Exploring Self-Teaching Game-Playing Algorithms

Abigail Santiago, Benjamin Santiago, Dr. Raghu Ramanujan

Introduction:

- We compare decision-making in self-taught algorithms to human-designed algorithms
- We compare decision-making in Chess using two different algorithms, Minimax and Monte Carlo Tree Search

Background:

Minimax and Stockfish-

- Minimax algorithm: creates a game tree (Fig. A) and assigns values to certain outcomes, maximizer and minimizers pick their preferred value (small or large) (Fig. B)
- Stockfish uses, minimax, which relies heavily on human-programmed heuristics to make decisions

Monte Carlo Tree Search (MCTS) and LeelaChess0-

- MCTS: Given a game state, selects a move based on the UCB1 formula (Fig. E) until a leaf node is reached, simulates an entire game, backpropagates an answer and assigns a score to the outcome
- Lc0 draws from machine learning, which is concerned with creating programs that learn and improve from experience
- MCTS does not rely on human-specific knowledge, and when combined with other machine learning techniques (regularization, neural networks), can create a self-teaching algorithm

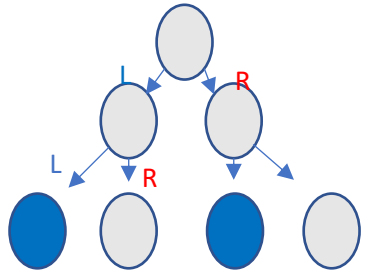


Fig. A: A simple turn based game tree, with the moves being either right or left

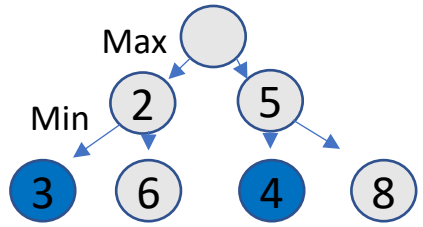


Fig. B: Minimax. Max will choose the higher value, min the smaller. Can be pruned to improve run time.

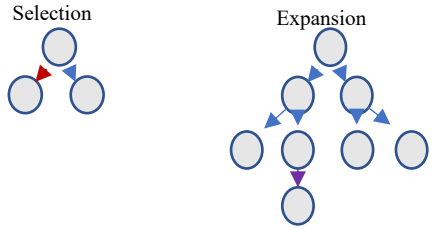
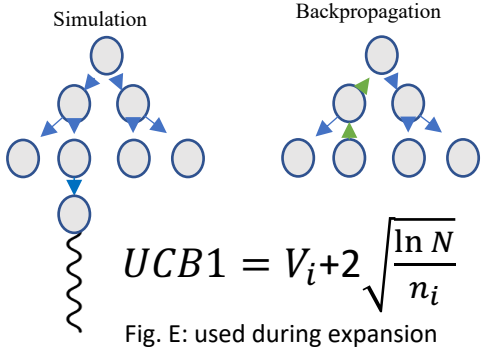


Fig. C: A Monte-Carlo tree search. Selects move, expands, simulates game, and backpropagates score



$$UCB1 = V_i + 2\sqrt{\frac{\ln N}{n_i}}$$

Fig. E: used during expansion to select moves until a leaf node is reached

Methodology and Results:

- Chess boards were sampled from games between human grandmasters (boards found on http://wtharvey.com/)
- The chess programs analyzed the boards to determine the best next move
- Lc0, Stockfish, and the human agreed 73.7% of the time (Fig. D)

Conclusion:

- Lc0, a community sourced program, produces solutions comparable to that of the current best chess engine
- Lc0 sometimes aligns with Stockfish (84.5%), and at other times it aligns with the human (90.5%), and rarely disagrees with both
- This suggests that Lc0 is a sort of "hybrid" in terms of decision making, combining features from both Stockfish and the human

Future Research:

- The hybrid process of Lc0 opens up the idea that perhaps one can combine the recommendations from Stockfish, Lc0 and/or a human to create an even stronger player

	Lc0	Human	Stockfish
Lc0	--------	90.5%	84.5%
Human	90.5%	--------	93.0%
Stockfish	84.5%	93.0%	--------

Fig. D: A table showing the percent of chess boards each program agreed upon