(0),0</td>
<td>-.75(-.5),1.5</td>
</tr>
<tr>
<td>Scissors</td>
<td>-.5(-.5),1.5</td>
<td>-1.25(-1.5),5</td>
<td>.25(.5),-1.5</td>
<td>1(1.5),-.5</td>
<td>-.125(0),0</td>
</tr>
</tbody>
</table>
Restricted Stackelberg Response with Safety

- Given a game, a prediction, a weight $w$, and a risk value $r$
- Construct a modified game from the game, prediction, and weight
- Play a Stackelberg equilibrium in the modified game while guaranteeing worst case performance within $r$ of Minimax value
Restricted Stackelberg Response Example

- Game: Rock/Spock/Paper/Lizard/Scissors
- Prediction: 100% Rock
- Safety Factor: .5
- Solution: .6 Spock, .4 Scissors

<table>
<thead>
<tr>
<th></th>
<th>Restricted Nash Response</th>
<th>Restricted Stackelberg Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>vs. Prediction</td>
<td>-.3</td>
<td>.7</td>
</tr>
<tr>
<td>vs. Best Response</td>
<td>.7</td>
<td>.7</td>
</tr>
<tr>
<td>vs. Worst Case</td>
<td>-.6</td>
<td>-.6</td>
</tr>
</tbody>
</table>
Calculating RSRS

- For each opponent move M, find strategy which maximizes payoff vs. prediction, subject to:
  - Worst case payoff no worse than Minimax-r
  - M is the best response for the opponent to the strategy
- Play the strategy with the highest payoff
Effect of Safety Value

- 0 - Play Minimax solution
- $\infty$ - Play Best Response
- Change in Safety value results in continuous change in strategy
Effect of Prediction Weight

- 0 - Play Stackelberg Response
- 1 - Play Best Response
- Change in Prediction Weight may result in discrete change in chosen strategy
- Safety value dominates prediction weight
## Sample Solutions

<table>
<thead>
<tr>
<th>Strategy</th>
<th>Distribution</th>
<th>vs. Prediction</th>
<th>vs. Best Response</th>
<th>vs. Worst Case</th>
</tr>
</thead>
<tbody>
<tr>
<td>r=0</td>
<td>(.2,.2,.2,.2,.2)</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>r=1.5,w=1</td>
<td>(0,1,0,0,0)</td>
<td>1.5</td>
<td>-.5</td>
<td>-1.5</td>
</tr>
<tr>
<td>r=1.5,w=0</td>
<td>(0,6,0,0,.4)</td>
<td>.7</td>
<td>.7</td>
<td>-.6</td>
</tr>
<tr>
<td>r=.1,w=.5</td>
<td>(.142,.233,.269,.051,.305)</td>
<td>.254</td>
<td>.045</td>
<td>-.1</td>
</tr>
<tr>
<td>r=.5,w=1</td>
<td>(0,.66,0,0,.33)</td>
<td>.8333</td>
<td>-.16</td>
<td>-.5</td>
</tr>
<tr>
<td>r=1.5,w=.8</td>
<td>(.6,0,0,.4,0)</td>
<td>.15</td>
<td>.7</td>
<td>-.6</td>
</tr>
<tr>
<td>r=1.5,w=1</td>
<td>(0,0,0,1,0)</td>
<td>.375</td>
<td>-.5</td>
<td>-1.5</td>
</tr>
<tr>
<td>r=.6,w=1</td>
<td>(0,.127,0,.355,.518)</td>
<td>.15</td>
<td>-.6</td>
<td>-.6</td>
</tr>
</tbody>
</table>
Stackelberg vs. Nash

![Graph showing expected payoffs in Stackelberg vs. Nash scenarios. The x-axis represents prediction weights, and the y-axis represents expected payoffs. Different lines represent different comparisons, such as RSRS vs. RNR, with various weight configurations.]
Performance Trade-off

- Strategies change discontinuously with $w$
- Increasing $w$ increases performance vs. prediction
- Strategies change when opponent’s best-response changes
- Relative change in payoffs is proportional to $w$ value
Choosing Prediction Weight

- Arbitrary value
- Equivalent safety value (RNR)
- Relative probability of best response and prediction
  - Best response based on logistic response function, to avoid a brittle opponent model
- Use Exponentially Weighted forecaster to blend multiple results
Choosing Safety Value

- Pick a maximum amount to risk
- Each round add $1/n$ to the budget
- Each round reduce the budget by the amount lost
- Same approach as SPS
Results

- Games repeated 100 times
- Results averaged across 100 runs
- Predictor is fictitious play (flawed, but simple)
- Performance compared between RNR, SPS, Best Response, RSR, and RSRS
Fixed Opponent
Random Opponent
Omniscient Opponent
Worst-Case Opponent
Switching Opponent

Outcomes vs Switcher Opponent

- Best Responder
- Safe Policy Selection
- Restricted Nash Response
- Restricted Stackelberg Response
- Restricted Stackelberg Response with Safety

Score vs Round

10 20 30 40 50 60 70 80 90 100
Battle of the Sexes
Chicken
Pursuer/Evader
Traveller’s Dilemma
Weight Learning vs Exponentially Weighted Average
Pessimistic RSRS
More Pessimistic RSRS
Future Work

● Extend attitude-based play
● Choosing parameter values
  ○ Exponentially Weighted Averages is efficient, but doesn’t provide a value
● Convergence Properties + Performance Guarantees
  ○ In self-play
  ○ What guarantees are appropriate? (Stackelberg Game)
More graphs