Multi-Player UCT

Nathan Sturtevant

DEPARTMENT OF COMPUTING SCIENCE
UNIVERSITY OF ALBERTA

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?

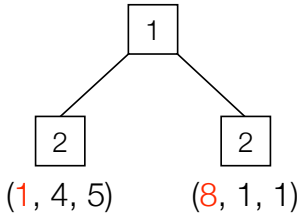
Background

- 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

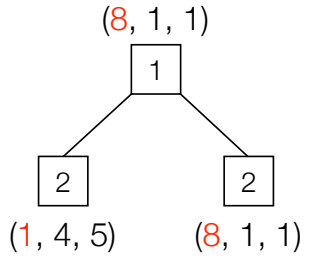
Background: UCT $X_i \quad C\sqrt{\frac{T}{T_i}}$

- 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

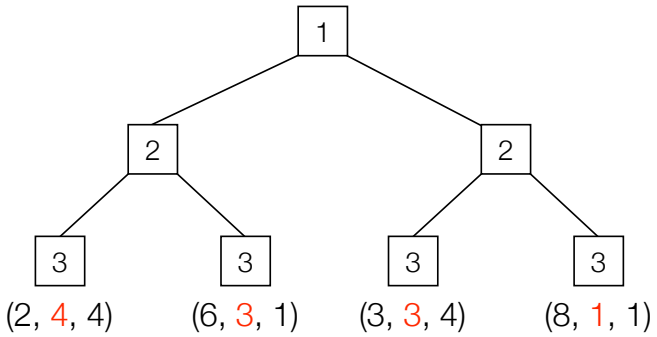
Simple Tree (1)



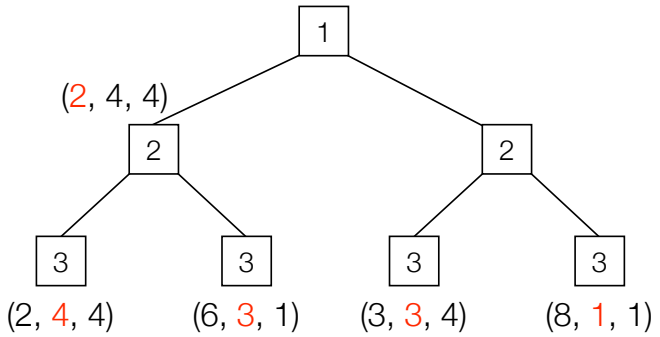
Simple Tree (1)



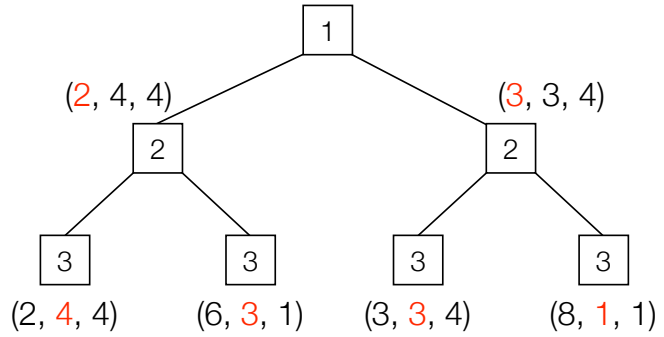
Simple Tree (2)



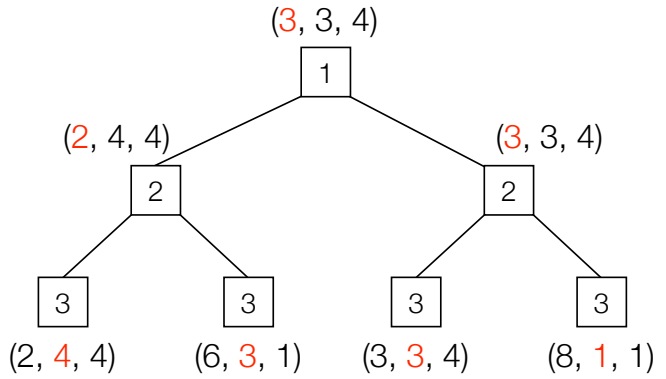
Simple Tree (2)



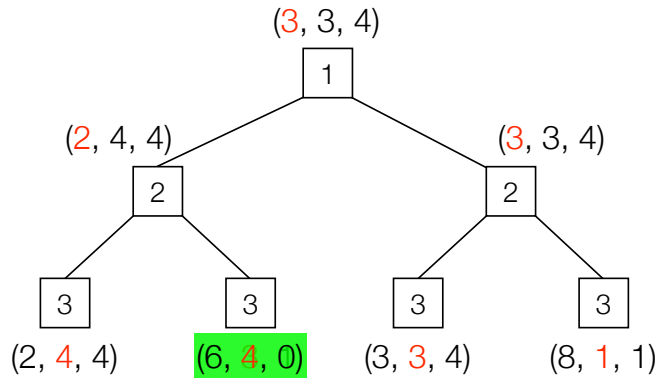
Simple Tree (2)



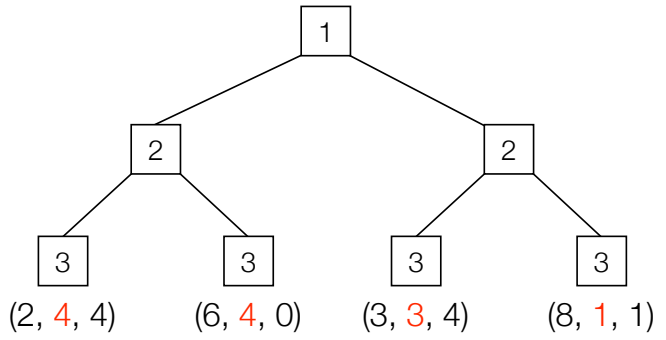
Simple Tree (2)



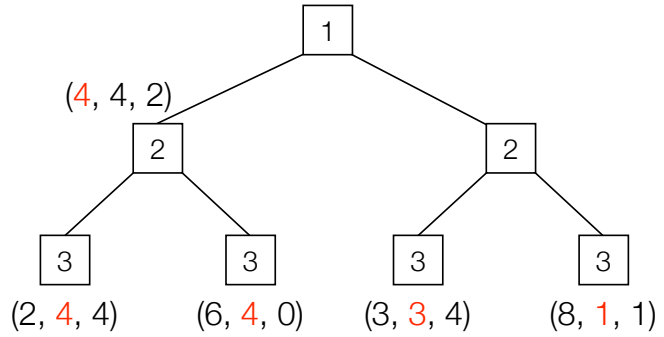
Simple Tree (2)



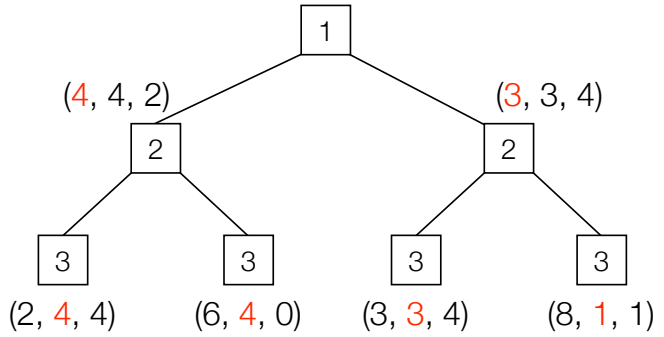
Complex Tree (1)



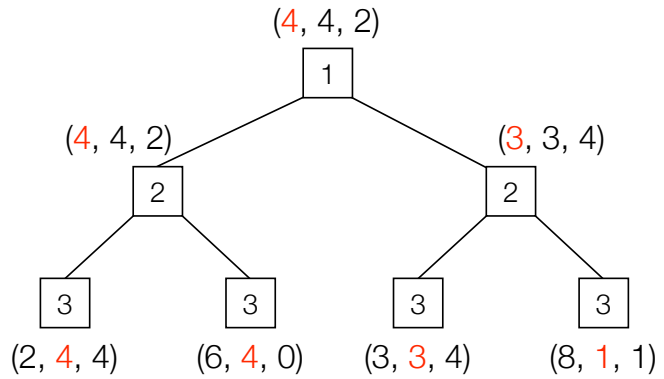
Complex Tree (1)



Complex Tree (1)



Complex Tree (1)



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

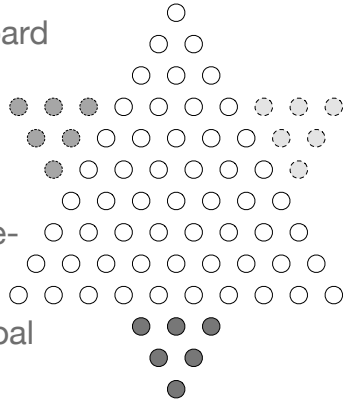
- Compare to existing (max^n, paranoid) algorithms
- Evaluate UCT enhancements (Gelly & Silver, 2007)
  - RAVE
  - Pre-initialization of data
  - Playout policies

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?



Chinese Checkers

	UCT	Par_{diff}	Par_{dist}	Max^n_{diff}	Max^n_{dist}
UCT	-	92.0	96.0	96.3	94.0
$Paranoid_{diff}$	8.0	-	53.7	75.0	63.3
$Paranoid_{dist}$	4.0	46.3	-	53.7	31.3
Max^n_{diff}	3.7	25.0	46.3	-	43.7
Max^n_{dist}	6.0	36.7	68.7	56.3	-

Chinese Checkers - 250k Node exp.

	UCT	Par_{diff}	Par_{dist}	Max^n_{diff}	Max^n_{dist}
UCT	-	92.0	96.0	96.3	94.0
$Paranoid_{diff}$	8.0	-	53.7	75.0	63.3
$Paranoid_{dist}$	4.0	46.3	-	53.7	31.3
Max^n_{diff}	3.7	25.0	46.3	-	43.7
Max^n_{dist}	6.0	36.7	68.7	56.3	-

Chinese Checkers

	UCT	Par_{diff}	Par_{dist}	Max^n_{diff}	Max^n_{dist}
UCT	-	92.0	96.0	96.3	94.0
$Paranoid_{diff}$	8.0	-	53.7	75.0	63.3
$Paranoid_{dist}$	4.0	46.3	-	53.7	31.3
Max^n_{diff}	3.7	25.0	46.3	-	43.7
Max^n_{dist}	6.0	36.7	68.7	56.3	-

Chinese Checkers

	UCT	Par_{diff}	Par_{dist}	Max^n_{diff}	Max^n_{dist}
UCT	-	92.0	96.0	96.3	94.0
$Paranoid_{diff}$	8.0	-	53.7	75.0	63.3
$Paranoid_{dist}$	4.0	46.3	-	53.7	31.3
Max^n_{diff}	3.7	25.0	46.3	-	43.7
Max^n_{dist}	6.0	36.7	68.7	56.3	-

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

	UCT	Max^n Learned	Random	Max^n Hand-tuned
total	250	312	411	1377
percentage	7.70%	9.62%	12.67%	42.45%

Quality of Play vs. UCT

	Learned	Max^n	Random
UCT score	46.12	51.77	16.31
vs. score	67.30	88.31	89.23
win%	83.9%	88.0%	100%

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?

Spades

Player 1	Player 2	P1 Avg	P2 Avg	P1 Win %
mOT_{MT}	MT_{mOT}	231.84	171.48	67.0%
mOT_{mOT}	MT_{MT}	179.19	212.76	43.0%
mOT_{gen}	prob-max^n	212.60	202.67	53.2%

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