Multi-Player UCT

Nathan Sturtevant

Research Overview

• How does UCT extend to multi-player games?
• How does UCT perform in multi-player games?
• How do UCT enhancements perform in multi-player games?

Background

• Max$^n$ (Luckhardt & Irani, 1985)
  • Computes an equilibrium strategy
• Paranoid (Sturtevant & Korf, 2000)
  • Reduces a game to two-player game
  • Improves pruning
  • Special case of max$^n$

Background: UCT

$X_i = C \sqrt{\frac{T}{T_i}}$

• UCT provides a rule for selecting the next node to explore in a monte-carlo simulation
  • Based only on the player to move at each node
  • “Trivial” to expand to multiple players
• Backup $n$-tuple of scores
• What computation is UCT performing?
  • Assume unlimited expansions
Multi-Player UCT

- UCT computes a strategy that is in equilibrium
  - No single player can gain by deviating, assuming payoffs are perfectly accurate
- Strategy may be mixed
  - May not actually play in a mixed way
  - Assumption of mixed play can change the strategy played
Experimental Results

- Compare to existing (max\(^n\), paranoid) algorithms
- Evaluate UCT enhancements (Gelly & Silver, 2007)
  - RAVE
  - Pre-initialization of data
  - Playout policies

Domains

- Chinese Checkers
- Hearts
- Spades

Chinese Checkers

- Race to get across the board
- Pre-computed table of shortest single-player distance
- 17 moves to solve single-agent problem optimally
- Minimize distance from goal or maximize difference in distance?

<table>
<thead>
<tr>
<th></th>
<th>UCT</th>
<th>Par(_{\text{diff}})</th>
<th>Par(_{\text{dist}})</th>
<th>Max(<em>0)(</em>{\text{diff}})</th>
<th>Max(<em>0)(</em>{\text{dist}})</th>
</tr>
</thead>
<tbody>
<tr>
<td>UCT</td>
<td>-</td>
<td>92.0</td>
<td>96.0</td>
<td>96.3</td>
<td>94.0</td>
</tr>
<tr>
<td>Par(_{\text{diff}})</td>
<td>8.0</td>
<td>-</td>
<td>53.7</td>
<td>75.0</td>
<td>63.0</td>
</tr>
<tr>
<td>Par(_{\text{dist}})</td>
<td>4.0</td>
<td>46.3</td>
<td>-</td>
<td>53.7</td>
<td>31.3</td>
</tr>
<tr>
<td>Max(<em>0)(</em>{\text{diff}})</td>
<td>3.7</td>
<td>25.0</td>
<td>46.3</td>
<td>-</td>
<td>43.7</td>
</tr>
<tr>
<td>Max(<em>0)(</em>{\text{dist}})</td>
<td>6.0</td>
<td>36.7</td>
<td>68.7</td>
<td>56.3</td>
<td>-</td>
</tr>
</tbody>
</table>
### Chinese Checkers - 250k Node exp.

<table>
<thead>
<tr>
<th></th>
<th>UCT</th>
<th>Par\text{diff}</th>
<th>Par\text{dist}</th>
<th>Max\text{diff}</th>
<th>Max\text{dist}</th>
</tr>
</thead>
<tbody>
<tr>
<td>UCT</td>
<td>-</td>
<td>92.0</td>
<td>96.0</td>
<td>96.3</td>
<td>94.0</td>
</tr>
<tr>
<td>Paranoid\text{diff}</td>
<td>8.0</td>
<td>-</td>
<td>53.7</td>
<td>75.0</td>
<td>63.3</td>
</tr>
<tr>
<td>Paranoid\text{dist}</td>
<td>4.0</td>
<td>46.3</td>
<td>-</td>
<td>53.7</td>
<td>31.3</td>
</tr>
<tr>
<td>Max\text{diff}</td>
<td>3.7</td>
<td>25.0</td>
<td>46.3</td>
<td>-</td>
<td>43.7</td>
</tr>
<tr>
<td>Max\text{dist}</td>
<td>6.0</td>
<td>36.7</td>
<td>68.7</td>
<td>56.3</td>
<td>-</td>
</tr>
</tbody>
</table>

### Chinese Checkers

<table>
<thead>
<tr>
<th></th>
<th>UCT</th>
<th>Par\text{diff}</th>
<th>Par\text{dist}</th>
<th>Max\text{diff}</th>
<th>Max\text{dist}</th>
</tr>
</thead>
<tbody>
<tr>
<td>UCT</td>
<td>-</td>
<td>92.0</td>
<td>96.0</td>
<td>96.3</td>
<td>94.0</td>
</tr>
<tr>
<td>Paranoid\text{diff}</td>
<td>8.0</td>
<td>-</td>
<td>53.7</td>
<td>75.0</td>
<td>63.3</td>
</tr>
<tr>
<td>Paranoid\text{dist}</td>
<td>4.0</td>
<td>46.3</td>
<td>-</td>
<td>53.7</td>
<td>31.3</td>
</tr>
<tr>
<td>Max\text{diff}</td>
<td>3.7</td>
<td>25.0</td>
<td>46.3</td>
<td>-</td>
<td>43.7</td>
</tr>
<tr>
<td>Max\text{dist}</td>
<td>6.0</td>
<td>36.7</td>
<td>68.7</td>
<td>56.3</td>
<td>-</td>
</tr>
</tbody>
</table>

### Chinese Checkers - Playout Policy

- Always play the move that makes the most progress across the board
- Decreases average playout length
  - 80 moves (27 per player)
  - 200 moves (67 per player)
- Increases player strength
  - 81% of games won by new policy given the same number of simulations
Chinese Checkers

- RAVE (History Heuristic)
  - Ineffective
  - Pre-initializing states
  - Use database
  - Also ineffective

Hearts

- Trick-based card game
  - 4 players
  - Every game is exactly 52 moves long
  - Every card is played exactly once in the game
  - Goal is to minimize the points taken
  - Get 0 points for “shooting the moon”

Hearts - Results

- Shooting the moon test
  - Which algorithm is most effective in stopping players from shooting the moon?
  - 3,244 test problems

Preventing Shooting the Moon

<table>
<thead>
<tr>
<th></th>
<th>UCT</th>
<th>Max(h) Learned</th>
<th>Random</th>
<th>Max(h) Hand-tuned</th>
</tr>
</thead>
<tbody>
<tr>
<td>total</td>
<td>250</td>
<td>312</td>
<td>411</td>
<td>1377</td>
</tr>
<tr>
<td>percentage</td>
<td>7.70%</td>
<td>9.62%</td>
<td>12.67%</td>
<td>42.45%</td>
</tr>
</tbody>
</table>
Quality of Play vs. UCT

<table>
<thead>
<tr>
<th></th>
<th>Learned</th>
<th>Max$^0$</th>
<th>Random</th>
</tr>
</thead>
<tbody>
<tr>
<td>UCT score</td>
<td>46.12</td>
<td>51.77</td>
<td>16.31</td>
</tr>
<tr>
<td>vs. score</td>
<td>67.30</td>
<td>88.31</td>
<td>89.23</td>
</tr>
<tr>
<td>win%</td>
<td>83.9%</td>
<td>88.0%</td>
<td>100%</td>
</tr>
</tbody>
</table>

Hearts - UCT Enhancements

- Playout policies
  - Most policies ineffective in increasing strength of play
- Pre-initialization
  - Only effective with very few simulations
- RAVE / History Heuristic
  - Also not effective

Spades

- Play 3-player version of Spades
- Bid on tricks that will be taken in the game
- Delayed penalty for overbidding
- Previous work dominated by opponent modeling
- What strategy do players use to cope with overbidding?

<table>
<thead>
<tr>
<th>Player 1</th>
<th>Player 2</th>
<th>P1 Avg</th>
<th>P2 Avg</th>
<th>P1 Win %</th>
</tr>
</thead>
<tbody>
<tr>
<td>mOT$_{MT}$</td>
<td>MT$_{mOT}$</td>
<td>231.84</td>
<td>171.48</td>
<td>67.0%</td>
</tr>
<tr>
<td>mOT$_{mOT}$</td>
<td>MT$_{MT}$</td>
<td>179.19</td>
<td>212.76</td>
<td>43.0%</td>
</tr>
<tr>
<td>mOT$_{gen}$</td>
<td>prob-max$^n$</td>
<td>212.60</td>
<td>202.67</td>
<td>53.2%</td>
</tr>
</tbody>
</table>
Conclusions

- UCT works very well in multi-player games
  - UCT enhancements not as well

Future work
- Find ways to improve UCT performance
- Better handle imperfect information